70

The Genome of the Ammonia Oxidizing Bacterium *Nitrosomonas europaea*: Iron Metabolism and Barriers to Heterotrophy

Xueming Wei, Neeraja Vajrala, Norman Hommes, Luis Sayavedra-Soto*, and Daniel Arp (arpd@science.oregonstate.edu)

Oregon State University, Corvallis, OR

Nitrosomonas europaea is an aerobic lithoautotrophic bacterium that uses ammonia (NH_3) as its energy source (3). As a nitrifier, it is an important participant in the N cycle, which can also influence the C cycle. The genome sequence of *N. europaea* has been annotated and consists of approximately 2460 protein-encoding genes (1). We are continuing to use the genome sequence to explore the genetic structure and mechanisms underlying the lithoautotrophic growth style of *N. europaea*. Currently, we are investigating its Fe requirements and its possible barriers to utilizing carbon sources different from CO_2.

Because *N. europaea* has a relatively high content of hemes, sufficient Fe must be available in the medium for it to grow. The genome revealed that approximately 5% of the coding genes in *N. europaea* are dedicated to Fe transport and assimilation. Nonetheless, with the exception of citrate, *N. europaea* lacks genes for siderophore production (1). We have initiated the study on this intriguing facet by determining the Fe requirements for growth and are characterizing the expression of the putative membrane siderophore receptors.

N. europaea changes its heme composition when Fe is at a relatively low concentration. Biochemical analyses showed that cytochrome and heme contents of cells grown in Fe-limited medium were 4 fold lower than those from Fe-rich medium. Cellular Fe contents (in both membrane and soluble fractions) showed the same trend. The activity of hydroxylamine oxidoreductase was over three fold lower in cells grown in Fe-limited medium than that in full medium. The growth yields at 0.1 μM Fe and at 0.2 μM Fe were about 35% and 65% respectively of that observed at 10 μM Fe (full medium). *N. europaea* has the mechanisms to cope and grow under Fe limitation.

The *N. europaea* genome revealed that there are over 26 sets of genes that are organized similarly to the genes in a *fecR/fecI* system. Through similarity searches, we have identified possible TonB-dependent receptor genes up- or downstream of these sets. Some of these are similar to genes encoding the siderophore receptors for desferrioxamine (desferal), ferrichrome, and coprogen.

The addition of desferal in Fe-limiting medium promoted the growth of *N. europaea*, though with a longer lag phase, suggesting a necessary induction period of the corresponding receptor. A gene for the putative desferal outer membrane receptor was identified by similarity searches (NE1097, a *foxA* homologue). NE1097 was expressed at a higher level (>10 fold) in Fe-limiting, desferal-containing cultures than in Fe-sufficient cultures. The expression of NE1097 required the presence of desferal, since typical lag phases were observed when inoculants from desferal cultures were used. Several membrane proteins detected only in the cells grown in Fe-limited medium may be involved in Fe

transport. For example, a membrane peptide with the calculated MW of the putative desferal receptor was observed only in the cells grown in desferal-containing medium. Ferric citrate had an effect similar to that of desferal on *N. europaea* growth in Fe-limiting medium, but with a longer lag phase and a higher final cell density than that in the full medium. Ferrichrome, on the other hand, did not prolong the lag phase, yet increased total cell growth, suggesting that the genes for the ferrichrome receptors were expressed constitutively.

Consistent with the genome sequence data, no siderophores were detected in *N. europaea* culture filtrates under either Fe-limiting or Fe-sufficient conditions using a standard siderophore assay. We considered the possibility that citrate serves as a Fe chelator/siderophore, since *N. europaea* has the necessary genes to produce it. Citrate was detected (2 to 5 μM) in cell-free filtrates from both, low- and full Fe cultures. Surprisingly, cell-free filtrates from full Fe cultures had relatively higher concentrations (5 μM) of citrate than in low Fe cultures (2 to 3 μM). The role of citrate in Fe acquisition, if any, is yet to be determined. *N. europaea* apparently expresses siderophore receptors (i.e. NE1097) under low Fe conditions to scavenge Fe more efficiently. These results reinforce the notion that *N. europaea* uses siderophores produced by other organisms in natural habitats.

Genes encoding the putative outer membrane desferal receptor (NE1097 and NE1088, *foxA* homologues) have been cloned, insertional mutant constructs made, and mutant strains obtained through homologous recombination. Physiological and genetic characterization of these mutants is in progress.

In addition to the Fe experiments, analysis of the *N. europaea* genome has led to experiments probing the possible barriers to heterotrophy in *N. europaea*. The genome contains genes that are similar to the genes encoding fructose transport systems (PTS-type) in other bacteria. Furthermore, *N. europaea* can use fructose as the only source of carbon for growth (2). However, not all the genes required for an active PTS system are present in the genome. The inactivation of the two identified PTS genes did not affect growth on fructose or cause any other growth phenotype. Fructose may enter the cells by some other means. The role of the existing PTS genes remains unclear.

Historically, the activity of the enzyme 2-oxoglutarate dehydrogenase has not been detected in *N. europaea*. The lack of this activity was believed to be the cause for the obligate autotrophy of *N. europaea*. However, the genomic sequence reveals that the three genes necessary to encode this enzyme are present. We inactivated the first gene (odhA) in the operon of this enzyme. The mutant strain grew similarly to wild-type cells during exponential growth. However, in late stationary phase or under ammonia starvation (i.e. energy-limiting conditions), mutant strains lost viability faster and recovered more slowly upon addition of more ammonia as compared to the wild type. This suggests that 2-oxoglutarate dehydrogenase may be involved in processes occurring during the stationary growth phase of *N. europaea*.

A gene encoding a putative ammonia transporter (*amt*) is present in the genome. However, the strain with this gene inactivated showed no difference in growth to wild-type cells over a wide range of ammonium concentrations. The function of *amt* in *N. europaea* is still unknown.

We are exploring the idea that one barrier to heterotrophic growth in *N. europaea* may be due to a lack of transporters for alternative growth substrates. The genes encoding the enzymes to utilize glycerol as the carbon source are present, but the genes encoding a glycerol transporter are not. The heterologous expression of the gene for the glycerol permease from *E. coli* in *N. europaea* permits *N. europaea* to utilize glycerol as the carbon source.

Acknowledgment: This research is funded by grants from the Department of Energy (ER63149-1017305-0007065 and ER63781-1023295-0009824) and resources from the Oregon Agricultural Experiment Station.

 * Presenting author

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. Hommes, N. G., L. A. Sayavedra-Soto, and D. J. Arp. 2003. Chemolithoorganotrophic growth of *Nitrosomonas europaea* on fructose. *J Bacteriol* 185:6809-6814.
3. 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.

71

Pelagibacter ubique: A Post-Genomic Investigation of Carbon Metabolism and Photochemistry in an Extraordinarily Abundant Oceanic Bacterium

Stephen J. Giovannoni[1] * (steve.giovannoni@oregonstate.edu), Lisa Bibbs[2], James Tripp[1], Scott Givan[1], Jang-Cheon Cho[1], Martha D. Stapels[3], Russell Desiderio[1], Mercha Podar[2], Kevin L. Vergin[1], Mick Noordeweir[2], Michael S. Rappé[4], Samuel Laney[1], Douglas F. Barofsky[1], and Eric Mathur[2]

[1]Oregon State University, Corvallis, OR; [2] Diversa Corporation, San Diego, CA; [3]Waters Corporation, Milford, MA; and [4]Hawaii Institute of Marine Biology, Kaneohe, HI

