Section 3

Regulatory Processes

110

VIMSS Computational Core: Comparative Analysis of Regulatory Systems in Environmental Microbes

Eric J. Alm[1,2], Inna Dubchak[1,2], Alex Gilman[1,2], Katherine Huang[1,2], Keith Keller[1,2], and **Adam P. Arkin**[1,2,3*] (APArkin@lbl.gov)

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

Background. The VIMSS Computational Core group is tasked with data management, statistical data analysis, modeling, and comparative and evolutionary genomics for the larger VIMSS effort. We have matured many of our analyses and VIMSS data into our flagship comparative functional microbial genomics tool MicrobesOnline (http://microbesonline.org). We have used this framework to interpret the data from the VIMSS physiological pipeline and in more global analysis of genome evolution and function.

Data Analysis. During the course of analysis of the various stress responses of DvH the computational core has developed a number of new statistical analyses of data that take advantage of the predicted regulatory structures (operons, regulons, etc.) from our comparative analyses. We have used these analyses this year to uncover unique responses to salt and heat shock, to exposure to heavy metals, and to elucidate how reducing power is moved among different pathways under nitrogen oxide stress. Both the new methods and the individual analyses have been published recently. We will report on the key finding of these studies.

Data Management. The Experimental Data Repository (http://vimss.lbl.gov/~jsjacobsen/cgi-bin/GTL/VIMSS/datarepository.cgi) has continued to grow in size and functionality and provides access to biomass production data, other growth curve data, synchrotron FTIR data, image data, phenotype microarray data, and transcriptome, proteome and metabolome data. Many of these data types now are supported by sophisticated analyses and visualizations that provide feedback and value to the project personnel and the wider public. This year we will be improving site navigation, project management tools and the range of analyses and visualizations available. It will also be expanded to accept data from collaborating projects in environmental sequences, plant/microbe mesocosms, and protein complex data. Further, a more automated transfer of data from the EDR to MicrobesOnline will be deployed. Many of the data management tools developed as part of this project have be translated to our collaborating projects both related to Genomics:GTL and not.

The MicrobesOnline Database. The MicrobesOnline database (http://microbesonline.org) currently hosts 243 genomes and features a full suite of software tools for browsing and comparing microbial genomes. Highlights include operon and regulon predictions, a multi-species genome browser, a multi-species Gene Ontology browser, a comparative KEGG metabolic pathway viewer and the VIMSS Bioinformatics Workbench for more in-depth sequence analysis. The Workbench provides gene carts that store user defined sets of genes found by searching MicrobesOnline and

* Presenting author

provides tools for multiple sequence alignments, phylogenetic trees construction, and, in prototype, cis-regulatory site detection. We are also incorporating and updating Mikhail Gelfand's highly curated RegTransDb of experimentally verified regulator site/transcription factor data for display on the gene pages of MicrobesOnline and to aid the Workbench cis-regulatory tools.

In addition, we provide an interface for genome annotation, which like all of the tools reported here, is freely available to the scientific community. To keep up with the ever-increasing rate at which microbial genomes are being sequenced, we have established an automated genome import pipeline. A number of outside groups are currently using the MicrobesOnline database for genome annotation projects. To facilitate the use of this community resource we have developed an access control system, so individual research groups can use the power of the VIMSS annotation tools, while keeping data from their own particular genome project private until their analyses are ready to be made public.

Also incorporated in this framework is our microbial microarray analysis suite including a number of quality control metrics, COG/TIGRFAM functional enrichment analysis, operon co-expression analysis and statistical significance tests developed based on this information. The microarray database currently holds data from nine bacteria and twenty-one conditions and it is growing rapidly. The Workbench will be expanded to allow microarray and comparative analysis to be combined. Functional genomics data of all sorts will be directly linked and visualized on the gene pages in MicrobesOnline for which such data exists. As data from our collaborating projects on protein complexes become available this data, too, will be served from this site. We welcome depositing of other peoples' data on MicrobesOnline which in turn we hope will allow users to better analyze their information.

Evolution of Microbial Genomes. We have also used the MicrobesOnline framework to support research on the evolution of specific pathways and genomic architecture in bacteria. We worked with Dmitry Rodionov and Mikhail Gelfand to reconstruct the metabolism of nitrogen oxides in diverse bacteria and later used this information to interpret our functional genomic data on nitrogen oxide stress response in DvH.

We have also used the MicrobesOnline framework for discovering core functionality. For example, we have been able to define a set of "signature" genes that are found in sulfate reducers as diverse of delta-proteobacteria and archaea. We can show that these genes, predicted entirely through comparative sequence analysis, also show coordinated gene expression patterns in DvH.

Additionally, we have derived a novel theory of the full life-cycle of operons in bacteria: how they are born, are tuned, and die. Our findings suggest that operon evolution is driven by selection on gene expression patterns. First, both operon creation and operon destruction lead to large changes in gene expression patterns. For example, the removal of lysA and ruvA from ancestral operons that contained essential genes allowed their expression to respond to lysine levels and DNA damage, respectively. Second, some operons have undergone accelerated evolution, with multiple new genes being added during a brief period. Third, although most operons are closely spaced because of a neutral bias towards deletion and because of selection against large overlaps, highly expressed operons tend to be widely spaced because of regulatory fine-tuning by intervening sequences. Although operon evolution seems to be adaptive, it need not be optimal: new operons often comprise functionally unrelated genes that were already in proximity before the operon formed.

In studying the comparative genomics of the two-component systems that lie at the heart of control of many of the stress responses we are studying we discovered that different organisms use different strategies for generation and acquisition of new sensory histidine kinases. We analyzed the

* Presenting author

phylogenetic distribution of nearly 5000 histidine protein kinases from 207 sequenced prokaryotic genomes. We found that many genomes carry a large repertoire of recently evolved signaling genes, which may reflect selective pressure to adapt to new environmental conditions. Both lineage-specific gene family expansion and horizontal gene transfer play major roles in the introduction of new histidine kinases into genomes; however, there are differences in how these two evolutionary forces act. Genes imported via horizontal transfer are more likely to retain their original functionality as inferred from a similar complement of signaling domains, while gene family expansion accompanied by domain shuffling appears to be a major source of novel genetic diversity. Family expansion is the dominant source of new histidine kinase genes in the genomes most enriched in signaling proteins, such as DvH and other environmental microbes, and detailed analysis reveals that divergence in domain structure and changes in expression patterns are hallmarks of recent expansions. These results lead us to conclude that in the ongoing evolution of bacterial signal transduction machinery, some organisms serve as 'producers' generating novel genetic diversity, while others serve as 'consumers' capitalizing on the existing diversity of their peers.

