Section 2

Metabolic Network Experimentation and Modeling

90 MEWG

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

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

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

Photobiological production of H_2 from water has great appeal as an environmentally sustainable, long-term means to meet large, projected increases in demand, to provide energy and economic security for the U.S. and other nations, and to relieve environmental stresses related to fossil fuel use

This project is using the cyanobacterial species *Synechocystis* PCC 6803 as a model to pursue two parallel lines of inquiry initially, with each line addressing one of the two main factors affecting H_2 production in PCC 6803: NADPH availability and O_2 sensitivity. H_2 production in PCC 6803 requires a very high NADPH:$NADP^+$ ratio, that is, that the NADP pool be highly reduced, which can be problematic because several metabolic pathways potentially can act to raise or lower NADPH levels. Also, though the [NiFe]-H_2ase in PCC 6803 is constitutively expressed, it is reversibly inactivated at very low O_2 concentrations, reportedly due to binding of O_2 to the active-site. Largely because of this O_2 sensitivity and the requirement for high NADPH levels, a major portion of overall H_2 production occurs under anoxic conditions in the dark, supported by breakdown of glycogen or other organic substrates accumulated during photosynthesis. Also, other factors, such as N or S limitation, pH changes, presence of other substances, or deletion of particular respiratory components, can affect light or dark H_2 production. Therefore, in the first line of inquiry, under a number of culture conditions with wild-type (WT) PCC 6803 cells and a mutant with impaired type I NADPH-dehydrogenase (NDH-1) function, we are using H_2 production profiling and metabolic flux analysis, with and without specific inhibitors, to examine systematically the pathways involved in light and dark H_2 production. Results from this work will provide rational bases for metabolic engineering to maximize photobiological H_2 production on a 24-hour basis. In the second line of inquiry, we are using site-directed and random mutagenesis to create mutants with H_2ase enzymes exhibiting greater O_2 tolerance (and perhaps higher H_2 production activity). The objectives of the research are addressed via the following four tasks:

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

Task 3 is being conducted in parallel with Tasks 1 and 2; Task 4 will reflect the convergence of research performed in Tasks 1-3.

* Presenting author

91

Geobacter Project Subproject V: *In Silico* Modeling of the Growth and Physiological Responses of *Geobacteraceae* in Complex Environments

Radhakrishnan Mahadevan[1]* (rmahadevan@genomatica.com), Maddalena Coppi[2], Daniel Segura[2], Steve Van Dien[1], Bernhard Palsson[1], Christophe Schilling[1], Abraham Esteve-Nunez[2], Mounir Izallalen[2], and **Derek Lovley**[2]

[1]Genomatica, Inc., San Diego, CA and [2]University of Massachusetts, Amherst, MA

The ultimate goal of the *Geobacter* Project is to develop genome-based *in silico* models that can be used not only to interpret environmental gene expression data in environments in which *Geobacteraceae* predominate, but also to predict with *in silico* studies, the outcome of various potential manipulations that might be made to optimize processes, such as *in situ* uranium bioremediation and harvesting electricity from waste organic matter, prior to conducting labor-intensive and expensive field experiments. Although the recently published *in silico* model of *Geobacter sulfurreducens* has been effective in providing explanations for important physiological phenomena and in guiding functional genomic studies, further development is necessary to improve predictions and make them more applicable to other *Geobacter* species.

For example, acetate is the key electron donor for *Geobacter* species during *in situ* uranium bioremediation and in the conversion of organic matter to electricity. Further analysis of the *in silico* metabolic model for *G. sulfurreducens* identified redundant pathways for acetate metabolism. . There are two acetate activation pathways encoded in the genome, the acetate kinase/phosphate transacetylase (Ack/Pta) pathway and the acetyl-CoA transferase (Ato), which plays a dual role in acetate activation and the TCA cycle. There are also, two enzymes catalyzing the synthesis of oxaloacetate, the TCA cycle enzyme, malate dehydrogenase (Mdh) and pyruvate carboxylase (PC), which catalyzes the conversion of pyruvate to oxaloacetate. Three reactions are present for the synthesis of acetyl-CoA from pyruvate: pyruvate dehydrogenase, pyruvate formate lyase and pyruvate ferredoxin oxidoreductase (Por) and three are possible pathways for synthesis of PEP involving pyruvate phosphate dikinase (PpdK), PEP synthase (PpS), and PEP carboxykinase (PpcK).

To evaluate the role of these pathways, five knockout mutant strains lacking elements of the various redundant pathways (Ato, Pta, Por, Mdh, PpcK) were constructed and evaluated along with the wild type for their ability to grow under twelve distinct environmental conditions (72 combinations) and the model predictions were compared to the results of the phenotypic analysis. The model predicted that *G. sulfurreducens* would be able to compensate for the absence of Ato by increasing flux through the Ack/Pta pathway and succinyl-CoA synthetase. However, failure of the Ato-knockout mutant to grow on acetate suggested that the succinyl-CoA synthetase was inactive. Similar constraints on metabolism were derived from the comparison of the *in vivo* phenotypes with the model predictions. Following the incorporation of these new constraints, the *in silico* model now correctly predicts the experimental result in 89% of the possible conditions providing highly accurate characterization of central metabolism in *G. sulfurreducens.*

Despite the augmentation of the constraints using the genetic data described above, the relative flux through the multiple PEP synthesizing pathways could not be resolved by the *in silico* model. In order to further characterize central metabolism and validate the model, carbon isotope (^{13}C) labeling studies were designed using the genome-scale *in silico* model and carried out initially in batch cultures of *G. sulfurreducens*. The *in silico* model was used to determine the optimal acetate labeling strategy for distinguishing the flux through the various phosphoenolpyruvate synthesizing pathways. The

predicted optimal labeling ratio (70% unlabeled and 30% labeled acetate) was utilized for preliminary studies in batch culture and the flux through key reactions in central metabolism was calculated based on the distribution of the label in the amino acids. Further comparison of the experimentally measured flux distribution with model prediction under steady state conditions in a variety of growth media will aid in validating and refining the model.