The alphaproteobacterium SAR11, now known as *Pelagibacter ubique,* is arguably the most abundant organism in the oceans, where it accounts for approximately 25% of all microbial plankton cells. During summer periods it may exceed 50% of the cells in the surface waters of temperate ocean gyres. *Pelagibacter* plays a key role in the oxidation of the oceanic dissolved organic carbon pool, which is approximately equivalent in size to the atmospheric carbon dioxide pool. The first cultured strains of *Pelagibacter* were isolated by high throughput methods for culturing cells by dilution into natural seawater media, and screening using a new cell array technology. *Pelagibacter* cultures are routinely propagated in autoclaved seawater, where they attain cell densities that are typical of native populations (ca. 10^6 cells/ml). During the sequencing of the 1.3 million base pair genome of *Pelagibacter ubique*, in collaboration with Diversa Corp., it was discovered that this organism has a proteorhodopsin (PR) gene. Liquid chromatography and tandem mass spectrometry were used to prove that the *Pelagibacter* PR gene is expressed in culture and that an identical protein is abundant in coastal Oregon seawater. Laser flash excitation experiments with whole, cultured cells revealed absorption transients with decay kinetics characteristic of retinylidene ion pumps, and light-dependent drops in pH provided confirmation that this PR is a light-dependent proton pump. *Pelagibacter ubique* is the first cultured bacterial isolate to exhibit the PR genes discovered by Bejá, Delong, and coworkers, and the only experimental choice at present for understanding how light-dependent proton pumps influence the efficiency of dissolved organic carbon (DOC) assimilation by heterotrophic bacteria in the ocean surface. The *Pelagibacter* genome is almost exactly the size of the genomes of the obligate intracellular parasites *R. conorii* and *W. pipientis*, but it appears to encode a relatively complete metabolic repertoire governed by unusually simple regulatory circuits. One objective of our current research is to predict the organic carbon sources used by *Pelagibacter* by metabolic reconstruction. Another major thrust of our research is the application of mass spectrometry methods to understand the regulatory responses of *Pelagibacter* to environmental variables, and to explore the proteome state

of *Pelagibacter* cells in the oceans, so that they can be used as proxies to report the biological state of the system. Metabolic modeling of *Pelagibacter* is an attractive long range goal because it is one of the smallest and simplest cells known. Its remarkable success may be attributable to the integration and optimization of metabolic processes for efficiency at low nutrient fluxes..

72

Does the Three Dimensional Organization of the Nucleoid of the *Deinococcaceae* Contribute to their Ionizing Radiation Resistance?

J. M. Zimmerman and J. R. Battista* (jbattis@lsu.edu)

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

Transmission electron micrographs of *Deinococcus radiodurans* R1 suggest that the nucleoid of this species exists as a toroidal ring, and have led to speculation that this structure facilitates the extreme radioresistance of this species. However, little direct evidence supports this contention. Since extreme radioresistance is characteristic of all the members of the *Deinococcaceae*, we hypothesize that if nucleoid morphology contributes to radioresistance, the genomic DNA of each species should form similar structures. Using epifluorescence and deconvolution microscopy, we evaluated the nucleoid morphologies of eight of the nine validly described species of *Deinococcus*, the radioresistant bacterium *Rubrobacter radiotolerans*, and the less radioresistant *Thermus aquaticus*, a distant relative of the deinococci. Although the nucleoids of *Deinococcus murrayi, Deinococcus proteolyticus, Deinococcus radiophilus,* and *Deinococcus grandis* have structures similar to *D. radiodurans*, the nucleoids of *Deinococcus radiopugnans* and *Deinococcus geothermalis* lack specific organization. The nucleoid of *R. radiotolerans* consists of multiple highly condensed spheres of DNA. Since only five of the seven recognized deinococcal species exhibit a structurally distinct nucleoid, we conclude there is no obvious relationship between the three dimensional organization of genomic DNA and extreme radioresistance. However, the genomic DNA of all extremely radioresistance species is highly condensed relative to the more radiosensitive species examined. We have examined nucleoid structure following the introduction of DNA double strand breaks and show that the shape of the nucleoid does not demonstrably change in radioresistant species even in strains incapable of repairing strand breaks, suggesting that DNA held in this tightly packed configuration contributes to the radioresistance of these bacteria.

* Presenting author

73

Large Scale Genomic Analysis for Understanding Hydrogen Metabolism in *Chlamydomonas reinhardtii*

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

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

Many taxonomically diverse microorganisms have the ability to produce H_2 anaerobically in pathways coupled either to dark fermentation, dark CO-oxidation, light-dependent N_2-fixation, or photosynthetic water oxidation. However, only certain photosynthetic organisms are able to directly couple water oxidation to the photoproduction of H_2. Among these *Chlamydomonas reinhardtii*, a green alga, is able to convert the low potential reductant generated from water by photosynthesis into H_2. A more fundamental understanding of the physiology, biochemistry and genetics of this prototype alga might enable the future development of a sustainable system for H_2 production. Recently, we and our collaborators announced the discovery of a physiological switch (sulfur deprivation), which attenuates photosystem II O_2-evolution activity (Wykoff et al., 1998) and allows *C. reinhardtii* to metabolize O_2 in sealed culture vessels (Melis et al., 2000). This produces an anaerobic environment and leads to the photoproduction of volumetric amounts of H_2 over a 4-day period in batch culture. Under sulfur-deprived conditions, algal cultures are subjected to a mixed metabolic state in which anaerobic fermentation, oxygenic photosynthesis and aerobic respiration co-occur. This physiological state provides us with a unique opportunity to further explore the gene-expression patterns and protein networks that sustain algal H_2 production.

With the completion of the *C. reinhardtii* genome sequence as a part of the DOE Office of Science's Genomics:GTL Program, it is now possible to thoroughly explore large-scale transcript profiles associated with H_2 metabolism in this alga using gene microarrays. High-density DNA microarrays are being used to examine the ways in which WT and mutant strains of *C. reinhardtii* acclimate to conditions that allow for H_2 production. Recently, an array with approximately 3,000 elements was used to examine sulfur deprivation responses in WT and mutant *C. reinhardtii* strains (Zhang et al., 2004). A new array based on specific synthetic 70 mers that represents approximately 10,000 genes has been developed at the Carnegie Institution (Stephan Eberhard and Arthur Grossman, unpublished); it should be ready for use by the beginning of 2005.

We have isolated several independent *C. reinhardtii* mutants with attenuated H_2-photoproduction activity at NREL using a rapid screening technique and will compare gene expression profiles from these mutants with the WT under conditions that facilitate H_2 production. One mutant lacks a functional *HydEF* gene (Posewitz et al., 2004), which encodes a radical SAM protein that is required to insert the metal catalytic center into the hydrogenase enzyme. The *hydEF-1* mutant is the only reported *C. reinhardtii* strain that is unable to produce any H_2 at all. The *hydEF* mutation specifically disrupts hydrogenase activity, but this mutant has no discernable phenotype in comparison to the WT when grown aerobically in the light. Nevertheless, the hydrogenase structural genes are induced during anaerobiosis. These data indicate that the necessary transcriptional, regulatory and signaling pathways required for hydrogenase induction remain intact in this mutant. Consequently, the genes differentially expressed in this mutant relative to the WT, should be a consequence of the mutant's inability to photoproduce H_2.

Another *C. reinhardtii* mutant, *sta7-10* (Posewitz et al., 2004), is unable to accumulate intracellular starch. Interestingly, this mutant shows aberrant induction of hydrogenase-gene transcription and attenuated H_2-photoproduction activity during anaerobiosis, which correlates with the redox state of its plastoquinone (PQ) pool. One reasonable hypothesis is that the redox state of the PQ pool may signal regulatory processes responsible for turning on or off specific genes involved in anaerobic fermentative pathways and H_2 production.

Previous physiological studies have linked photosynthetic electron transport and fermentation to H_2 production in *C. reinhardtii*. However, a more general knowledge of the metabolic and regulatory context that facilitates H_2 production will be necessary to understand current limitations in H_2-production yields. We will begin to develop a global understanding of the factors that promote H_2 production during anaerobiosis by analyzing transcript profiles from WT and mutant cultures of *C. reinhardtii*. We will also investigate whether intracellular energy stores and/or redox carriers modulate this activity and/or influence the expression of genes needed for H_2 generation. This work will more fully elucidate the biochemical pathways utilized by *C. reinhardtii* during anaerobiosis and provide insights into how mutants altered in normal fermentative metabolism acclimate to anaerobiosis. Genome-wide expression data will also facilitate the modeling of carbon and reductant fluxes, guiding future molecular and metabolic engineering approaches to improve H_2 output by *C. reinhardtii*.