111

Genome-Wide Mapping of Transcriptional Networks in *Escherichia coli* and *Shewanella oneidensis* MR-1

G. Cottarel, M.E. Driscoll* (mdriscol@bu.edu), J. Faith, M.A. Kohanski, B. Hayete, J. Wierzbowski, C.B. Cantor, **J.J. Collins**, and T.S. Gardner

Boston University, Boston, MA

Both *E. coli* and *Shewanella* possess versatile respiratory networks for energy production, but owing to their distinct ecological niches, this versatility is found at different ends in their redox pathways. *E. coli* thrives on a wide range of electron donors, yet can respire only with oxygen and a few organic compounds. Conversely, *Shewanella*'s use of electron donors is comparatively narrow, but it can respire using dozens of electron acceptors, including solid metals.

The ability to survive in a range of nutritive environments is aided by having the right respiratory enzymes active at the right times. These regulatory responses are conferred by rich transcriptional networks in both *E. coli* and *Shewanella*, which translate changes in environment into changes in gene expression.

To further identify these transcriptional regulatory networks in *E. coli* we have assembled a data set of 505 Affymetrix *E. coli* microarray expression profiles representing over 200 unique conditions and mutants. We assayed 269 of the arrays in our lab, and compiled the other 236 from studies in other laboratories. To our knowledge, this data set represents the largest uniformly collected and normalized microarray dataset for a prokaryote.

We are currently generating a comparable number of expression profiles for *Shewanella*, representing approximately 300 unique growth conditions, as well as mutants generously provided by the *Shewanella* Federation, using an Affymetrix chip we have designed for this organism.

To infer networks from this microarray data, we have developed a novel algorithm based on Bayesian network theory. The algorithm is similar to our NIR algorithm in that it determines the topology of the regulatory network by looking for the most likely regulators of each gene accord-

ing to gene expression data. Key improvements over the NIR algorithm enable application at the genome-scale and produce more accurate and biologically meaningful regulatory models. These include: (1) the ability to capture nonlinear combinatorial regulatory relationships such as Boolean or thermodynamic gene regulation functions; (2) the ability to incorporate prior information to improve the accuracy of the reverse-engineered network model; and (3) implementation of a Multi-Chain Markov-Chain Monte Carlo numerical optimization algorithm and a model averaging procedure to enable more reliable identification of network topology.

We compared our algorithm's predictions for *E. coli* to RegulonDB, a database of validated transcription factor/promoter interactions. Though RegulonDB is not a complete description of *E. coli* transcription regulation, it is the most comprehensive available. Our algorithm identified 383 regulatory connections, 114 of which have been identified in RegulonDB, and 269 of which are novel.

The wealth of existing knowledge about *E. coli* transcriptional regulation has made it a model organism for improving the performance and validating the results of our network inference algorithms. Though less well-studied, *Shewanella*'s unique metal-reducing capabilities presents an opportunity to explore our transcription network predictions in both bioremediation and microbial fuel cell applications.

112

Computational Approaches to Investigating Transcription Regulatory Networks and Molecular Evolution of the Metal-Reducing Family *Geobacteraceae*

Julia Krushkal[1]* (jkrushka@utmem.edu), Marko Puljic[1], Ronald M. Adkins[1], Jeanette Peeples[1], Bin Yan[1,2], Radhakrishnan Mahadevan[3], and **Derek R. Lovley**[4]

[1]University of Tennessee Health Science Center, Memphis, TN; [2]National Institutes of Health, Bethesda, MD; [3]Genomatica, Inc., San Diego, CA; and [4]University of Massachusetts, Amherst, MA

The evolutionary dynamics of gene regulatory mechanisms among living organisms is one of the fundamental questions that is key to our understanding of their ability to adapt to diverse environments and for our ability to manipulate their biology for human needs. We are investigating these mechanisms in *Geobacteraceae*, a metal-reducing family of delta-*Proteobacteria* that participate in bioremediation of contaminated environments and in energy harvesting. We are employing several complementary computational strategies to unveil transcriptional regulatory networks of this environmentally important group of microorganisms and to better understand how genetic differences allow members of this family to adapt to a variety of environmental conditions.

Our transcriptome analysis of delta-*Proteobacteria* involves the prediction of the operon organization of their genomes. We used information obtained from genome sequencing projects of species of delta-*Proteobacteria* to predict operon organization of completed genomes and partial assemblies of *Geobacter sulfurreducens, G. metallireducens, G. uraniumreducens*, *Desulfuromonas acetoxidans* and *D. palmitatis, Pelobacter carbinolicus* and *P. propionicus, Desulfotalea psychrophila, Desulfovibrio desulfuricans, D. vulgaris,* and *Bdellovibrio bacteriovorus*. Operon organization was also predicted for an environmentally important member of beta-*Proteobacteria, Rhodoferax ferrireducens.* The fit of sequence based operon predictions to results of genome wide measurement of gene tran-

* Presenting author

Figure 1. An example of scored predicted regulatory sites in a user-specified region of the *G. sulfurreducens* genome.

scription and protein production is being evaluated using statistical tests such as ANOVA and permutation analyses.

Building upon our earlier analyses of operon prediction, comparative genomic information, gene expression microarray data, and sequence similarity comparisons, we developed a database containing thousands of predicted transcription regulatory regions of *G. sulfurreducens*. Computational clustering techniques were employed to group those sites that had nearly identical genome locations. Subsequent sorting approaches allowed us to identify those individual sites and regulatory clusters that were predicted by the highest number of gene expression and sequence comparison data sets. These genome-wide predictions pinpoint the regulatory sites and transcription factors that have high likelihood of involvement in important regulatory pathways. For example, the highest scoring regulatory cluster independently predicted from 9 different microarray and sequence prediction sources was located upstream of an operon encoding proteins GSU2941 and GSU2942 that correspond to a transcription regulator from the LysR family and a methyl-accepting chemotaxis protein, respectively. Another high-scoring regulatory cluster predicted from 7 different microarray and sequence analyses sources was located upstream of GSU0655 that corresponds to the σ32 subunit of RNA polymerase (RpoH). These regulators and the upstream regions of their operons likely affect multiple regulatory pathways of *G. sulfurreducens*.