The *in silico* model has also been used to predict the metabolic adaptation to several environmental perturbations including nitrogen limitation, electron acceptor limitation, and phosphate limitation. These predictions are currently being compared to experimental data, including physiological measurements and global gene expression data, with the goal of improving our understanding of the metabolism of *Geobacteraceae* under non-optimal but environmentally relevant conditions. For example, *in silico* modeling of growth under electron acceptor limiting conditions suggested that either formate or hydrogen production might serve as electron sinks. Although, formate was not detected in the media in response to acceptor limitation, there was increased hydrogen production. Another metabolic adaptation to this condition included an increase in acetate flux through the TCA cycle. Comparison of global gene expression data with model predictions of flux distribution revealed some evidence of increased TCA cycle flux, including upregulation of isocitrate dehydrogenase and downregulation of the phosphotransacetylase. However, there were several exceptions (downregulated acetate transporter, citrate synthase) indicating that flux changes may not be well correlated with transcription profiling data.

Another subject of current investigation is the incorporation of regulatory constraints in the metabolic model, as recent studies in *Escherichia coli* have shown that the addition of such constraints can improve the quality of predictions. The analysis of the existing gene expression data for *G. sulfurreducens* revealed putative regulatory interactions involved in heat shock response, and sulfate metabolism. These initial regulatory interactions will be integrated with the *in silico* metabolic model to design experiments that will optimally perturb the regulatory network. The integrated model will be used to pick the most informative environment changes and the transcription factors for deletion. The gene expression data obtained from these environments will be used to assemble the regulatory network and this process will be repeated to obtain a refined and integrated regulatory and metabolic network model. However, further investigation will be required to extend the results from the *G. sulfurreducens* model to characterize metabolism in other members of *Geobacteraceae*.

Although, recent comparative studies of *Geobacteraceae* genomes have indicated that the electron transport chain components are not fully conserved across the different members of this family, the majority of the elements of central metabolism are conserved among the various family members. Hence, the development of models of metabolism for other *Geobacteraceae* can be accelerated by leveraging the existing curated *G. sulfurreducens* model. An automated modeling pipeline has been redesigned from earlier prototypes to reconstruct the metabolic network for a new organism through a comparison of its genome with the organisms for which a high quality model is available. This has enabled the rapid construction of draft genome-scale models that can be manually curated further to obtain a complete model. We have utilized this pipeline to construct models of other *Geobacteraceae* including *G. metallireducens*, and *Pelobacter carbinolicus* based on the *G. sulfurreducens* model. The initial model of *G. metallireducens* contains about 566 genes, 514 reactions, whereas *P. carbinolicus* model contains 444 genes and 527 reactions. These models are currently being manually curated to ensure that the model can synthesize all essential biomass components. The construction of a comprehensive and physiologically validated *in silico* models of these and other organisms will create a database of metabolic functions that will be valuable for predicting the metabolic capabilities of environmental isolates and optimizing strategies for bioremediation.

It is expected that *Geobacter* species might readily be genetically modified to improve electricity generation because there has been no previous evolutionary pressure to select for this property. Model-

* Presenting author

based analysis suggested that the respiration rate and subsequently, the rate of electron transfer to electrodes could be increased with the introduction of an energy draining futile cycle. An additional ATP consuming reaction was added to *G. sulfurreducens* by introducing the gene for the cytosolic portion of the ATP synthase under the control of an IPTG-inducible promoter. When IPTG was added to culture media, cells had higher respiration rates and the current generation doubled. Electricity production might also be enhanced if *Geobacter* species could utilize more electron-dense fuels. Model simulations indicated that the addition of glycerol transport capability to *G. sulfurreducens* would enable glycerol utilization and this prediction was experimentally confirmed. These studies demonstrate that genome-based *in silico* modeling of microbial physiology can significantly aid in experimental design for strain improvement for practical applications.

92 MEWG

Development of Computational Tools for Analyzing and Redesigning Biological Networks

Priti Pharkya[1], Madhukar Dasika[1], Vinay Satish Kumar[1], Narayanan Veeraghavan[1], Patrick Suthers[1], Anthony Burgard[2], and **Costas D. Maranas**[1]* (costas@psu.edu)

[1]Pennsylvania State University, University Park, PA and [2]Genomatica, Inc., San Diego, CA

The incredible growth in recent years of biological data at all levels has provided a major impetus for developing sophisticated computational approaches for unraveling the underlying complex web of protein, DNA and metabolite interactions that govern the response of cellular systems to intracellular and environmental stimuli. Even partial knowledge of these interconnections and interactions can facilitate the targeted redesign of these systems in response to an overproduction objective. In this poster, we will highlight our progress towards the development of computational frameworks aimed at analyzing and redesigning metabolic and signaling networks.