References

1. Melis, A., Zhang, L., Forestier, M., Ghirardi, M. L. and Seibert, M. (2000). Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. Plant Physiol **122**, 127-36.
2. Posewitz, M. C., King, P. W., Smolinski, S. L., Zhang, L., Seibert, M. and Ghirardi, M. L. (2004). Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. J Biol Chem **279**, 25711-20.
3. Posewitz, M. C., Smolinski, S. L., Kanakagiri, S., Melis, A., Seibert, M. and Ghirardi, M. L. (2004). Hydrogen Photoproduction Is Attenuated by Disruption of an Isoamylase Gene in *Chlamydomonas reinhardtii*. Plant Cell **16**, 2151-63.
4. Wykoff, D. D., Davies, J. P., Melis, A. and Grossman, A. R. (1998). The regulation of photosynthetic electron transport during nutrient deprivation in *Chlamydomonas reinhardtii*. Plant Physiol **117**, 129-39.
5. Zhang, Z., Shrager, J., Jain, M., Chang, C.W., Vallon, O. and Grossman, A. R. (2004). Insights into the survival of *Chlamydomonas reinhardtii* during sulfur starvation based on microarray analysis of gene expression. Eukaryot Cell.**3**, 1331-48.

* Presenting author

74

Exploring the Genome and Proteome of *Desulfitobacterium hafniense* DCB2 for its Protein Complexes Involved in Metal Reduction and Dechlorination

James M. Tiedje[1]*, Sang-Hoon Kim[1], Christina Harzman[1], John Davis[2], Brett Phinney[1], Michael Ngowe[1], Washington Mutatu[1], William Broderick[1], David DeWitt[1], Joan Broderick[1], and Terence L. Marsh[1] (marsht@msu.edu)

[1]Michigan State University, East Lansing, MI and [2]Columbus State University, Columbus, GA

The strictly anaerobic bacterium *Desulfitobacterium hafniense* DCB-2 grows by pyruvate fermentation and by alternate respiration using a wide range of electron acceptors including sulfur compounds, chlorinated compounds, and oxidized metals. This organism also grows with N_2 as the sole nitrogen source. The metabolic versatility makes this organism useful for studies of the mechanism of dechlorination and metal reduction, and potentially useful in bioremediation. The sequence of the genome has been determined and seven ORFs similar to the *cprA* (reductive dehalogenase or RDase) gene of *D. dehalogenans* and four ORFs similar to *nifH* of nitrogenase have been detected. To identify the RDase genes induced during alternate respiration with 3-chloro-4-hydroxybenzoate (3C4HBA), 3,5-dichlorophenol (DCP), or *ortho*-bromophenol (*o*-BP), Xeotron® microarrays of the genome of DCB-2 were prepared. Competitive hybridization of cDNA prepared from cultures grown by pyruvate fermentation and under three dehalorespiration conditions indicated that three RDase genes (designated as MENN, MFRS, and MSGV) were induced by 3C4HBA, and two genes (MSSA and VKMN) were induced by both DCP and *o*-BP. Also induced were genes within putative RDase operons, transporter/permease genes, and genes involved in electron transport systems. RT-PCR analysis targeting the seven *cprA* homologs revealed the same patterns of RDase gene expression. Northern hybridization assays targeting three RDase genes (MENN, MSSA, and VKMN) with RNAs from the four culture conditions showed a single mRNA species transcribed from MENN (1.8 kb) and VKMN (2 kb) that is long enough to contain genetic information for two linked genes, the RDase (*cprA*) and the adjacent docking protein (*cprB*) genes. However, Northern hybridization was not adequate as confirmation of the microarray results due to the cross-hybridization of probes to mRNAs from the multiple RDase genes that are phylogenetically related to varying degrees.

The physiology of heavy metal reduction under conditions where the oxidized metal was the only available electron acceptor was investigated with *D. hafniense* DCB-2 growing on a defined minimal freshwater media. The ability of DCB-2 to reduce Fe(III), Cu(II), U(VI), and Se(VI) were each tested separately. Bacterial growth under these metalorespiration conditions were observed for Fe(III), Cu(II), and U(VI), but not Se(VI). Reduction of Se(VI) did occur by DCB-2 when grown by pyruvate fermentation. SEM of DCB-2 morphology under fermentative growth with Se(VI) revealed that selenium is concentrated in polyps attached to the outside of the cell. The oxidation state of this selenium is unknown. Biofilm formation was observed for DCB-2 under conditions of fermentation (DCB-1 media) and respiration (ferric citrate media) when grown on either one of two different beads (Dupont and Siran™). Preliminary morphological examination under light microscope of planktonic cells grown in different heavy metal reduction conditions revealed considerable diversity in cellular morphologies, depending on the specific growth factors. This suggests a diverse adaptive repertoire to varying environmental conditions.

* Presenting author

75

An Integrative Approach to Energy, Carbon, and Redox Metabolism in the Cyanobacterium *Synechocystis* sp. PCC 6803

Wim Vermaas[1,*] (wim@asu.edu), Robert Roberson[1], Allison van de Meene[1], Bing Wang[1], Sawsan Hamad[1], Zhi Cai[1], Julian Whitelegge[2], Kym Faull[2], Sveta Gerdes[3], Andrei Osterman[3], and Ross Overbeek[3]

[1]Arizona State University, Tempe, AZ; [2]University of California, Los Angeles, CA; and [3]Fellowship for the Integration of Genomes (FIG), Argonne, IL

The goal of this project is to merge knowledge from genomic, bioinformatic, proteomic, metabolic, ultrastructural and other perspectives to understand how cyanobacteria live, adapt and are regulated. This project focuses on the cyanobacterium *Synechocystis* sp. PCC 6803, which is spontaneously transformable and has a known genome sequence. Cyanobacteria contribute greatly to global photosynthetic CO_2 fixation and are related to the ancestors of chloroplasts.

Electron tomography and structure determinations. Cyanobacteria have a comprehensive internal membrane system, the thylakoids, where photosynthetic reactions take place. Using electron tomography, we have determined the 3-D structure of cyanobacterial cells with about 4 nm resolution in all directions (also see http://lsweb.la.asu.edu/synechocystis/). Also, we have been successful with high-resolution freeze fracture demonstrating the layered character of the thylakoid membranes and the presence of particles on the outer, cytoplasmic and thylakoid membranes that represent integral and peripheral membrane proteins. The glycocalyx of the cell wall was also exposed. Together these images further expand our 3-D insights in the structure of the cyanobacterial cell and the biogenesis of its components.

Studies on mutants that cannot synthesize chlorophyll in darkness show that such mutants do not have thylakoids in darkness, but will develop thylakoids within hours of illumination. Before illumination, cellular structures are observed that may be thylakoid precursors. In addition, the content of the storage compound polyhydroxyalkanoate (PHA) was determined in wild type and different mutants from 70 nm sections. Initial results were compared with the biochemical data obtained from GC/MS. Particularly in mutants that cannot respire due to the lack of terminal oxidases, polyhydroxybutyrate accumulates, suggesting that this bioplastic serves as a storable fermentation product of this organism. High-resolution, 3-D structural investigations have also led to the discovery of new structures in cyanobacteria. For example, a fine, reticulate network of filaments in the cytoplasm was observed. The nature of these filaments is being determined.

Metabolic fluxes. Another important aspect that is investigated is the physiology of the organism in terms of its carbon fixation and utilization. The relative fluxes of the various central carbohydrate utilization pathways (glycolysis and the pentose phosphate pathway) are investigated by means of ^{13}C-labeling in wild type and mutants impaired in either glycolysis or the pentose phosphate pathway, and grown under various conditions. ^{13}C-glucose is rapidly taken up and converted. Upon isolation of metabolites at different times after ^{13}C-glucose addition, labeled and unlabeled metabolite products are separated by LC and analyzed by MS. The resulting patterns are then used to determine metabolic fluxes through central carbon metabolism pathways. Labeling processes occur on the timescale of 5-40 min, with different compounds showing different labeling kinetics, demonstrating the usefulness of this approach to follow the path and rate of carbon metabolism in vivo.

* Presenting author

Proteome studies. Relative expression and protein turnover studies traditionally employ complete substitution of stable isotopes. This approach has limitations, and we are investigating the use of labeling with 2-4% ^{13}C to code samples for expression proteomics and turnover measurements. Altering the ^{13}C abundance to ~2% yields a measurable effect on the peptide isotopic distribution and the inferred isotope ratio. Elevation of ^{13}C abundance to 4% leads to extension of isotopic distribution and background peaks across every unit of the mass range.