We have developed an online tool that allows a user to query the entire database of predicted regulatory elements in order to identify transcription regulatory sites in a genome region of interest or upstream of an operon of interest (Figure 1). These regulatory elements are ranked based on the number of sources of their prediction. We have utilized this resource to provide ranked predictions of transcription regulatory sites that may regulate expression of dozens of genes important for environmental bioremediation and energy production by *G. sulfurreducens.* Additionally, our database provides thousands of ranked predictions of components of transcription regulatory networks that involve transcription factor-operon interactions. These predicted genome-wide regulatory interactions are now being verified using gene expression microarray data from multiple experiments.

We are using comparative phylogenetic analyses to investigate the molecular evolutionary changes that allowed representatives of *Geobacteraceae* to have a diverse range of physiological and metabolic properties in a variety of environmental conditions. For example, we are investigating the molecular evolution of multiple protein sequences from *Pelobacter carbinolicus,* a cytochrome-poor relative of *G. sulfurreducens.* The pipeline of phylogenetic analyses includes: 1) Blastp sequence similarity searches of microbial proteome data using *P. carbinolicus* and *G. sulfurreducens* protein sequences as queries, 2) ClustalW automated alignments of homologous sequences derived from complete genomes at three levels of phylogenetic divergence: a) other species of delta-*Proteobacteria*, b) all species of *Proteobacteria*, and c) all species of *Bacteria* and *Archea* and 3) calculation of distance matrices and phylogenetic tree inference based on each aligned data set using software packages MEGA, PHYLIP, and PAUP* (Figure 2). In another comparative

phylogenetic study of transcription regulatory interactions, we are investigating the molecular evolution of each predicted transcription factor of *G. sulfurreducens* by inferring phylogenetic trees containing their homologs from other bacterial and archaeal species.

Figure 2. Pipeline of comparative evolutionary analyses of proteome data of representatives of *Geobacteraceae*.

113

Geobacter Project Subproject IV: Regulatory Networks Controlling Expression of Environmentally Relevant Physiological Responses of *Geobacteraceae*

Byoung-Chan Kim[1]* (bckim@microbio.umass.edu), Katy Juarez-Lopez[1], Hoa Tran[1], Richard Glaven[1], Toshiyuki Ueki[1], Regina O'Neil[1], Ching Leang[1], Gemma Reguera[1], Allen Tsang[1], Robert Weis[1], Marianne Schiffer[2], and **Derek Lovley**[1]

[1]University of Massachusetts, Amherst, MA and [2]Argonne National Laboratory, Argonne, IL

In order to predictively model the physiological response of *Geobacteraceae* under different environmental conditions and to rationally optimize practical applications of these organisms in bioremediation and harvesting electricity from biomass, it is necessary to understand how the expression of genes encoding important physiological functions is regulated. As reported last year, several global regulatory systems, such as RpoS, have now been well characterized in *Geobacter sulfurreducens*. Additional global regulatory systems are being elucidated and more detailed studies on the regulation of key genes are being conducted.

We have recently discovered that *G. sulfurreducens* produces novel, electrically conductive pili that appear to be the electrical conduits to Fe(III) and Mn(IV) oxides and possibly electrodes. The production of these 'microbial nanowires' appears to be highly regulated. For example, at optimal growth temperatures, the nanowires are produced during growth on insoluble Fe(III) or Mn(IV) oxides, but not during growth on soluble electron acceptors. A combination of genetic, microarray, and proteomic investigations revealed that expression of pilA, which encodes the structural protein for the nanowires, is regulated by a two-component regulatory system in which the response regulator, PilR, functions as a sigma 54-EBP (Enhancer Binding Protein). Furthermore, PilR regulates the expression of several c-type cytochromes known to be important in electron transfer to Fe(III). This is the first instance of in which a sigma 54-EBP has been implicated in regulating respiratory functions. In addition, microarray analysis of a mutant incapable of producing acetyl-phosphate suggested that acetyl-phosphate concentrations modulate the activity of PilR, establishing a link between central metabolism and expression of important respiratory genes. These results are important not only for understanding how electron transfer to insoluble electron acceptors is controlled, but also for providing insights into strategies for over-expressing microbial nanowires for practical applications.

* Presenting author

OmcS is an outer membrane c-type cytochrome that is essential for the reduction of Fe(III) oxide and the production of electricity. Expression of omcS was found to be regulated via multiple mechanisms. Evaluation of gene expression in the appropriate mutants revealed that the global regulators Fur, FNR1, and FNR2 all affect the expression of the omcS operon. In addition, omcS expression also appears to be regulated by a two component signal transduction system encoded upstream of the omcS operon. The response regulator associated with this two component system was demonstrated to bind upstream of the omcS start site. We have constructed a knockout mutant lacking the sensor/regulator pair as well as an inducible vector for overexpressing the response regulator in this mutant with the goal of overexpressing the omcS operon for the purposes of optimizing electrical energy harvesting.

The outer-membrane c-type cytochrome OmcB appears to be an essential intermediary electron carrier for extracellular electron transfer. Last year we reported that several outer-membrane c-type cytochromes influence OmcB expression. Some were required for transcription and others were required for translation and/or maturation. In addition, a gene encoding a putative regulatory protease located adjacent to omcF, an outer membrane cytochrome required for omcB transcription, was also found to be required for OmcB production and Fe(III) reduction. These results coupled with recent findings that expression of OmcB is also regulated by RpoS, levels of pppGpp, the transcriptional regulator OrfR, as well PilR demonstrate that multiple regulatory circuits modulate the expression of this central electron transfer component.

Phosphate is often a limiting nutrient in environments in which *Geobacteraceae* predominate. The pho operon, which responds to inorganic phosphate, has been well characterized in other microorganisms, and we have found a homologous two-component signal transduction system that regulates phosphate uptake in *G. sulfurreducens*. We have constructed a knockout mutant lacking the sensor/regulator pair as well as an inducible vector that will overexpress the response regulator. Microarray analysis is being performed to determine the regulatory network that responds to inorganic phosphate availability and to improve regulation inputs for this aspect of the in silico model of *G. sulfurreducens* metabolism. Interestingly, this signal transduction system may act as a global regulator as it influences the transcription of a number of other transcription factors.