(1) Metabolic Network Gap Filling: Existing stoichiometric metabolic reconstructions, even for well studied organisms such as *E. coli*, include "unreachable" or blocked reactions due to the inherently incomplete nature of the reconstructed metabolic maps. These blocked reactions cannot carry flux under any uptake conditions. In this project we first identify all such blocked reactions and subsequently pinpoint which reactions to add to the existing model to bridge the maximum number of such gaps. The minimal set that accomplishes this task is chosen from an encompassing list of candidate reactions constructed from databases such as KEGG and Metacyc. The developed framework is demonstrated on genome-scale metabolic models of *Escherichia coli* and *Saccharomyces cerevisiae.* Reactions with higher BLAST scores against the genome of the curated model are preferentially selected. In addition, information as to which metabolites are present (e.g., CE-MS measurements) can be integrated into the gap-filling procedure.

(2) Assessing Objective Functions Driving Metabolic Responses to Perturbations: Genome-scale metabolic reconstructions are increasingly being used to predict the response of metabolic networks to genetic (e.g., gene knock-outs) and/or environmental (e.g., high/low glucose) perturbations. This is accomplished by optimizing an objective function that abstracts the dominant factors driving flux reallocation. These postulated hypotheses include biomass formation maximization, minimization of metabolic adjustment (MOMA)[1] , regulatory on/off minimization (ROOM) [2], etc. In this project, we assess the quantitative performance of these hypothesized objective functions in response to genetic and/or environmental perturbations and propose a new one based on flux ratios rather than

absolute values. A comprehensive comparison using experimental data for wild-type and perturbed networks alludes to the use of composite objective functions as the best predictors.

(3) Elucidating Fluxes in Genome-scale Models Using Isotopomer Labeling Experiments: Isotopic label tracing is a powerful experimental technique that can be combined with the constraint-based modeling framework to quantify metabolic fluxes in underdetermined systems. The calculation of intracellular fluxes by 13C-MFA is based on the fact that when cells are fed a growth substrate with certain carbon positions labeled with 13C, the distribution of this label in the intracellular metabolites can be precisely determined based on the known biochemistry of the participating pathways. Most labeling studies focus on skeletal representations of central metabolism and ignore many flux routes that could contribute to the observed isotopic labeling patterns. In addition, often times a wide range of flux values could explain the experimentally observed labeling patterns in network areas where the experimental measurements provide low resolution. In this work, we investigate the importance of carrying out isotopic labeling studies at the genome-scale. Specifically, we explore how the activity of multiple alternative pathways could in many cases adequately explain the experimentally measured labeling patterns and also suggest methods for improving the resolution of quantified fluxes. Finally, we investigate the effects of introducing global metabolite balances on cofactors such as ATP, NADH, and NADPH as their inclusion in labeling analysis is often neglected but may be important for obtaining biologically realistic flux distributions.

(4) Optimal Redesign: Our research group developed the OptKnock[3] and Optstrain[4] procedures for microbial strain redesign through targeted gene additions and deletions. Both procedures use the maximization of biomass to predict flux reallocations in the face of genetic perturbations. Here we will present how to extend these optimization frameworks to account for popular quadratic objective functions such as MOMA[1] and contrast the obtained results. In addition, we will discuss how to computationally integrate modulations (i.e., up or down gene regulations) in addition to knock-in/outs in the palette of allowed genetic manipulations for microbial strain optimization[5].

(5) Signaling Networks: The same pathway modeling concepts that have been extensively applied to analyze and optimize metabolite flows in metabolic networks can also be used to analyze and redirect information flow in signaling networks. Here we describe optimization-based frameworks for elucidating the input-output structure of signaling networks and for pinpointing targeted disruptions leading to the silencing of undesirable outputs while preserving desirable ones. The frameworks are demonstrated on a large-scale reconstruction of a signaling network composed of nine signaling pathways. Results reveal that there exist two distinct types of outputs in the signaling network that either can be elicited by many different input combinations or are highly specific requiring dedicated inputs. Furthermore, identified targeted disruptions are not always in terminal steps. Many times they are in upstream pathways that indirectly negate the targeted output by propagating their action through the signaling cascade.

References

1. Segre D, Vitkup D, Church GM (2002) "Analysis of optimality in natural and perturbed metabolic networks," *PNAS* 99: 15112-15117.
2. Sholmi T, Berkman O, Ruppin E (2005) "Regulatory on/off minimization of metabolic flux changes after genetic pertubations," *PNAS* 102: 7695-7700.
3. Burgard AP, Pharkya P, Maranas CD (2003) "OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Micorbial Strain Optimization," *Biotechnology and Bioengineering* 84: 647-657.
4. Pharkya P, burgard AP, Maranas CD (2004) "OptStrain: A Computational Framework for Redesign of Microbial Production Systems," *Genome Research* 14: 2367-2376.
5. Pharkya P, Maranas CD "An optimization framework for identifying reaction actvation/inhibition or elimination candidates for overproduction in microbial systems," *Metabolic Engineering* (*In press*).