Subtle modification of the isotope ratio (~1-2% increase in ^{13}C) had no effect upon either the ability of data-dependent acquisition software or database searching software to trigger tandem mass spectrometry or match MSMS data to peptide sequences, respectively. More severe modification of the isotope ratio caused a significant drop in performance of both functionalities. Software for deconvolution of isotope ratio concomitant with protein identification using LC-MSMS has been developed (Isosolv). Subtle modification of isotope ratio proteomics (SMIRP) offers a convenient approach to *in vivo* isotope coding.

Bioinformatics. The work of the FIG research team resulted in the deployment of a new open source genomic platform, the SEED. The SEED represents a new generation of software for genome comparative analysis containing one of the largest (and permanently growing) genome collections. The complete system with annotations and tools is freely available. The SEED supports: (1) semi-automated genome comparative analysis and annotation, (2) pathway and subsystem reconstruction and analysis across multiple species, (3) community annotations and alternative assignments from major public integrations, (4) gene discovery using genome context analysis techniques, and (5) integration, comparative analysis and interpretation of functional genomics data. The current integration (http://theseed.uchicago.edu/FIG/index.cgi) contains data from 470 bacterial (of these 261 are complete or near completion), 32 archaeal (21 more or less complete), 558 (16 complete) eukaryotic, and 1272 viral genomes, including complete and nearly complete genomes of 14 cyanobacteria, as well as anoxygenic phototrophic bacteria and higher plants – invaluable for comparative genomic studies of energy and carbon metabolism in *Synechocystis* sp. PCC 6803. An important unique feature of the SEED is the support of metabolic reconstruction and comparative genome analysis via encoding and projection of functional subsystems. The FIG research team has validated the new software by developing over 150 core subsystems, covering many aspects of central metabolism. Another important SEED feature is that it is readily editable and expandable by an inexperienced user. The editing of existing subsystems and the construction of new ones is straightforward and does not require any programming skills. This provides experimental biologists with unique opportunities of fully interactive *in silico* analysis of metabolic pathways at a whole-genome scale and distinguishes the SEED from other valuable resources, such as KEGG.

The SEED platform provided the foundation for the development of CyanoSEED (to be released in January 2005), a specialized portal to comparative analysis, community-based annotation, and metabolic reconstruction of all available cyanobacterial genomes. Many new subsystems covering areas specific for cyanobacteria were added to the CyanoSEED, including: cyanobacterial photosynthetic and respiratory membrane complexes, inorganic carbon concentration and fixation mechanisms, several pigment and cofactor biosynthetic pathways, etc. A comprehensive metabolic reconstruction effort supported by the CyanoSEED provides all components required for compiling stoichiometric matrices and starting flux-balance modeling.

This combination of molecular and cell biology, genomics, proteomics and metabolome analysis leads to comprehensive insight in cyanobacterial physiology and structure, and helps to elucidate the workings of this ecologically and evolutionarily important group of organisms.

76

Role of Cellulose Binding Modules in Cellulose Hydrolysis

David B. Wilson* (dbw3@cornell.edu) and Shaolin Chen

Cornell University, Ithaca, NY

Cellulose binding modules (CBM) are found on most cellulases that catalyze the hydrolysis of crystalline cellulose, as well as on many hemicellulases produced by cellulolytic microorganisms. Removal of a CBM usually does not reduce activity on soluble substrates, such as CMC or cellulodextrins, but significantly reduces activity on crystalline cellulose (1). There is clear evidence that a major role of a CBM is to keep the catalytic domain close to its substrate, thus increasing its ability to bind to and hydrolyze individual cellulose chains (2). However, there also have been several reports that CBMs can disrupt the surface of cellulose, presumably by breaking some of the hydrogen bonds, which hold chains together (3,4). However, other workers have not seen this and thus this second role is still controversial (2, 5).

We had shown that addition of a hundred fold molar excess of a family 2 CBM to the *Thermobifida fusca* endocellulase Cel6A catalytic domain (lacking its CBM) did not give any stimulation of its activity on filter paper and thus I did not think that CBMs could disrupt cellulose (5). However, in recent experiments, we found that while the activity of Cel6Acd on filter paper is not stimulated by free CBM, the activity of native Cel6A was doubled by a one hundred-fold molar excess of a family 2 CBM. Furthermore, a twenty-fold excess of *T. fusca* E7 also doubled the activity of Cel6A on filter paper. E7 is a *T. fusca* 18K extracellular protein which is induced by growth on cellulose, is present in large amounts in the culture supernatant and that binds well to cellulose and to chitin. It has weak homology to some family 3 CBMs. These results do provide strong evidence that CBMs can alter the structure of cellulose in a way that makes it more easily hydrolyzed by an endocellulase. Experiments are under way to see if other cellulases are stimulated by E7 and if we can detect the nature of the change in cellulose structure caused by E7 binding.

References

1. Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382: 769-781 (2004).
2. Bolam, D.N., Ciruela, A., McQueen-Mason, S., et al. Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem. J. 331: 775-781 (1998).
3. Din, N., Gilkes, N.R., Tekant, B., Miller, R.C. Jr., Warren, R.A.J., and Kilburn, D.G. Non–hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. Bio/Technology 9: 1096-1099 (1991).
4. Levy, I., Shani, Z., and Shoseyov, O. Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains. Biomol. Eng. 19: 17-30 (2002).
5. Barr, B.K. Hydrolysis, specificity and active site binding of Thermobifida fusca cellulases. PhD Thesis Cornell University (1997)

 * Presenting author

77

Three *Prochlorococcus* Cyanophage Genomes: Signature Features and Ecological Interpretation

Matthew B. Sullivan[1]* (mbsulli@mit.edu), Maureen Coleman[1], Peter Weigele[1], Forest Rohwer[2], and Sallie W. Chisholm[1]

[1]Massachusetts Institute of Technology, Cambridge, MA and [2]San Diego State University, San Diego, CA

The oceanic cyanobacteria *Prochlorococcus* are globally important, ecologically diverse primary producers. Their viruses (phages) are thought to mediate population sizes and affect the evolutionary trajectories of their hosts. Here we present an analysis of genomes from three *Prochlorococcus* phages — a podovirus and two myoviruses. Although *Prochlorococcus* are evolutionarily and physiologically distinct from the majority of hosts of previously sequenced phages, the morphology, overall genome features and gene content suggest these phages are quite similar to the T7-like (P-SSP7) and T4-like (P-SSM2 and P-SSM4) phages. Using the existing phage taxonomic framework as a guideline, we examined genome sequences to establish 'core' genes for each phage group and found our cyanophages contained 15 of 26 'core' T7-like genes (P-SSP7) and 43 and 42 of 75 'core' T4-like genes (P-SSM2 and P-SSM4). Outside of these 'core' phage genes, taxonomy of best hits analyses suggest each genome contains a significant number of 'cyanobacterial' genes – i.e., genes which are common in cyanobacteria, but have not been observed among phages outside the cyanophages – some of which we speculate represent 'core' cyanophage genes. For example, all three phage genomes contain photosynthetic genes (*psbA, hliP*) that are thought to help maintain host photosynthetic activity during infection, as well as an aldolase family gene (*talC*) that suggests alternative routes of carbon metabolism are important during cyanophage infection. The podovirus P-SSP7 genome also contains an integrase gene (*int*) and genome features that suggest it is capable of integrating into its host. If functional, this would be the first report of a cultured temperate T7-like phage or a temperate marine cyanophage, and would have significant evolutionary and ecological implications for both phage and host. Further, both myoviruses P-SSM2 and P-SSM4 contain phosphate-inducible genes (*phoH* and *pstS*) that may be important for phage and host responses to phosphate stress, a commonly limiting nutrient for marine cyanobacterial growth. Thus, these marine cyanophages appear to be variations of two well-known phages, but contain genes that, if functional, may document how cyanophages have specialized for infection of photosynthetic hosts in low nutrient oceanic environments.