The surprising findings that *Geobacter* species are highly planktonic in subsurface environments and that chemotaxis may be an important mechanism for localizing Fe(III) oxides has led to further evaluation of their chemotaxis mechanisms. There are 6 major clusters of chemotaxis genes in *G. sulfurreducens* and 7 in *Geobacter metallireducens*, and these organisms possess large number of chemoreceptors: 20 in *G. metallireducens* and 34 in *G. sulfurreducens*. A novel high throughput signal screening method was developed in order to identify the signals to which these receptors respond. In this assay, individual *Geobacter* chemoreceptors, or chimeric receptors consisting of *Geobacter* sensing domains linked to the cytosolic fragment of the *E. coli* chemoreceptor Tar, are expressed in two *E. coli* strains with defects in chemotaxis, the chemoreceptor-deficient strain UU1250 and the adaptation-deficient strain RP1273, and screened for changes in motility in response to an array of potential signaling molecules. Attractants for a *G. sulfurreducens* chemoreceptor were identified using both the native chemoreceptor and a *Geobacter/E. coli* chimera and included acetate and other organic acids. This is significant because acetate is the primary electron donor supporting the growth and activity of *Geobacter* species in subsurface environments and on energy-harvesting electrodes. By fusing signal sensor of histidine kinases to cytosolic fragment of the *E. coli* chemoreceptor, Tar, this strategy can also be used to elucidate the signal specificity of abundant two component systems of the *Geobacteraceae*. We are currently utilizing this approach to identify the environmental signal that the omcS-regulating two component system, described above, responds to.

Additional studies on global regulators including: the sigma factors RpoH, RpoE, and RpoN; the iron response regulators, Fur and IdeR; Fnr; secondary messengers; and other two-component regulatory systems, including novel systems with c-type heme binding motifs in the sensor, are in progress and will be summarized in the presentation.

114

Advances in System Level Analysis of Bacterial Regulation

Lucy Shapiro[1], Michael Laub[2], Ken Downing[3], Alfred Spormann[1], and **Harley McAdams**[1]* (hmcadams@stanford.edu)

[1]Stanford University, Stanford, CA; [2]Harvard University, Cambridge, MA; and [3]Lawrence Berkeley National Laboratory, Berkeley, CA

In the last year we have made major advances in identification of the complete genetic circuitry that runs the cell cycle of the bacterium *Caulobacter crescentus*. We have shown that DnaA acts as a third global regulator that, with CtrA and GcrA, controls the temporal sequencing of the genetic modules that implement the cell cycle. Using custom designed Affymetrix chips, we have identified 769 transcription start sites and 27 conserved promoter motifs used by cell cycle-regulated genes and by genes responding to heavy metal exposure. Two component signal transduction proteins are the primary phosphotransfer system that regulates and coordinates the expressions many cell cycle events. The Laub lab has constructed deletions of each of the 106 *Caulobacter* two component genes and developed a novel technique to identify the targets of the histidine kinases allowing the identification of novel phosphor-regulatory pathways. We are using state-of-the-art fluorescence microscopy and cryo-EM tomography to demonstrate the role of dynamic positioning of regulatory proteins, structural complexes, and chromosomal loci in the overall regulation of the cell as a three dimensional system.

DnaA coordinates replication initiation and cell cycle transcription in *Caulobacter crescentus*. The McAdams and Shapiro labs have shown that DnaA, a critical protein involved in initiation of chromosome replication, also coordinates DNA replication initiation with cell cycle progression by acting as a global transcription factor (Hottes et al. 2005). DnaA functions as a critical transcription factor required for the expression of the gene encoding the GcrA master regulator whose levels oscillate with CtrA during the cell cycle. The redundant control of *gcrA* transcription by DnaA (activation) and CtrA (repression) forms a robust switch controlling the decision to proceed through the cell cycle or to remain in the G1 stage. Other genes in the DnaA regulon include those encoding nucleotide biosynthesis enzymes and components of the DNA replication machinery. Thus, DnaA not only initiates DNA replication, but also promotes the transcription of the components necessary for successful chromosome duplication. DnaA activation of transcription of *ftsZ* and *podJ* starts the cell division and polar organelle development processes that, in addition to DNA replication, prepare the cell for asymmetric division.

Degradation of CtrA requires a dynamically localized ClpXP protease complex and a specificity factor (Iniesta et al.; McGrath et al. 2005). The McAdams and Shapiro labs have shown that the ClpXP protease, which is responsible for the degradation of multiple bacterial proteins, is dynamically localized to specific cellular positions in *Caulobacter,* where it degrades co-localized substrates. For example, the CtrA cell cycle master regulator co-localizes with the ClpXP protease at the stalked cell pole at the swarmer to stalked cell transition and, in the stalked daughter cell compartment

* Presenting author

immediately after cytoplasmic compartmentalization (McGrath, Iniesta et al. 2005), well before daughter cell separation (Judd et al. 2003). C-localization of CtrA with ClpXP is essential for CtrA degradation that enables initiation of chromosome replication. By a combination of bioinformatic, biochemical, genetic, and fluorescent microscopy techniques, we identified two conserved proteins, RcdA and CpdR, that are essential for CtrA degradation. RcdA directly interacts with CtrA and ClpX *in vivo* to mediate CtrA degradation (McGrath, Iniesta et al. 2005). Unphosphorylated CpdR, a response regulator, acts to localize ClpXP at the cell pole (Iniesta, McGrath et al.).

Distinct Constrictive Processes, Separated in Time and Space, Divide *Caulobacter* Inner and Outer Membranes (Judd et al. 2005). Tomographic cryoEM images of the cell division site show separate constrictive processes closing first the inner membrane (IM) and then about 20 minutes later, the outer membrane (OM) in a manner distinctly different from that of septum-forming bacteria. In the early stages of cell division, the inner and outer membranes constrict simultaneously, maintaining their 30-nm separation as seen in regions distant from the constriction. As cell division progresses, the IM constricts faster, creating a growing distance between the inner and outer membranes near the division plane until fission of the inner membrane creates a cell containing two inner membrane-bound cytoplasmic compartments surrounded by a single continuous outer membrane.

High-throughput identification of 769 transcription start sites and 27 DNA regulatory motifs (McGrath et al.). Using 62 data sets of transcription profiles obtained with a custom-designed Affymetrix chip, the McAdams lab identified transcriptional start sites of 769 genes (53 transcribed from multiple start sites. Transcriptional start sites were identified by analyzing the cross-correlation matrices created from the probes tiled every 5 bp upstream of the gene. Motif-searching upstream of the start sites within co-expressed promoters yielded 14 cell cycle regulator binding motifs (8 previously unknown) and 13 heavy metal response regulator motifs (10 previously unknown). This is a ten-fold increase in known *Caulobacter* transcription start sites and a doubling of known binding motifs.