* Presenting author

93

Next Generation Computational Tools for Biochemical Network Models

Ravishankar Rao Vallabhajosyula, Frank Bergmann, and **Herbert M. Sauro*** (hsauro@kgi.edu)

Keck Graduate Institute, Claremont, CA

Introduction

In this contribution we wish to introduce three developments that will assist in the simulation of large scale biochemical networks that are of interest to the GTL community. Firstly we have developed a new algorithm to compute the conserved moieties of large biochemical networks. Secondly we have developed a new PC based high speed simulator that is at least an order of magnitude faster than current simulators. Finally we wish to describe a new real-time visualization tool that permits users to view simulations of biochemical systems in a more natural and meaningful way.

Conservation Analysis of Large Biochemical Networks

Conservation Analysis of biochemical models is an important step in determining the dependent and independent species in a biochemical network leading to a dimensional reduction of the model. This is a numerically intensive task that becomes error-prone with existing methods for large networks. The computation of the correct conserved cycles is also essential for the computation of the reduced Jacobian, which is non-singular. This reduced Jacobian is very important for a number of subsequent analyzes such as Bifurcation Analysis[1] and Metabolic Control Analysis[2].

Our new method makes uses of the Householder QR factorization of the stoichiometric matrix to obtain the relevant conservation relations for biochemical networks. Its advantage lies in much greater enhanced reliability and accuracy over other methods currently in use. The underlying algorithm is described in detail in our recent paper[4]. It has been integrated into Systems Biology Workbench (SBW)[3] and accepts models in standard SBML. SBW integration allows other developers to easily access the capability of this method. A separate graphical interface for this tool is also provided (See Figure 1).

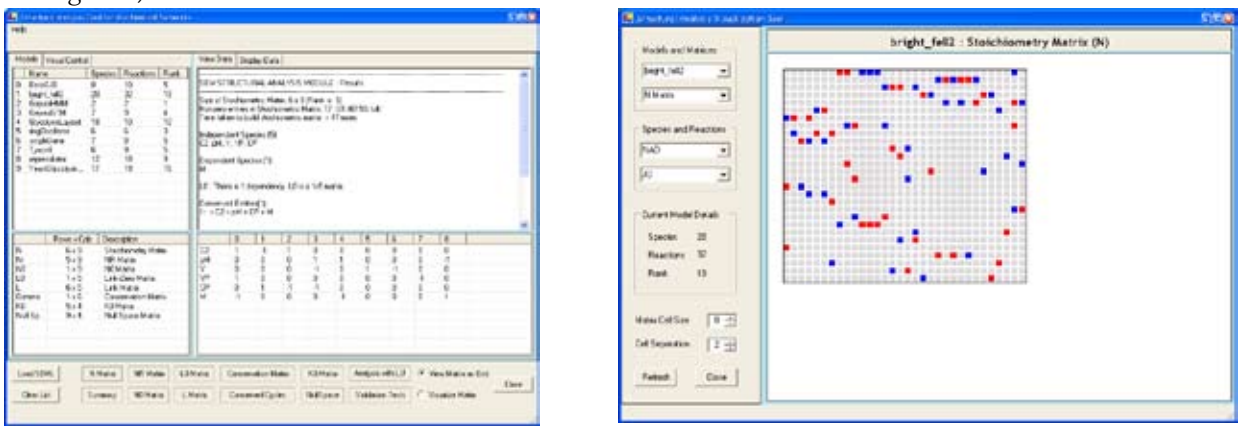


Figure 1: A Graphical tool for conservation analysis of biochemical network models

* Presenting author

High-Performance Simulation: roadRunner simulator

One of our aims is to create a rich interactive user experience for the modeling of biochemical systems. A critical requirement for this is the availability of high performance simulation software. The bulk of existing simulation tools rely on an interpretative method for evaluating the model equations which in turn is slow and inefficient. Instead we have been experimenting with just-in-time compilation of models using Java and .NET. A new high-performance simulator, codenamed 'roadRunner', drives this effort. This simulator uses the conservation analysis via SBW to construct the reduced model. It also implements most of the features specified in SBML level 2, including discrete events and user defined functions. This simulator has been developed using C# and relies on Sundials CVODE and NLEQ. The powerful reflection features of C# together with on-the-fly code generation allow 'roadRunner' to outperform a Java implementation. Initial tests also show an order of magnitude increase in performance over existing tools such as Jarnac and SBML odesolver. In addition to simulation, roadrunner also implements many other analyzes such as metabolic control analysis, model fitting and frequency analysis.

3D Time-Course Visualization of Simulations

Modeling and analysis of reaction-networks relies heavily on simulation tools. Traditionally the simulation results are available either as data tables or X-Y plots. Data tables are helpful for further processing by other computational tools. X-Y plots however tend to get complex even for a limited number of species. In creating a new visualization tool we had two goals in mind. The first goal was to tie the simulation results closely to the model and the other was to enable the user to view the simulation in real time.