* Presenting author

78

The Alternative Sigma Factor RpoN Regulon of *Rhodopseudomonas palustris*

Yasuhiro Oda[1]* (yasuhiro-oda@uiowa.edu), Sudip K. Samanta[1], Frank W. Larimer[2], and Caroline S. Harwood[1]

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

The alternative RNA polymerase sigma factor RpoN activates a wide range of genes involved in various cellular processes in many different bacterial species. These include nitrogen and carbon metabolism, flagellar biosynthesis, and virulence. The photosynthetic bacterium *Rhodopseudomonas palustris* is an extremely metabolically versatile species. Under anaerobic conditions it can generate energy from light and convert nitrogen gas to ammonia and hydrogen (a biofuel) by nitrogen fixation. It can also degrade lignin monomers. This metabolic versatility is reflected in the genome sequence of *R. palustris* strain CGA009. To address the question of what physiological processes are controlled by RpoN in *R. palustris*, we isolated an *rpoN* Tn5 transposon mutant. The mutant grew in medium containing ammonium as a nitrogen source with the same growth rate as wild-type, but did not grow under nitrogen-fixing conditions. In addition, *rpoN* appears to be involved in motility, hydrogen recycling, and biofilm formation. Introduction of a plasmid containing the *rpoN* gene into the mutant *in trans* complemented all of these phenotypes. To assess the RpoN regulon of *R. palustris* in more detail with an emphasis on nitrogen metabolism, the whole genome gene expression profile of wild-type cells was compared to that of the *rpoN* mutant. The two strains were grown in medium containing yeast extract as the nitrogen source, which derepresses nitrogenases in wild-type cells and allows growth of the *rpoN* mutant. Wild-type cells expressed over 400 genes at levels of 2-fold or higher as compared to the *rpoN* mutant. Among these were the genes involved in nitrogen (e.g., nitrogenases, nitrogen regulatory proteins, ammonium transporters, and glutamine synthetases) and carbon (e.g., lignin monomers, fatty acids, and dicarboxylic acids) metabolism, flagellar biosynthesis, and various transport systems. These results and a computational analysis of the RpoN regulon have elucidated the functional role of the alternative sigma factor RpoN in the successful metabolic opportunist *R. palustris*.

* Presenting author

79

Integrative Control of Key Metabolic Processes in *Rhodopseudomonas palustris* for the Enhancement of Carbon Sequestration and Biohydrogen Production

F. Robert Tabita[1]* (Tabita.1@osu.edu), Janet L. Gibson[1], Caroline S. Harwood[2], Frank Larimer[3], J. Thomas Beatty[4], James C. Liao[5], and Jizhong (Joe) Zhou[3]

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

The nonsulfur purple (NSP) photosynthetic (PS) bacteria (1) are the most metabolically versatile organisms found on Earth and they have become model organisms to understand the biology of a number of important life processes. One bacterium, *Rhodopseudomonas palustris*, is unique in that it is able to catalyze more processes in a single cell than any other member of this versatile group. Thus, this organism probably catalyzes more fundamentally and environmentally significant metabolic processes than any known living organism on this planet. *R. palustris* is a common soil and water bacterium that can make its living by converting sunlight to cellular energy and by absorbing atmospheric carbon dioxide and converting it to biomass. It is often the most abundant NSP PS bacterium isolated in enrichments. Its abundance is most probably related to one of its unique characteristics; i.e., unlike other NSP PS bacteria, *R. palustris* can degrade and recycle components of the woody tissues of plants (wood contains the most abundant polymers on earth). *R. palustris* can do this both aerobically in the dark and anaerobically in the light. Recent work has shown that regulation of the processes of CO_2 fixation, N_2 fixation, and H_2 metabolism is linked in NSP bacteria (2). Moreover, a different, yet uncharacterized regulatory mechanism operates under aerobic conditions (unpublished results). Now that its genome sequence is available through the efforts of the JGI and the members of this consortium (3), interactive metabolic regulation of the basic CO_2, hydrogen, nitrogen, aromatic acid, and sulfur pathways of *R. palustris*, as well as other important processes, can be probed at a level of sophistication that was not possible prior to the completion of the genomic sequence. We have pooled the collective expertise of several investigators, using a global approach to ascertain how all these processes are regulated in the cell at any one time. These studies take advantage of the fact that *R. palustris* is phototrophic, can fix nitrogen and evolve copious quantities of hydrogen gas, and is unique in its ability to use such a diversity of substrates for both autotrophic CO_2 fixation (i.e., H_2, H_2S, $S_2O_3^{2-}$, formate) and heterotrophic carbon metabolism (i.e., sugars, dicarboxylic acids, and aromatics, plus many others) under both aerobic and anaerobic conditions.

With regard to the integrative control of metabolism, we have shown that the control of CO_2 fixation is 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 H_2 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 in bio-reactor systems, the advent of these newly isolated strains, in which hydrogen production is not subject to the normal control mechanisms that diminish the wild-type stain, 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 hydrogen gas, by combining the properties of the hydrogen- producing derepressed strains, with waste organic carbon degradation, there is much potential to apply these basic molecular manipulations to practical advances. To maximize this capability, the coordinated application of gene expression profiling (transcriptomics), proteomics, carbon flux analysis and bioinformatics approaches have been combined with traditional studies of mutants and physiological/biochemical characterization of cells. During the course of these studies, novel genes and regulators were identified from investigating control of specific processes by conventional molecular biology/biochemical techniques. These studies, along with the microarray studies discussed above, have shown that there are key protein regulators that control many different processes in this organism. In many instances, further surprises relative to the role of known regulators, such as the Reg system and CbbR, were noted in *R. palustris*. A novel phospho-relay system for controlling CO_2 fixation gene expression was also identified and biochemically characterized and the means by which this system influences other aspects of metabolism is also under study. This latter system, where key regulators contain motifs that potentially respond to diverse metabolic and environmental perturbations, suggests an exquisite means for controlling key processes such as CO_2 fixation. Likewise, interesting and important genes and proteins that control sulfur oxidation, nitrogen fixation, hydrogen oxidation, and photochemical energy generation have been identified and characterized, and the biochemistry of these systems is under intense study.

In summary, functional analysis of the *R. palustris* proteome and transcriptome, along with traditional biochemical/physiological characterization, has led to considerable progress, placing our group in excellent position to address long term goals of computational modeling of metabolism such that carbon sequestration and hydrogen evolution might be maximized.

References

1. Tabita, F. R., and Hanson, T. E. Anoxygenic photosynthetic bacteria. 2004. In: Microbial Genomics. C. M. Fraser, K. E. Nelson, and T. D. Read (eds.). Humana Press, Inc., Totowa, NJ, pp. 225-243.

2. Dubbs, J. M., and Tabita, F. R. Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO_2 assimilation, nitrogen fixation, hydrogen metabolism and energy generation. FEMS Microbiol. Rev. **28** (2004) 353-376.

3. Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F. R., Gibson, J.L., Hanson, T. E., Bobst, C., Torres y Torres, J., Peres, C., Harrison, F.H., Gibson, J., and Harwood, C.S. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nature Biotechnology **22** (2004) 55-61.

 * Presenting author

80

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

Gary L. Andersen*[1] (GLAndersen@lbl.gov), Ping Hu[1], Eoin L. Brodie[1], and Harley H. McAdams[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA and [2]Stanford University School of Medicine, Stanford, CA

Potentially hazardous levels of heavy metals have dispersed into subsurface sediment and groundwater in a number of metal contaminated DOE sites and represent a challenge for environmental restoration. Effective bioremediation of these sites requires knowledge of genetic pathways for resistance and biotransformation by component organisms within a microbial community. The aquatic bacterium *Caulobacter crescentus* is a ubiquitous organism with a distinctive ability to survive in low nutrient environments. It has been selected for extensive study by DOE because of its ability to survive in broad environmental habitats where contamination may be present. The recently completed sequence of the strain CB15 has provided the information to study genome wide response to heavy metal stress. A customized 500,000-probe Affymetrix array has been designed by the McAdams laboratory at Stanford University to measure transcription levels of all 3763 putative ORFs, both strands of hypothetical proteins as well as the intervening intergenic regions. This study used this microarray to study transcriptional response to heavy metal stress.