Identification of a Novel Cell Cycle Checkpoint (Spangler et al.).Using DNA microarrays, the Laub lab mapped the response of wild type *Caulobacter* cells to DNA damage and identified two major transcriptional responses: induction of an SOS regulon and repression of genes activated by CtrA, the cell cycle master regulator. Included in the SOS regulon is a novel, but highly conserved gene named *cciA* which is responsible for preventing cell cycle progression after DNA damage. CciA co-localizes with the chromosome segregation machinery component topoisomerase IV and directly inhibits both gyrase and topo IV. These results suggest that DNA damage leads to a delay in cell division by induction of CciA to directly inhibit chromosome segregation which blocks completion of division. The CciA system is the first cell cycle checkpoint identified in *Caulobacter*.

Systematic Analysis of Two-Component Signal Transduction (Skerker et al. 2005). The Laub lab deleted each of the 106 *Caulobacter* two-component signal transduction genes. Thirty-nine of the genes are required for growth, morphogenesis, or cell cycle progression; 9 are essential. A novel technique to identify cognate pairs, called phosphotransfer profiling was developed to identify response regulator targets of histidine kinases. This *in vitro* biochemical technique successfully identifies the specific, *in vivo*-relevant targets of histidine kinases. A library of *Caulobacter* two-component deletion strains was constructed such that each strain harbors a pair of unique bar-codes (20mers) which enable high-throughput fitness analyses to complement microarray and genetic experiments.

Dissection of Genetic Diversity in Subpopulation of *Caulobacter crescentus* Biofilms. The Spormann lab has shown that *C. crescentus* biofilms exhibit a bi-phasic architecture: flat, monolayer biofilms containing interspersed mushroom-like structures that are the result of clonal growth (Entcheva-Dimitrov et al. 2004). Cells grown from the mushroom structures auto-aggregate in liquid

static or shaken culture, carry increased cell surface hydrophobicity, and show reduced swarming motility. These traits are inheritable. The mushroom phenotype results from a single point mutation in ORF CC3629 (*rfbB*), encoding dTDP-D-glucose-4,6-dehydratase, which is involved in the O-antigen biosynthesis pathway of lipopolysaccharides (LPS), showing that LPS components are critical determinants of *Caulobacter* biofilm architecture. Our results show genetic variants accumulate rapidly in biofilms and represent an important pool and mechanism for generating microbial diversity.

References

1. Entcheva-Dimitrov, P. and A. M. Spormann (2004). "Dynamics and control of biofilms of the oligotrophic bacterium *Caulobacter crescentus*." *J Bacteriol* **186**(24): 8254-66.
2. Hottes, A. K., L. Shapiro and H. H. McAdams (2005). "DnaA coordinates replication initiation and cell cycle transcription in *Caulobacter crescentus*." *Mol Microbiol* **58**(5): 1340-53.
3. Iniesta, A. A., P. T. McGrath, A. Reisenauer, H. H. McAdams and L. Shapiro "A phospho-signal cascade coupled to cytokinesis controls the dynamic positioning and activity of a proteolysis machine critical for bacterial cell cycle progression." In preparation.
4. Judd, E. M., L. R. Comolli, J. C. Chen, K. H. Downing, W. E. Moerner and H. H. McAdams (2005). "Distinct constrictive processes, separated in time and space, divide *Caulobacter* inner and outer membranes." *J Bacteriol* **187**(20): 6874-82.
5. Judd, E. M., K. Ryan, W. E. Moerner, L. Shapiro and H. H. McAdams (2003). "Fluorescence bleaching reveals asymmetric compartment formation prior to cell division in *Caulobacter*." *Proc. Natl. Acad. Aci. USA* **100**(14): 8235-8240.
6. McGrath, P. T., A. A. Iniesta, K. R. Ryan, L. Shapiro and H. H. McAdams (2005). "Controlled degradation of a cell cycle master regulator requires a dynamically localized protease complex and a polar specificity factor." *Cell*: In press.
7. McGrath, P. T., H. Lee, A. A. Iniesta, A. K. Hottes, P. Hu, L. Shapiro and H. H. McAdams "Identification of transcriptional start sites using high density arrays." In preparation.
8. Skerker, J. M., M. S. Prasol, B. S. Perchuk, E. G. Biondi and M. T. Laub (2005). "Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis." *PLoS Biol* **3**(10): e334.
9. Spangler, J. E., M. S. Prasol and M. T. Laub "A Novel Checkpoint System Regulates Cell Cycle Progression after DNA Damage in *Caulobacter crescentus.*" In preparation.

This work was funded in part by the U.S. Department of Energy's Genomics:GTL program under two projects, "Global Characterization of Genetic/Regulatory Circuitry Controlling Adaptive Metabolic Pathways" and "Advanced methods for high-resolution imaging of dynamic cellular and subcellular processes in complex bacterial systems".

 * Presenting author

115

A System-Level Analysis of Two-Component Signal Transduction

Michael T. Laub[1] (laub@cgr.harvard.edu), Lucy Shapiro[2], and Harley H. McAdams[2]

[1]Harvard University, Cambridge, MA and [2]Stanford University, Stanford, CA

Two-component signal transduction systems, comprised of histidine kinases and their response regulator substrates, are the predominant means by which bacteria sense and respond to signals. These systems allow cells to adapt to prevailing conditions by modifying cellular physiology, including initiating programs of gene expression, catalyzing reactions, or modifying protein-protein interactions. These signaling pathways have also been demonstrated to play a role in coordinating bacterial cell cycle progression and development. We have initiated a system-level investigation of two-component pathways in the tractable model organism *Caulobacter crescentus*, which encodes 62 histidine kinases and 44 response regulators. Comprehensive deletion and overexpression screens have identified more than 40 of these 106 two-component genes as required for growth, viability, or proper cell cycle progression in standard laboratory growth conditions. In addition, the creation of a comprehensive library of deletion mutants enables the identification of two-component signaling genes required for survival in alternative growth conditions and in response to environmental changes. These studies are done in a quantitative and high-throughput manner using a pooled, bar-coding strategy similar to that used for the yeast genome-wide deletion project. Comparison of these experiments with DNA microarray experiments under identical conditions demonstrate that the signal transduction genes *required* for a response show virtually no overlap with the set of signaling genes whose expression level changes.

As with most bacterial species, the majority of genes encoding histidine kinases in *Caulobacter* are orphans – i.e. not encoded in an operon with their cognate response regulator – demanding other approaches for mapping signaling pathways. To address this need we have developed a novel systematic biochemical approach, called phosphotransfer profiling, to map the connectivity of histidine kinases and response regulators. By combining genetic and biochemical approaches, we have begun mapping pathways critical to growth and cell cycle progression. This includes a novel essential two-component signaling pathway, CenK-CenR, which controls cell envelope biogenesis, as well as a complex phosphorelay controlling CtrA, the master regulator of the *Caulobacter* cell cycle. Specific examples will be presented to demonstrate how the combination of techniques, applicable to any bacterial species, can be used to rapidly uncover regulatory networks.