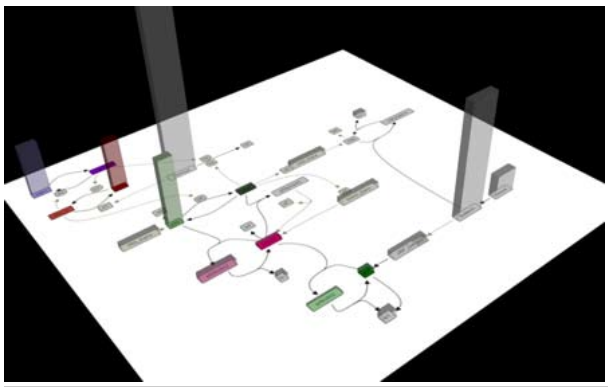


Figure 2: The 3D visualization tool showing simulation data in real-time

SBML is an established standard for the exchange of model-information. Recently the SBML group adopted a layout standard. This means it is now possible to store information about species, compartments and positions or dimensions of the model. This has allowed us to make the layout of a model portable. This layout implements the basis of a 3D visualization tool and is projected onto a 3D plane. Furthermore all positions of species are recognized and their concentrations are rendered as columns on top of the 3D plane. The heights of these columns represent the current species concentration values that vary during the time-course simulation. Since the time course simulation is performed continuously it is possible to dynamically focus on specific aspects of the model (see Figure 2).

The simulator and the visualization tool that implements the conservation analysis described earlier will be showcased at the DOE GTL meeting, and the results presented in a poster. A test version of the 3D Time-course Visualization can be found under: http://public.kgi.edu/~fbergman/Simulate3D.htm. The software for the latest version of Systems Biology Workbench is freely available at http://www.sys-bio.org.

References

1. V. Chickarmane, S.R. Paladugu, F. Bergmann and H.M. Sauro. "Bifurcation discovery tool," *Bioinformatics*, 21(18), 3688-3690, (2005).

* Presenting author

2. J-H.S. Hofmeyr. Metabolic control analysis in a nutshell, *Proc. Intl. Conf. Systems Biology*, Pasadena, California, 291-300, (2000).
3. H.M. Sauro, M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, and H. Kitano. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* Winter, 7(4), 355-372, (2003).
4. R.R. Vallabhajosyula, V. Chickarmane, and H.M. Sauro. Conservation analysis of large biochemical networks. *Bioinformatics Advance Access* published on November 29, 2005, DOI 10.1093/bioinformatics/bti800.

94

Cyclic AMP-Dependent Regulatory Networks of *Shewanella oneidensis* MR-1 Involved in Anaerobic Energy Metabolism

Alex S. Beliaev[1]* (alex.beliaev@pnl.gov), Daad A. Saffarini[2], Lee Ann McCue[1], Amoolya H. Singh[3], Yang Zhang[1], Matthew J. Marshall[1], Grigoriy E. Pinchuk[1], and Jim K. Fredrickson[1]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]University of Wisconsin, Milwaukee, WI; and [3]University of California, Berkeley, CA

Shewanella oneidensis MR-1 is a facultative metal-reducing bacterium with extensive respiratory versatility. Unlike many bacteria studied to date, the ability of *S. oneidensis* to grow anaerobically with several electron acceptors is regulated by the cAMP-receptor protein (CRP). CRP-deficient mutants of MR-1 are impaired in anaerobic reduction and growth with Fe(III), Mn(IV), fumarate, nitrate, and DMSO. Loss of anaerobic respiration in Crp^- mutants is due to loss of terminal anaerobic reductases and not due to deficiency in carbon metabolism. To further elucidate the role of CRP and to understand the mechanisms of cAMP-dependent gene expression under anaerobic conditions in *S. oneidensis* MR-1, a combination of experimental and computational approaches have been applied.

To study the evolution of Crp in *S. oneidensis* MR-1 and its functional divergence from other closely related γ-*Proteobacteria*, we examined the conservation of the *crp* coding region as well as its upstream promoter region. Early results show that the coding region is conserved to almost 100% identity in *S. oneidensis* MR-1 and five other species of *Shewanella* (*S. amazonensis* SB2B, *S. baltica* OS155, *S. denitrificans* OS-217, S. *frigidimarina* NCIMB 400, and *Shewanella* sp. PV-4), and conserved to 95-97% identity with several closely related γ -*Proteobacteria* (*Escherichia*, *Salmonella*, *Shigella*, *Yersinia*, and *Vibrio*). Outside of this clade, *crp* appears to have diverged significantly in sequence since the last common ancestor. The ratio of synonymous to non-synonymous nucleotide substitutions (dN/dS, or K_a/K_s) of *crp* in these same species indicates that the gene is under strong selective pressure not to undergo mutation.

Microarray analyses mRNA expression profiles of wild-type and crp mutant cells grown anaerobically with different electron acceptors indicated that CRP positively regulates the expression of genes involved in energy generation and transcriptional regulation. These include the periplasmic nitrate reductase (*napBHG*), the polysulfide reductase (*psrAB*), anaerobic DMSO reductase (*dmsAB*) genes, as well as the nitrate/nitrite sensor protein *narQ*. Mobility shift assays using purified CRP suggest that this protein activates gene expression directly by binding to promoter regions of anaerobic reductase genes. Furthermore, our experiments indicate that cAMP is required for CRP activation. Mobility shifts were observed only when cAMP was added to CRP-DNA reaction mix and cAMP

addition to aerobically growing *S. oneidensis* cells resulted in the induction of fumarate and Fe(III) reductase activities.