We studied the toxic effect of six heavy metals (seven compound: methylmercury chloride, cadmium sulfate, sodium selenite, lead nitrate, potassium chromate, potassium dichromate and uranyl nitrate) on growth, survival and cell morphology. We unexpectedly found that strain CB15N was not significantly affected for growth at 1 mM uranium concentration. The highest level of uranium currently observed in ground water at the Oakridge FRC is 200 μM. Under the same conditions in our laboratory, growth of *E. coli* K-12 was completely stopped and the growth of *Pseudomonas putida* (*Pseudomonas spp*. has been reported to accumulate uranium) was drastically reduced. We believe this is the first study to identify *C. crescentus* as a uranium-resistant bacterium. Whole genome transcriptional analysis using the Affymetrix *C. crescentus* microarray revealed groups of genes, operons and pathways, which were up regulated under different heavy metal stresses. Some of the up-regulated pathways (such as DNA repair, removal of superoxide radicals, thio-group protection) confirmed what is known about heavy metal stress on other organisms. Nine transcripts were commonly up-regulated when the cells were stressed with four different toxic metals. We also observed the up-regulation of specific regulatory genes as well as genes and operons of unknown function in response to specific metal stresses. In cells stressed with uranium we observed the up-regulation of four proteins that belong to two different two-component signal transduction systems. Their involvement in uranium stress was confirmed in phenotypic studies by deletion mutants of one signaling pathway. We also identified groups of genes and operons of unknown functions, including transcripts from antisense strand of a predicted gene. Further studies may elucidate function of these transcripts and, ultimately, the mechanism used by *C. crescentus* to overcome uranium toxicity. Whole genome transcriptional analysis provided a powerful tool for the detection of candidate genes, with no prior knowledge, that may be involved in metal stress survival. Such analysis will be increasingly necessary as more microbial genome sequences are completed with only computational annotation to suggest function.

81

Systematic Analysis of Two-Component Signal Transduction Systems Regulating Cell Cycle Progression in *Caulobacter crescentus*

Michael Laub* (Laub@CGR.Harvard.edu)

Harvard University, Cambridge, MA

Progression through the cell cycle requires the precise coordination of DNA replication, chromosome segregation, cell division, and cell growth. How these processes are coordinated and regulated can be studied in the experimentally tractable model bacterium Caulobacter crescentus. Cell cycle progression in Caulobacter is accompanied by a series of morphological transitions which culminate in the production of two asymmetric daughter cells, a sessile stalk cell that immediately initiates a new round of DNA replication after cell division and a motile swarmer cell that must differentiate into a stalked cell before initiating DNA replication. We have begun a systematic analysis of the signaling and regulatory genes controlling the Caulobacter cell cycle, focusing primarily on the two-component signal transduction systems, comprised of histidine kinases and response regulators. Two-component signaling systems are ubiquitous regulatory pathways in prokaryotes that provide a versatile means of detecting and responding to changes in environmental, cellular, and developmental conditions.

We systematically generated deletion strains for each of the 107 two-component signaling genes (63 histidine kinases and 44 response regulators) encoded in the Caulobacter genome. The systematic phenotypic characterization of these mutants has identified four new two-component genes essential for viability and 16 others required for proper cell cycle progression. The deletion mutants generated were individually bar-coded to enable high-throughput analysis of growth and fitness under a variety of environmental conditions; use of this assay to identify the stimuli for specific two-component systems will be presented. Finally, we have developed a technique, termed kinase-substrate profiling, which allows the rapid and accurate delineation of phosphate flow through two-component signaling pathways. This technique has been employed to identify kinase-regulator pairs among the newly identified cell cycle regulatory proteins. Examples will be presented to demonstrate how the combination of these systematic genetic, biochemical, and genomic approaches can quickly lead to the identification of signaling pathways controlling key cellular processes and metabolic changes.

 * Presenting author

82

The U.S. DOE Joint Genome Institute Microbial Program

David Bruce[1]* (dbruce@lanl.gov), Alla Lapidus[2], Patrick Chain[3], Jeremy Schmutz[4], Frank Larimer[5], Nikos Kyrpides[2], Paul Gilna[1], Eddy Rubin[2] and Paul Richardson[2]

[1]JGI-Los Alamos National Laboratory, Los Alamos, NM; [2]JGI-Production Genomics Facility and Lawrence Berkeley National Laboratory, Berkeley, CA; [3]JGI-Lawrence Livermore National Laboratory, Livermore, CA; [4]JGI-Stanford Human Genome Center, Palo Alto, CA; and [5]JGI-Oak Ridge National Laboratory, Oak Ridge, TN

The Department of Energy initiated the Microbial Genome Program (MGP, http://microbialgenome.org) in late 1994 as a spin-off of the Human Genome Program. The principle goal of the MGP is to fund research into microbes related to DOE interests. The DOE Joint Genome Institute (JGI, www.jgi.doe.gov) is composed of affiliates from a number of national laboratories including Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, as well as the Stanford Human Genome Center. The JGI's Production Genomics Facility in Walnut Creek, California is a high throughput sequencing center and the principal engine for the JGI. Originally designed to sequence the DOE commitment to the Human Genome Program (Chromosomes 5, 16, and 19), this facility now routinely produces over 2.5 billion raw bases per month from a wide variety of organisms including microbes and microbial communities. The JGI is responsible for sequencing, assembling, and annotating microbial genomes of interest to the DOE through the MGP and GTL programs.

The JGI Microbial Program was recently established to better coordinate and leverage the capabilities of its partner organizations. A work flow procedure has been formalized to process samples from DNA prep through sequencing, assembly, finishing, quality assurance, annotation and analysis. To date, the JGI has sequenced over 100 microbes to draft quality, finished over 30 and is currently working on more than 60 additional microbial projects. Most projects are now targeted for finishing at one of three JGI locations.

Virtually all microbial projects are sequenced by the whole genome shotgun method. The JGI randomly shears the purified DNA under different conditions and selects for three size populations. Fragments are end repaired and selected for inserts in the range of 3kb, 8kb, and 40kb. These are cloned into different vector systems and checked for quality by pcr or sequencing. Once libraries have passed the QC step, a total coverage of approximately 8.5X sequencing is performed on colonies from the three libraries. The resulting reads are trimmed for vector sequences and assembled. This production sequencing assembly is quality checked and then released to the collaborating PI as the initial Quality Draft assembly and is automatically annotated by the group at Oak Ridge National Laboratory. At this point, the draft assembly is then assigned to a finishing group that will close all gaps, resolve repeat discrepancies, and improve low quality regions. The final assembly is then passed to the Stanford Quality Assurance group to assess the integrity and overall quality of the genome sequence. The finished sequence then receives a final annotation and this package is used as the basis for analysis and publication.

* Presenting author

83

Identification of Genes that are Required for Recycling Reducing Power during Photosynthetic Growth

Christine L. Tavano, Angela M. Podevels, and Timothy J. Donohue* (tdonohue@bact.wisc.edu)

University of Wisconsin, Madison, WI

Photosynthetic organisms have the unique ability to conserve the energy in light and generate reducing power. We are studying what is required for photosynthesis in the α-proteobacterium, *Rhodobacter sphaeroides*. Global gene expression analysis has shown that RNA levels from ~50 genes of unknown function were regulated by changes in light intensity and oxygen tension, like those of known components of the *R. sphaeroides* photosynthetic apparatus. Several of these uncharacterized genes were located in the RSP4157-4164 gene cluster on plasmid P004. A mutant containing a polar insertion in RSP4157, CT01, was able to grow via photosynthesis under autotrophic conditions using H_2 as an electron donor and CO_2 as a carbon source. However, CT01 was unable to grow photoheterotrophically in a succinate based medium unless compounds that can be used to recycle reducing power (the external electron acceptor DMSO or CO_2) were present, suggesting that this mutant was defective in recycling reducing power during photosynthetic growth when a fixed carbon source was present. CT01 had decreased levels of RNA for genes encoding putative glycolate degradation functions. Exogenous glycolate also rescued photoheterotrophic growth of CT01, leading us to propose that CO_2 produced from glycolate metabolism can be used by the Calvin cycle to recycle reducing power generated in the photosynthetic apparatus. The ability of glycolate, CO_2, or DMSO to support photoheterotrophic growth of CT01 suggests that products of the RSP4157 gene cluster serve a heretofore unknown role in recycling reducing power under photosynthetic conditions..