The ability of our *in vitro* phosphotransfer profiling method to identify signaling pathways that are relevant *in vivo* takes advantage of an observation that histidine kinases are endowed with a global, kinetic preference for their cognate response regulators. This system-wide selectivity helps insulate two-component pathways from one another, preventing unwanted cross-talk. Moreover, it suggests that the specificity of two-component signaling pathways is determined almost exclusively at the biochemical level. We are using computational and mutagenesis methods to map the amino acids which confer specificity. The results may enable computational prediction of two-component pairings in any organism and the rational design of novel signaling pathways for constructing biosensors or synthetic genetic circuits.

116

The Bacterial Birth Scar, a Spatial Determinant for the Positioning of a Polar Organelle at the Late Cytokinetic Site

Edgar Huitema, Sean Pritchard, David Matteson, Sunish Kumar Radhakrishnan, and Patrick H. Viollier* (Patrick.Viollier@case.edu)

Case Western University, Cleveland, OH

Many prokaryotic protein complexes underlie polar asymmetry. In *Caulobacter crescentus* a flagellum is built exclusively at the pole that arose from the previous cell division. The basis for this pole-specificity is unclear, but could involve a cytokinetic birth scar that marks the newborn pole as the flagellum assembly site. We identified two novel developmental proteins, MadX and MadR, which localize to the division septum and the newborn pole after division. We show that septal localization of MadX/R depends on cytokinesis. Moreover MadR, a c-di-GMP phosphodiesterase homolog, is a flagellum assembly factor that relies on MadX for proper positioning. In the absence of MadX, flagella are assembled at ectopic locations and MadR is mislocalized to such sites. Thus MadX and MadR establish a link between bacterial cytokinesis and polar asymmetry, demonstrating that division indeed leaves a positional mark in its wake to direct the biogenesis of a polar organelle.

This work was funded in part by the U.S. Department of Energy's Genomics:GTL program under two projects, "*Global Characterization of Genetic/Regulatory Circuitry Controlling Adaptive Metabolic Pathways"* and *"Advanced methods for high-resolution imaging of dynamic cellular and subcellular processes in complex bacterial systems"*.

117

Anaerobic Respiration Gene Regulatory Networks in *Shewanella oneidensis* MR-1

Jizhong Zhou[1,2]* (jzhou@ou.edu), Haichun Gao[1,2,3], Xiaohu Wang[5], Yunfeng Yang[2], Xiufeng Wan[6], Zamin Yang[2], Dawn Klingeman[2], Alex Beliaev[4], Soumitra Barua[1,2], Ting Li[2], Christopher Hemme[1,2], Yuri A. Gorby[4], Mary S. Lipton[4], Steve Brown[2], Margaret Romine[4], Kenneth Nealson[7], James M. Tiedje[3], Timothy Palzkill[5], and James K. Fredrickson[4]

[1]University of Oklahoma, Norman, OK; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Michigan State University, East Lansing, MI; [4]Pacific Northwest National Laboratory, Richland, WA; [5]Baylor College of Medicine, Houston, Texas; [6]Miami University, Oxford, OH; and [7]University of Southern California, Los Angeles, CA

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

* Presenting author

ArcA. In *Escherichia coli*, metabolic transitions between aerobic and anaerobic growth states occur when cells enter an oxygen-limited condition. Many of these metabolic transitions are controlled at the transcriptional level by the activities of the global regulatory proteins ArcA (aerobic respiration control) and Fnr (fumarate nitrate regulator). A homolog of ArcA (81% amino acid sequence identity) was identified in *S. oneidensis* MR-1, and *arcA* mutants with either MR-1 or Dsp10 (a spontaneous rifampicin-resistant mutant) as the parental strains were generated. The *arcA* deletion mutant grew slower than the wild type and was hypersensitive to H_2O_2. Microarray analysis indicated that S. *oneidensis* ArcA regulates a large number of genes that are not within the *E. coli* ArcA regulon although a small set of genes were found overlapping.

The *S. oneidensis arcA* gene and a mutated *arcA* gene carrying the point mutation of D54N were cloned and expressed in *E. coli*. Both the wild type and the modified ArcA proteins were purified and their DNA binding properties were analyzed by electrophoretic motility shift (EMS) and DNase I footprinting assays. The results indicated that the phosphorylated ArcA proteins were able to bind to a DNA site in the MR-1 genome similar in sequence to the *E. coli* ArcA binding site. The common feature of the binding site is the presence of a conserved 15 bp motif with 2-3 mismatches to the *E. coli* ArcA-P consensus binding motif. Additional EMS experiments revealed that the *S. oneidensis* ArcA proteins can bind to the promoter region of genes whose products are involved in the anaplerotic shunt (*sfcA*), hydrogen metabolism and the terminal DMSO reductase in a phosphorylation-dependent manner. Genome scale computational predictions of binding sites were also performed and 331 putative ArcA regulatory targets were identified. Although the computational screen is in need of refinement, the results suggest the *S. oneidensis* and *E. coli* ArcA-P proteins may differ significantly in terms of the regulation of energy metabolism/respiration. For example, the *ptsG* and *aceBA* are the only two overlapping operons shared between the *E. coli* and *S. oneidensis* ArcA-P modulons. Despite the fact that both ArcA proteins bind to a similar DNA motif, the regulation of aerobic/anaerobic respiration may be more complex than expected in *S. oneidensis*.

EtrA. A homolog of *E. coli Fnr* was identified in MR-1, termed *etrA* (electron transport regulator), which showed a high degree of amino acid identity (51%). An *etrA* deletion mutant was generated but no obvious phenotypic differences under nitrate and fumurate conditions between the wild type and the *etrA* mutant were observed except that the mutant is hypersensitive to H_2O_2. Both strains were individually grown in continuous culture for 410 h under fully aerobic followed microaerobic steady state conditions and then transitioned to anaerobic steady state growth with fumarate as the electron acceptor. Microarray analysis revealed that about 20% of the ORFs showed significant differences in expression between the *etrA* mutant and the parental strains under these growth conditions. The study also revealed that expression profiles of aerobic and microaerobic cultures shared a high level of similarity but greatly differed from that of anaerobic culture. Among genes whose upstream region contains a conserved *E. coli* EtrA-binding motif, only a small number were found significantly affected by the mutation. In addition, the protein expression analysis of the steady state cultures indicated that approximately 20% of all proteins showed significant changes under the steady state conditions. While some genes were significantly changed in both mRNA and protein profiles, significant differences in expression profiles were also noticed. While the exact role of EtrA in *S. oneidensis* remains unknown, it is evident that EtrA of S. *oneidensis* differs from Fnr of *E. coli* significantly in functionalities.