The genome sequence of *S. oneidensis* MR-1 contains three putative adenylate cyclase genes, designated *cyaA*, *cyaB*, and *cyaC*. Deletions of both *cyaA* and *cyaC* resulted in anaerobic growth deficiency with DMSO, nitrate, Fe(III), Mn(IV), and fumarate. These phenotypes are similar to the phenotypes of the CRP-deficient mutants. The function of both *CyaA* and *CyaC* as adenylate cyclases was confirmed by complementing an *E. coli* cyaA mutant. It is interesting to note that *CyaC* contains a predicted membrane-bound domain, similar to eukaryotic adenylate cyclases that are involved in signaling. One hypothesis to be tested is that the membrane domain of *CyaC* is involved in oxygen sensing, and therefore cAMP synthesis by this protein occurs under anaerobic conditions. *CyaC* consists of a membrane domain and a catalytic domain that is predicted to reside in the cytoplasm. Surprisingly, deletion of the *CyaC* membrane domain leads to loss of enzyme activity, suggesting that it may play a role in the stability or activation of the catalytic domain. Further work to identify the cAMP/CRP -dependent regulatory networks in *S. oneidensis* MR-1 is underway.

95 MEWG

Engineering *E. coli* to Maximize the Flux of Reducing Equivalents Available for NAD(P)H-Dependent Transformations

Patrick C. Cirino* (cirino@engr.psu.edu), Costas D. Maranas, and Jonathan W. Chin

Pennsylvania State University, University Park, PA

Biocatalysis offers the opportunity for unmatched reaction specificity and product diversity, and is integral to realizing a future of cost-effective "green" chemistry. The ever-growing ease with which we are able to manipulate cellular metabolism and to design enzymes with specified properties presents the opportunity to develop biocatalytic systems of increased complexity and efficiency. Transformations of primary importance are those catalyzed by reduced nicotinamide cofactor-dependent reductases, dehydrogenases and oxygenases. While significant improvements have been made in the biocatalytic properties of these enzymes, many issues remain unresolved regarding the preferred approach for implementing their transformations and the accompanying cofactor regeneration requirement. The primary objective of this research is to develop microbial production strains which will serve to host NAD(P)H-dependent, heterologous reactions with maximized efficiency in the generation and subsequent utilization of reduced cofactors derived from glucose or other renewable energy sources. *E. coli* is the microbial host chosen for these studies, and we are initially studying the reduction of xylose to xylitol (by heterologous xylose reductases with different cofactor preferences) to serve as an experimental platform that allows us to systematically characterize the individual and synergistic influences of select genetic modifications on strain performance, measured as the yield on xylitol produced (from xylose reduction) per glucose consumed as co-substrate. Maximizing this yield translates to uncoupling carbon metabolism from respiration or fermentation and effectively "respiring" on xylose.

In addition to studying metabolic parameters that are *expected* to impact our objectives (e.g., fermentative pathways, global regulators, transhydrogenase function), stoichiometric network analysis and the strain optimization framework OptKnock are being used to understand the influence of enzymes with potentially critical or ambiguous physiological functions on theoretical yields and to suggest knockout strategies that will constrain the network such that cell growth is coupled to

* Presenting author

xylitol production. For example, xylose transport via the ATP-dependent transporter was shown to significantly reduce xylitol yields compared to xylose transport via proton symport or facilitated diffusion. Furthermore, the physiological roles of the two native transhydrogenases (i.e., whether they are reversible) critically influence flux distributions and yields. Experimental results for wild-type and deletion strains are being used to better understand network architecture, supplement existing metabolic models, and improve prediction accuracy and fidelity.

Maximizing xylitol production inherently leads to very low growth rates because ATP yields must be low if reducing equivalents are to be directed towards xylose reduction. Consequently, deletion of ATP synthase is chosen by Optknock to be a key genetic modification required for coupling xylitol production to biomass formation. Deletion of *atpA* increased the xylitol yield by ~40% in shake-flask cultures (normalized to cell density). In order to more accurately reflect conditions of low growth, we are also using metabolically active but non-growing "resting cells" to evaluate strain performance. Use of resting cells additionally eliminates growth as a variable that alters partitioning of carbon and reducing equivalents, and provides a more reliable method for determining maximum experimental yield. The yield on xylitol per glucose consumed in our "base" strain expressing an NADPH-dependent xylose reductase improved from 1.5 in batch culture to 4.3 under non-growing conditions. Measuring these yield values from knockout strains enables us to illuminate the contributions of various pathways and reactions to NADPH-dependent xylose reduction, characterize the effects of overexpression of an NADPH sink (a common scenario in whole-cell biocatalysis) on partitioning of co-substrate (i.e., glucose) carbon and reducing equivalents, and identify combinations of genetic modifications that lead to improved strain efficiency and productivity.

96

A New MILP Based Approach for *in Silico* Reconstruction of Metabolic Networks and Its Application to Marine Cyanobacterium *Prochlorococcus*

Xiaoxia (Nina) Lin* (xiaoxia@genetics.med.harvard.edu), Aaron Brandes, Jeremy Zucker, and **George M. Church**

Harvard Medical School, Boston, MA

http://arep.med.harvard.edu/DOEGTL/

The increasing availability of annotated genomes has rendered the possibility of applying systems approaches, e.g. flux balance analysis (FBA), to the study of a large variety of organisms. To achieve this high-throughput goal, we have been developing an automatic bioinformatics pipeline to facilitate the process of generating *in silico* whole-cell metabolic models from genome annotations. In this work, we present a new computational framework for the key step in this pipeline which constructs metabolic networks by integrating genome annotation, reaction database, and phylogenetic information.