84

A Tightly-Regulated Oscillatory Circuit Formed by Conserved Master Regulator Proteins Controls the *Caulobacter* Cell Cycle

Harley McAdams* (hmcadams@stanford.edu) and Lucy Shapiro

Stanford University School of Medicine, Stanford, CA

A complex oscillatory genetic circuit controls *Caulobacter crescentus* cell-cycle progression and asymmetric polar morphogenesis. Two tightly regulated master regulatory proteins, CtrA and GcrA, were recently shown to form the core oscillator.[1] Two switches triggered by the progression of chromosome replication and cytoplasmic compartmentalization, respectively, act to create a type of escapement mechanism that paces and synchronizes progression of the cell cycle.[1,2] The intracellular concentrations of GcrA and CtrA exhibit both temporal and spatial oscillations that act to activate or repress numerous cell cycle regulated genes. Many of these genes are themselves top-level regulators of modular functions that execute the functions involved in cell cycle progression (for example, replicating the chromosome, initiating the FtsZ ring, or constructing polar structures)

 * Presenting author

The architecture of the bacterial cell's regulatory control system is a hierarchical, modular, asynchronous, self-timed control system.[3] In this system, synchronization of the ordering of cell cycle events is dependent on cross module dependencies, that is, checkpoints, often involving signaling via two-component systems (see Michael Laub's abstract). Recent results elaborating this circuit include characterization of the regulons of two additional key *Caulobacter* cell cycle regulatory proteins, DnaA and DivK (publications in preparation). We have also identified and characterized two long-sought proteins centrally involved in active regulation of CtrA proteolysis (publication in preparation).

Caulobacter divides asymmetrically, producing daughter cells with differing polar structures, different cell fates, and asymmetric regulation of the initiation of chromosome replication. Complex intracellular signaling is required to keep the organelle developmental processes at the cell poles synchronized with other cell cycle events. Two recently characterized switch mechanisms controlling cell cycle progress are triggered by relatively large scale developmental events in the cell: the progress of the DNA replication fork and the physical compartmentalization of the cell that occurs well before division. These mechanisms invoke rapid, precisely timed, and even spatially differentiated regulatory responses at important points in the cell cycle. Assays of the relative timing of cytoplasmic compartmentalization and cell division coupled with high resolution 3D tomographic images of the end stages of cell division has shown how breaking of phosphor-signaling paths by blocking the cytoplasmic diffusion of phosphorylated signal proteins leads to differential regulatory programming of the daughter cells.[2,4]

1. Holtzendorff, J., Hung, D., Brende, P., Reisenauer, A., Viollier, P. H., McAdams, H. H., Shapiro, L. Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304**, 983-7 (2004).
2. McGrath, P.T., Viollier, P. & McAdams, H.H. Setting the Pace: Mechanisms tying *Caulobacter* cell-cycle progression to macroscopic cellular events. *Current Opinion in Microbiology*, **7(2)**:192-7. (2004).
3. McAdams, H.H. & Shapiro, L. A bacterial cell-cycle regulatory network operating in time and space. *Science* **301**, 1874-7 (2003).
4. Judd, E.M., Ryan, K., Moerner, W.E., Shapiro, L. & McAdams, H.H. Fluorescence bleaching reveals asymmetric compartment formation prior to cell division in *Caulobacter*. *Proc. Natl. Acad. Aci. USA* **100**, 8235-8240 (2003).

85

Dynamics and Control of Biofilms of the Oligotrophic Bacterium *Caulobacter crescentus*

Alfred M. Spormann (spormann@stanford.edu) and Plamena Entcheva-Dimitrov

Stanford University, Stanford, CA

Caulobacter crescentus is an oligotrophic α-proteobacterium with a complex cell cycle involving sessile stalked and piliated, flagellated swarmer cells. Because the natural lifestyle of *C. crescentus* intrinsically involves a surface-associated, sessile state, we investigated the dynamics and control of *C. crescentus* biofilms developing on glass surfaces in a hydrodynamic system. In contrast to biofilms of the well studied *Pseudomonas aeruginosa*, *E. coli*, and *Vibrio cholerae*, *C. crescentus* CB15 cells form biphasic biofilms, consisting predominantly of a cell monolayer biofilm and a biofilm containing densely-packed, mushroom-shaped structures. Based on comparisons between the *C. crescentus* strain CB15 wild type and its holdfast (*hfsA*, ΔCC0095), pili (Δ*pilA-cpaF*::Ω*aac3*), motility (*motA*), flagellum (*flgH*) mutants and a double mutant lacking holdfast and flagellum (*hfsA*; *flgH*), a model for biofilm formation in *C. crescentus* is proposed: For both biofilm forms, the holdfast structure at the tip of a stalked cell is crucial for mediating the initial attachment. Swimming motility by means of the single polar flagellum enhances initial attachment and enables progeny swarmer cells to escape from the monolayer biofilm. The flagellum structure contributes also to maintaining the mushroom structure. Type IV pili enhance but are not absolutely required for the initial adhesion phase. However, pili are essential for forming and maintaining the well-defined three-dimensional mushroom-shaped biofilm. The involvement of pili in mushroom architecture is a novel function for type IV pili in *C. crescentus*. These unique biofilm features demonstrate a spatial diversification of the *C. crescentus* population into a sessile, as 'stem cells' acting subpopulation (monolayer biofilm), which generates progeny cells capable of exploring the aqueous, oligotrophic environment by swimming motility, and a subpopulation accumulating in large mushroom structures.

86

Widespread and Abundant CelM Endoglucanases of Marine *Cytophaga*-like Bacteria Revealed by Whole Genome Shotgun Sequencing and Fosmid Cloning

Matthew T. Cottrell and David L. Kirchman* (kirchman@cms.udel.edu)

University of Delaware, Lewes, DE

Culture-independent approaches for investigating the composition of marine bacterial communities have revealed that the most abundant members of natural bacterial communities differ from those that have been isolated in culture. Data obtained with a variety of approaches, including fluorescence in situ hybridization (FISH) and fosmid cloning of environmental DNA, have identified *Cytophaga*-like as an abundant constituent of marine bacterial consortia. In our work we have been focusing the role of *Cytophaga*-like bacteria in carbon cycling in the ocean and their adaptations for

* Presenting author

the consumption of dissolved organic material (DOM) in the form of high molecular weight polysaccharides. Previous investigations revealed that in natural marine consortia *Cytophaga*-like bacteria are superior competitors for the consumption of radiolabeled protein, chitin and possibly other polysaccharides. In this study we sought to identify what types of endoglucanases are employed by marine *Cytophaga*-like bacteria to utilize high molecular weight polysaccharides. Our approach was to search contigs bearing *Cytophaga*-like 16S rDNA in the recently published whole genome shotgun sequence database of the Sargasso Sea for endoglucanase genes.

We examined 27 contigs bearing *Cytophaga*-like 16S rDNA, including 11 annotated as such, plus 16 additional contigs we identified as having *Cytophaga*-like 16S rDNA by BLAST analysis of a 16S rDNA data base. BLAST analysis of the *Cytophaga*-like sequences revealed two contigs with open reading frames encoding proteins that are most similar to CelM in the *Cytophaga hutchinsonii* genome sequence. Subsequent BLAST analysis of the 1,001,987 conceptual peptides against a database of 113 cellulases (EC 3.2.1.4) obtained from Swiss-Prot revealed 30 contigs encoding CelM-like cellulases, which was the most prevalent cellulase detected. Cellulases belonging to glycosyl hydrolase family 5 were detected on 27 contigs, while 10 or fewer contigs possessed cellulases similar to those in glycosyl hydrolase families 8, 9, 10 and 12.

PCR primers that amplify a 1,132 bp fragment from selected environmental CelM sequences were used to screen a fosmid library of bacterial DNA from the Arctic Ocean. The CelM-positive fosmid bears *Cytophaga*-like 16S rDNA that places it in a separate cluster from the Sargasso Sea *Cytophaga*-like bacteria with CelM. Phylogenetic analysis of CelM genes from cultivated and uncultivated microbes revealed that the marine environmental and the *Cytophaga hutchinsonii* genes cluster together and are distinct from the CelM genes identified in a wide variety of microbes, including Archaea and Gram-positive bacteria. Sequence analysis of the Arctic CelM revealed the expected conserved domains, including COG1363 and Pfam 05343, and evidence of a signal peptide.