Regulation of nitrate respiration. *S. oneidensis* MR-1 has a NarQ/NarP two-component system. Genome predictions indicate that NarP is the regulator belonging to the LuxR/UhpA family while NarQ is the membrane-anchored sensor. NarP controls the expression of several genes involved in anaerobic respiration and fermentation in *E. coli*. In the presence of nitrate, NarP can be phosphor-

ylated by phospho-NarQ in *E. coli*. In this activated state, phospho-NarP can act as an activator of transcription of the nitrate reductase systems and as a repressor of expression of genes for other anaerobic systems. Two in–frame deletion mutants, *narP* and *narP/narQ*, were generated in MR-1. The *narP* mutant failed to grow on nitrate and nitrate reduction was not evident. In contrast, the *narP/narQ* double mutant grew well on nitrate with much higher biomass yields than MR-1. Nitrate was reduced to ammonia in the *narP/narQ* double mutant whereas nitrate was reduced to nitrite in MR-1. A parallel study on the *S. oneidensis* NAP system demonstrated that NAP was essential for growth of MR-1 on nitrate. These results suggest in *S. oneidensis* that: (1) the NarP/NarQ two-component system controls the NAP system directly; (2) there may be another nitrate reduction pathway through which nitrate can be reduced to ammonia; (3) other regulatory systems controlling the alternative nitrate reduction pathway exist when the NarP/NarQ two-component system is absent. In addition, his-tagged NarP protein has been purified. Results of electrophoretic mobility shift assays (EMSA) showed that phosphorylated NarP was able to bind to its own promoter. The identity of the binding site for the *narQP* operon has been narrowed down to a 74 base pair region in the promoter. Investigations on NarP are focused on defining the binding sites and thereby the NarP regulon of *S. oneidensis*.

Small regulatory RNAs. Small regulatory RNAs have been hailed as the key to coordinate global regulatory circuits. However, essentially nothing is known regarding the potential role of regulatory RNAs in *S. oneidensis*. As an exploratory study, we investigate a small RNA named RyhB, whose counterpart in *E. coli* is known to be regulated by extracellular iron and Fur. RyhB was predicted in *S. oneidensis* and other *Shewanella* species, and then experimentally validated by Reverse Transcription-PCR. Preliminary results suggest that alike *E. coli* RyhB, the expression of *S. oneidensis* homolog is also regulated by iron and Fur. Interestingly, RyhB expression is specifically induced under iron-reducing condition.

In summary, although several global regulatory genes were identified in *S. oneidensis* MR-1 with high degree of amino sequence identity to those in *E. coli*, regulatory circuits in *S. oneidensis* MR-1 appear to have distinct roles in MR-1. This could reflect the complexity of lifestyles of *Shewanella* as well as the diverse environments where the bacterium lives.

 * Presenting author

118

Uncovering the Regulatory Network of *Shewanella*

Gary D. Stormo* (stormo@genetics.wustl.edu), Jiajian Liu, and Xing Xu

Washington University School of Medicine, St. Louis, MO

We are working to determine the regulatory sites that control the expression of gene in *Shewanella* species. We are interested in both transcription and post-transcriptional regulation, so we are considering both DNA motifs that bind to transcription factors and RNA motifs that regulate mRNA translation, attenutation and degradation.

Our analysis is being done in two phases. Because much is known about the *E. coli* genome and its regulatory network, we are first determining which factors and binding sites are conserved between the species, allowing us to infer conserved regulatory interactions in those cases. Comparing known and predicted transcription factors (TFs) from *E. coli* to the 5 *Shewanella* species with available genome sequences, we identify 41 TFs that occur in every genome. There are varying numbers that occur in some *Shewanella* species but not all (at least in part this is due to the incompleteness of some of the genome sequences). But for a large number of *E. coli* TFs we find no clear orthologous sequences (based on reciprocal best BLAST matches) in any of the *Shewanella* species. To follow up that study we are also determining which TFs occur only in *Shewanella* and what their distributions are across the different species.

In order to determine the set of conserved intergenic motifs, which represent putative regulatory sites, we are using the program PhyloNet (Wang and Stormo, PNAS 102:17400-5). This program uses profiles from aligned orthologous intergenic regions and compares them across the entire genome to identify those which occur multiple times and are likely to represent regulatory sites for TFs that control multiple genes. Because of the preponderance of dimeric TFs in bacteria, the parameters are the program are being optimized for that type of data. Preliminary results show that most of the regulatory sites for TFs that are conserved with *E. coli* can be identified by this approach. In addition many novel predicted motifs are obtained that probably correspond to TFs found only in the *Shewanella* species.

Our studies of post-transcriptional regulatory sites are initially focusing on identifying RNA structural motifs that are conserved with *E. coli*. Many such post-transcriptional regulatory sites are known in *E. coli*, including several recently described riboswitches. Some of these can be easily identified in *Shewanella*, whereas others are more difficult to detect. We are also beginning the search for novel post-transcriptional regulatory motifs in the 5'UTRs of *Shewanella* genes using software we have developed to identify conserved secondary structures.

119

Integration of Control Mechanisms and the Enhancement of Carbon Sequestration and Biohydrogen Production by *Rhodopseudomonas palustris*

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

[1]Ohio State University, Columbus, OH; [2]University of Washington, Seattle, WA; [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 integration and control of basic metabolic processes, we have shown that there is reciprocal regulation of CO_2 fixation and nitrogen fixation/hydrogen metabolism in this organism. Indeed, by blocking the processes by which *R. palustris* removes excess reducing equivalents generated from the oxidation of organic carbon, strains were constructed such that reducing equivalents could be 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

* Presenting author

coupled to the generation of hydrogen gas[4], 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[5]. During the course of these studies, novel genes and regulators were identified from investigating control of specific processes by conventional molecular biology/biochemical techniques[6]. These studies, along with the microarray and proteomics 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-transfer system for controlling CO_2 fixation gene expression was also identified and biochemically characterized and a unique signaling process was revealed[7]. Moreover, the key regulator was shown to possess motifs that potentially respond to diverse metabolic and environmental perturbations, suggesting an exquisite means for controlling 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.
4. Oda, Y., Samanta, S. K., Rey, F. E., Wu, L., Liu, X., Yan, T., Zhou,J., and C. S. Harwood. "Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*." *J. Bacteriol.* **187** (2005) 7784-7794
5. VerBerkmoes, N. C., M. B. Shah, P.K. Lankford, D. A. Pelletier, M. B. Strader, D. L. Tabb, W. H. McDonald, J. W. Barton, G. B. Hurst, L. Hauser, B. H. Davison, J. T. Beatty, C. S. Harwood, F. R.Tabita, R. L. Hettich, and F. W. Larimer. "Determination and comparison of the baseline proteomes of the versatile microbe *Rhodopseudomonas palustris* under its major metabolic states." *J. Proteome Res.* (in press).
6. Braatsch, S.,J. Bernstein, F. Harrison, J. Morgan, J. Liao, C. Harwood, and J. T. Beatty. "Both of the two *ppsR* genes of *Rhodopseudomonas palustris* CGA009 encode repressors of photosynthesis gene expression." Submitted for publication.
7. Romagnoli, S., and F. R. Tabita. "A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the cbb_I region of *Rhodopseudomonas palustris* CGA010." Submitted for publication.