Our goal is to construct metabolic pathways/networks for a new species based on its genome annotation and a multiple-species pathway/reaction database (e.g. BioCyc databases). Using a mixed-integer linear programming (MILP) optimization framework, the new algorithm selects a set of reactions from a universal super-network which can achieve the functionality of a pathway or network to convert specific metabolites or to enable the cell to live and grow. The solution includes not only reactions already identified in the genome annotation but also additional ones required to achieve the

functionality which are the most possible phylogenetically. Alternative and/or sub-optimal solutions can also be systematically generated to increase the likelihood of identifying the real biological network. In addition, quantitative data such as nutrient condition can be readily incorporated to improve the predictions.

The above approach has been applied to the study of a marine cyanobacterium *Prochlorococcus marinus*, which dominates the phytoplankton in the tropical and subtropical oceans and contributes to a significant fraction of the global photosynthesis. As a proof-of-concept, the algorithm automatically generated novel TCA pathways which do not exist in the pathway databases and are consistent with partial knowledge of cyanobacteria. We have also successfully reconstructed the networks for central carbon metabolism, amino acid biosynthesis, and nucleotide biosynthesis. We are currently moving towards the whole-genome metabolic network of *Prochlorococcus* as well as using the results from this approach to identify missing genes/enzymes and to refine the genome annotation.

97 MEWG

Optimizing Central Metabolic Pathways in Yeast

Thomas W. Jeffries[1,2]* (twjeffri@wisc.edu), Chenfeng Lu[1], Karen J. Mansoorabadi[1], Jennifer R. Headman[1], Ju-Yun Bae[1], and Haiying Ni[3]

[1]University of Wisconsin, Madison, WI; [2]USDA, Forest Products Laboratory, Madison, WI; and [3]University of Chicago, Chicago, IL

Optimization of metabolite flux relies primarily on deletion, overexpression, or regulated expression of one or more target genes with the assumption that these changes will be reflected in altered levels of enzymatic activities. Predictive models are essential, but in practice, genetic engineering must be carried out with empirical trials and concomitant modification of environmental variables. Targeted changes do not always have the desired effects.

Our research has concentrated on engineering xylose metabolism in yeasts. We explore the metabolism of *Pichia stipitis*, which is capable of converting xylose to ethanol, and we modify *Saccharomyces cerevisiae*, which is not. The two approaches are complementary. No one feature of *P. stipitis* has as yet proven to be the "magic key" that enables xylose fermentation. *S. cerevisiae* possesses orthologs of all the known steps for xylose assimilation (*GRE3, SOR1*, *XKS1*), but does not naturally coordinate their expression in a way that permits the conversion of xylose to ethanol. Overexpression of the corresponding *P. stipitis* genes (*XYL1, XYL2, XYL3*) can improve both assimilation and ethanol production in *S. cerevisiae*, but elevated expression of *XYL3* (or *XKS1*) can inhibit growth on xylose unless accompanied by the overexpression of either Ps or Sc*TAL1* (transaldolase).

We have conducted a systematic search for genes that will enhance growth on xylose when deleted or overexpressed. One unpredicted finding was *PHO13*, a *p*-nitrophenyl phosphatase that shows some specificity for histone dephosphorylation. Deletion of this gene relieves growth inhibition on xylose and increases *ScTAL1* expression approximately two-fold. Our overexpression studies have identified three additional genes that appear to relieve inhibition when expressed along with *XYL1, XYL2* and *XYL3*.

Coordinated expression – even at low levels – appears to be important, so we have developed a series of *S. cerevisiae* promoters that enable regulated expression over a dynamic range of about 80-fold. We have also developed methods that enable the relatively rapid switching of promoter/gene pairs to

* Presenting author

screen the effects of regulated expression with multiple genes simultaneously. The effects of multi-gene expression depend greatly on the environmental conditions and the physiological state of the cells such that one combination of promoters can be appropriate for one condition (e.g. high aeration), but less desirable for another (e.g. limited oxygen). Such effects are expected to extend to other regulatory signals as well.

As reported previously at this conference, heterologous expression in plants can be affected by gene orientation. We have explored this phenomenon and have concluded that gene orientation and the repeated use of a single promoter (promoter dilution) does not seem to have a significant effect in *S. cerevisiae*.

The genetic background can greatly affect results. As mentioned, mutations can be introduced – or they can arise spontaneously. To further explore this aspect, our research has examined the effects of *XYL1, XYL2,* and *XYL3* overexpression in several laboratory and industrial yeast strains.

98 MEWG

Multi-Scale Models for Gene Network Engineering

Yiannis N. Kaznessis* (yiannis@cems.umn.edu)

University of Minnesota, Minneapolis, MN

Armed with increasingly fast supercomputers and greater knowledge of the molecular mechanisms of gene expression, it is now practical to numerically simulate complex networks of regulated biological reactions, or gene circuits. It is also becoming feasible to calculate the free energy of noncovalent binding of regulatory proteins to specific DNA target sites. Using a hybrid stochastic-discrete and stochastic-continuous simulation algorithm, we obtain an accurate time-evolution of the behavior of complex gene circuits, including a clear picture on the role of highly dilute, but significant, regulatory proteins. These regulatory proteins are responsible for the non-linear control used by biological organisms to regulate their most important processes. The network simulations provide insight, which can guide rational engineering of regulatory proteins and DNA operator sequences using molecular mechanics simulations. In this presentation we examine two important gene circuits, the bistable switch and the oscillator. We study the role of specific biomolecular interaction phenomena on the dynamics of these gene circuits. Using models that span multiple time and space scales, from atomistic, to molecular, to interaction networks we develop design principles for high quality bistable switch and oscillator circuits.