The physiological role of CelM and related proteins is uncertain because there is experimental evidence for genes encoding similar proteins with endoglucanase activity in *Clostridium thermocellum* and peptidase activity in *Lactococcus lactis*. In order to test the hypothesis that CelM of *Cytophaga*-like bacteria plays a role in the consumption of high molecular weight DOM by *Cytophaga*-like bacteria, we subcloned the Arctic *Cytophaga*-like CelM gene into an expression vector for endoglucanase and peptidase activity screening. We expect that the Arctic CelM will have endoglucanase activity because it is more similar to CelM in *C. thermocellum* gene than in *L. lactis*. In addition, the Arctic *Cytophaga*-like CelM is highly similarity to the CelM gene in *C. hutchinsonii*, which is an aggressive cellulose degrader. The activity of Arctic CelM is now being characterized and the complete fosmid is being sequenced by DOE-Joint Genome Institute.

87

Data Analysis and Protein Identification Strategy for the Systems-Level Protein-Protein Interaction Networks of *Shewanella oneidensis MR-1*

Gordon A. Anderson*[1] (gordon@pnl.gov), James E. Bruce[2], Xiaoting Tang[2], Gerhard Munske[2], and Nikola Tolic[1]

[1]Pacific Northwest National Laboratory, Richland, WA and [2]Washington State University, Pullman, WA

The Protein-Protein Interaction Networks research program aims to identify proteome wide protein interaction using cross-linking and high performance mass spectrometry. Cross-linking coupled with mass spectrometry is a widely used technique in protein interaction research. This technique can present a number of informatics challenges. Techniques that involve cross-linking of the proteins and then enzymatic digestion of these proteins into complex mixtures of peptides and cross linked peptides present a significant analysis challenge. The number of possible cross-linked peptides is the square of the number of tryptic peptides in the organism under study. This problem is amplified by other factors such as incomplete digestion and posttranslational modification.

This research project has developed unique cross-linker molecules called Protein Interaction Reporters (PIRs) that contain bonds that can be cleaved with high specificity in the mass spectrometer. This allows the detection of the cross-linked peptide mass and then, after low energy CID, detection of the individual peptide masses. Using this mass information and the amino acid specificity of the cross-linker molecule peptide identification is possible in many cases without further mass spectrometry. An additional feature of this technique is the spacer chain of the cross-linker that is detected in the low energy CID spectrum. This provides a mass "signature" in the CID spectrum indicating cross-linked peptide data is present. This presentation will outline the data analysis steps required and outline an identification strategy.

High performance mass spectrometry will be used to enable high throughput identification of cross linked proteins. The cross-linked proteins will be extracted from the organism and purified. The proteins will be digested and prepared for analysis by mass spectrometry. High performance FTICR mass spectrometry will be used because of its high mass measurement accuracy. FTICR allows peptide mass measurements of 1 to 5 ppm. This mass measurement accuracy coupled with the peptide constraints due to this PIR approach will enable peptide and protein identification based on mass measurement accuracy alone.

The analysis of cross linked peptides consists of a mass spectrometry experiment that involves capturing a pre-cursor ms scan containing potential cross linked peptides. The next ms spectrum is a low energy CID of the same ions detected in the first pre-cursor spectrum. The second spectrum contains the cross-linked peptides with the cross-linker fragmented. This fragmentation occurs without affecting the peptide backbone. This pair of spectra (the pre-cursor and low energy CID) allows identification of the peptide masses that were cross linked. The low energy CID spectrum will also contain a "signature" indicating a cross linked peptide pair was in the pre-cursor spectrum. This "signature" is the mass of the cross linker's core or spacer mass. This "signature" identifies spectra that potentially will identify a cross linked peptide pair. The two spectra can then be analyzed to identify the two peptide masses. The mass of the cross-linker core plus the two cross linked peptides must

* Presenting author

equal the mass of a ion detected in the pre-cursor spectrum. This analysis step will insure peptide identification will only be attempted for masses resulting from cross linked peptides.

The next step in the process is the identification of these peptides from their accurate masses. This identification is assisted by peptide constraints imposed by the cross-linker molecule. This presentation will show the mass measurement accuracy needed for identification with and without the peptide constrains this methodology provides. This analysis is performed using the genomic data from *Shewanella oneidensis*. This data will be used to calculate all of the possible cross linkable peptides. These peptide sequences will be used to calculate the peptide masses. The resultant masses will be used to predict there uniqueness at various levels of mass measurement accuracy to determine the feasibility of identifying the peptides using mass accuracy alone.

This presentation, through simulation, demonstrates the analysis algorithms that will be used to identify cross-linked peptides and illustrate the mass spectrometry performance necessary for high throughput performance.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

88

A Protein Interaction Reporter Strategy for Systems-Level Protein Interaction Networks of *Shewanella oneidensis MR-1*

James E. Bruce[1]* (james_bruce@wsu.edu), Xiaoting Tang[1], Harry Zhu[1], Saiful Chowdhury[1], Devi Adhikari[1], Gerhard Munske[1], Gordon A. Anderson[2], and Nikola Tolic[2]

[1]Washington State University, Pullman, WA and [2]Pacific Northwest National Laboratory, Richland, WA

A key challenge inherent in the utilization of genomic data is related to deciphering the network of protein-protein interactions that enable biological function. Chemical cross-linking has gained increasing interest as a tool for protein interaction profiling, yet the examples of successful application are relatively scarce. The difficulty inherent in mass spectral interpretation and protein identification resultant from cross-linking reaction mixtures is a significant barrier that has hindered many such efforts. The analysis of complex MS and MS/MS patterns resultant from various types of cross-linker products and multiple fragmentation pathways can present levels of complexity that preclude protein and protein-protein interaction identification. However, a general technique that can identify proteins based on a physical property common to protein-protein interactions, namely, the proximity of multiple protein species within a complex mixture, is still very desirable.

Our efforts have been devoted to the development of a novel approach for chemical cross-linking that can enable improved identification of protein interactions in complex systems. A key component of this research is the development of new compounds that can provide advanced features and additional information from cross-linking reaction mixtures. We call our approach that employs mass spectrometry-cleavable cross-linkers a "Protein Interaction Reporter" (PIR) strategy, since the fragments of the cross-linker themselves are encoded with additional information that enables improved analytical capabilities for protein interaction profiling. For example, our first-generation PIR structures have been developed with low-energy CAD cleavable bonds that, when activated, release a

* Presenting author

reporter ion of specific *m/z*. These bonds can be efficiently fragmented at energy levels that preclude fragmentation of nearly all peptide amide backbone bonds. Thus, our initial MS/MS analyses of PIR-labeled products are less congested by complex multiple fragmentation pathways that are commonly observed in most cross-linked peptide MS/MS spectra. Next-generation PIR structures employ additional fragmentation schemes and features to allow even more information to be encoded in the compound. In all cases, the measured *m/z*'s of fragment ions resultant from PIR-peptide complex activation provide information that enables improved cross-link type and cross-linked peptide identification.

This presentation will highlight our initial proof-of-principle PIR experiments that were performed with model noncovalent complexes. These applications of PIR technology showed that MS/MS data could be used to differentiate various product types from cross-linking reactions, and help pinpoint ions that are resultant from protein interactions. Since the complexity of products normally poses a significant impediment to successful cross-linker application even for model noncovalent complexes, PIR advancements allow improved capabilities for analysis of protein interactions with mass spectrometry. This approach was able to produce protein-protein interaction structural data in excellent agreement with the known X-Ray crystal structure of Ribonuclease S, our model noncovalent complex. In addition, second-generation PIR structures were synthesized to incorporate affinity capture capabilities to allow enrichment of cross-linking products from complex mixtures. This feature will allow PIR cross-linked products to be enriched from digestion mixtures of proteins. Our initial investigations with this compound demonstrated that the sites of incorporation of a biotinylated PIR were nearly identical to those of the first generation structure. Thus, the affinity label seems to pose no major limitation to the reactivity of the PIR. We have also begun application of this PIR structure to complex protein mixtures, including *Shewanella oneidensis MR-1* cell lysates in an effort to better define conditions for cross-linking studies with our compounds. The results of these initial applications, additional compound features, and our envisioned implementation of the PIR strategy with tandem accurate mass analyses for the characterization of protein interactions in *Shewanella oneidensis MR-1* will be presented.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

 * Presenting author