120

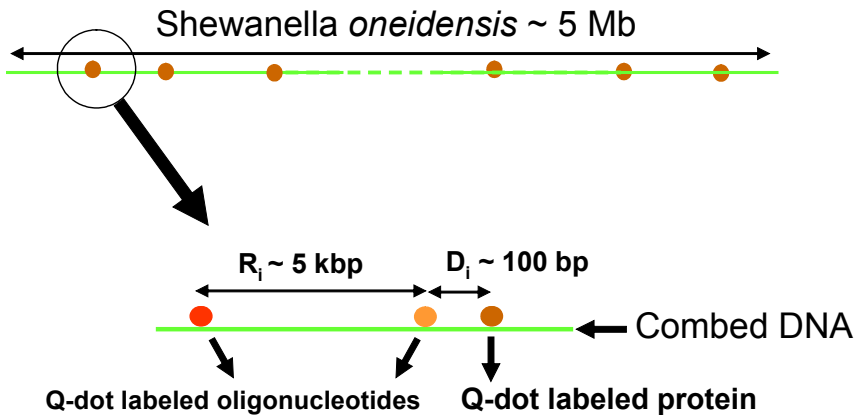
High-Resolution Physical Mapping of Transcription Factor Binding Sites in Whole Bacterial Genomes

J. Antelman[1], X. Michalet[1]* (michalet@chem.ucla.edu), S.B. Reed[2], M. Romine[2], and **S. Weiss**[1]

[1]University of California, Los Angeles, CA and [2]Pacific Northwest National Laboratory, Richland, WA

We propose to develop a novel technique for identifying and mapping regulator protein binding sites on bacterial genomes with very high-resolution, using quantum dots to uniquely tag both DNA and proteins. Our strategy to perform high-resolution physical mapping combines three technologies: DNA molecular combing[1], single-quantum dot labeling and detection[2], and ultrahigh-resolution multicolor colocalization[3]. The dissimilatory metal-reducing bacterium *Shewanella oneidensis* MR-1 will be used as a model system because its genome has been completely sequenced and variety of resources (protein-specific antibodies, purified proteins, clones for expressing and purifying proteins, cloning encoding promoter-encoding DNA) will be accessible from ongoing GTL funded work[4]. We will specifically study the binding sites of the catabolite repressor protein (CRP) protein, for which expression data is already available, but little DNA-protein interaction data. The localization precision (within ~25 bp) of this technique will provide significantly improved data for computational prediction of conserved binding sequences. Furthermore, the positioning of the binding sites relative to the RNA polymerase binding site will be used to predict whether the interaction leads to repression or enhancement of transcriptional activity.

The following diagram summarizes our approach:

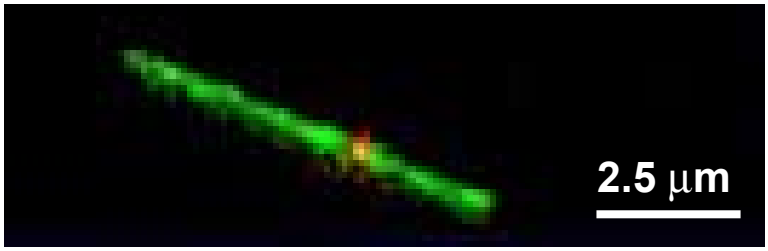


DNA binding proteins bound to DNA *in vivo* are cross-linked prior to DNA extraction. The DNA, with bound protein, is then site-specifically labeled with either a digoxigenin- or a biotin-labeled oligonucleotide probe. Following this, the DNA is stained using a fluorescent intercalating dye and stretched onto a hydrophobic glass surface using DNA molecular combing. After combing, the oligonucleotides are detected with antidigoxigenin-coated or streptavidin-coated quantum dots of various emission wavelength. Additionally, a quantum dot coated with antibodies against the protein of interest is added to visualize the bound protein. Precise measurement of the respective distances between the individual quantum dots (within a few dozen base pair) is then performed using multicolor stage-scanning confocal microscopy as described in ref. [3].

* Presenting author

To identify different oligonucleotide-labeled regions simultaneously, corresponding to different binding sites, unique nearby oligonucleotide probes labeled with digoxigenin and a biotin are used. To distinguish between regions, the probes are separated by different distance (R_i on the diagram above). Simultaneous AFM imaging is used as a control in the initial stages to ensure that both oligonucleotides and protein are attached to the same combed DNA strand. The end result is a complete, visual map of all binding sites for the studied protein.

Presently, two important accomplishments have been achieved towards the development of this technique. First, a novel method for rendering glass coverslips hydrophobic has been developed. These modified hydrophobic glass coverslips are required for the molecular combing process. The new approach is cheaper, faster and more robust than the previously published technique, and will allow any lab to use this protocol with standard laboratory materials. Secondly, we have modified a RecA mediated oligonucleotide-labeling technique, which allows for the site-specific labeling of dsDNA. These probes can be inserted at any sequence in genomic DNA. Finally, we have demonstrated the detection of this type of probes with single streptavidin- or antidigoxygenin-labeled quantum dots. Shown below is a streptavidin-labeled quantum dot bound to a biotin-labeled probe, hybridized on locus 4404-4434 of lambda phage DNA, counterstained with the DNA intercalating dye YOYO-1.



References

1. Michalet, X., et al., *Science* **277** (1997) 1518
2. Michalet, X., et al., *Science* **307** (2005) 538
3. Lacoste, T., et al., *Proc. Natl. Acad. Sci. USA* **91** (2000) 9461
4. Kolker, E., et al., *Proc. Natl. Acad. Sci. USA* **102** (2005) 2099