99

Generalized Computer Models of Chemoheterotrophic Bacteria: A Foundation for Building Genome-Specific and Chemically-Detailed Bacterial Models

Jordan C. Atlas, Evgeni V. Nikolaev, and **Michael L. Shuler*** (mls50@cornell.edu)

Cornell University, Ithaca, NY

A significant challenge in systems biology is to better understand fundamental design principals of cellular organization and function by taking advantage of information encoded in annotated DNA sequences. While bioinformatics tools and related technology will continue to dominate within the next decade, these efforts are, by themselves, insufficient. New tools are necessary to explicitly relate genomic and molecular information to cellular physiology and population response. The release of the 1000th microbial genome is expected within next few years (Overbeek *et al.,* 2005), and this further accelerates the development of technology to build accurate metabolic reconstructions and balanced stoichiometric models of bacterial cells. Stoichiometry correctly defines overall barriers for intracellular steady-state fluxes under fixed 'defined medium' constraints, and genome-scale stoichiometric models have been very successful in many instances in fundamental and applied research (Reed and Palsson, 2003). However, the predictive capability of static stoichiometric models is limited to the calculation of 'instant snapshots' of different phenotypes and therefore such models cannot capture dynamic changes in protein machinery, metabolite concentrations, and cell geometry. The advent of abundant data coupled with the limitations of current modeling approaches necessitates the development of novel modeling frameworks to rapidly build completely functional bacterial cell models.

The availability of complete metabolic reconstructions for a variety of bacterial species can facilitate the development of molecularly-detailed bacterial subsystems models, named here 'modules.' Such *chemically detailed* modules can then be combined within a *coarse-grained* model to produce a completely functional *hybrid* single cell model (Castellanos *et al.,* 2004). The initial step of our 'hybrid model approach' was to construct a coarse-grained model with lumped 'pseudochemical species.' The coarse-grained model explicitly links DNA replication to metabolism, cell cycle, cell geometry and external environment (Browning, *et al.*, 2004). Such models can include known or putative regulation, relations capturing key events in the production and utilization of cell's energy and redox equivalents, RNA transcripts, proteins, lipids, and different forms of nucleotides. These coarse-grained models correctly predict sustained bacterial reproduction (*i.e.*, chromosome replication and cell division following one another in the right order and timing), which can be viewed as an *obligatory* test for all whole cell models. Modest-sized coarse-grained models allow for the application of rigorous mathematical analyses such as bifurcation and stability analyses to verify the model's robustness.

This hybrid-modular framework has been successfully applied to create Cornell's Minimal Cell Model (MCM). A 'minimal cell' is a hypothetical free living organism possessing the functions required for sustained reproduction in a maximally supportive culture environment. The 'modularity' has been demonstrated by constructing a genomically and chemically detailed model of nucleotide metabolism within the MCM (Castellanos *et al.,* 2004), utilizing statistical mechanics methods for parameter estimation (Brown and Sethna, 2003). We are currently in the process of incorporating detailed genome-specific information into the *E. coli* model. Our focus will be on the core carbon metabolic subsystems such as energy metabolism, nucleotide biosynthesis, and biomass precursor formation. The carbon metabolism provides 12 key precursors for all biochemicals formed within

* Presenting author

the bacterium and allocates larger fluxes of all intracellular fluxes unevenly distributed throughout metabolism. The 'core' hybrid model will allow us to gain fundamental insight into how the dynamic events in the central carbon metabolism are controlled by changes in the chromosome and external environment. The core model will also include an improved model of chromosome replication and lumped modules for subsystems with diminished fluxes. This modeling framework will serve as a platform for the development of a variety of bacterial cell models. Specifically, the approach will be applied to the development of a functionally complete model of *Shewanella oneidensis*, a microorganism closely related to *E. coli*. The constructed models will be publicly available as Matlab modules and via the System Biology Markup Language (SBML) format. The current project is conducted in the cooperation with Gene Network Science, Inc., and utilizes the GNS VisualCell™ modeling platform.

References

1. Brown, K.S., and J.P. Sethna. 2003. "Statistical Mechanics Approaches to Models with Many Poorly Known Parameters." *Phys. Rev. E* 68, 021904.
2. Browning, S.T., M. Castellanos, and M.L. Shuler. 2004. "Robust Control of Initiation of Prokaryotic Chromosome Replication: Essential Considerations for a Minimal Cell." *Biotechnol. Bioeng.* 88(5):575-584.
3. Castellanos, M., D.B. Wilson, and M.L. Shuler. 2004. "A Modular Minimal Cell Model: Purine and Pyramidine Transport and Metabolism." Proc. Natl. Acad. Sci. (USA). *PNAS* 101(17): 6681-6686.
4. Overbeek, R., T.Begley *et al* 2005. "The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes," *Nucleic Acids Res*. 33(17): 5691-5702
5. Reed, J.L. and Palsson, B.Ø. 2003. "Thirteen Years of Building Constraint-Based *in silico* Models of *Escherichia coli*," *Journal of Bacteriology*, 185(9): 2692-2699.

100

Determination of the Most Probable Objective Function for Flux Analysis of Metabolism Using Bayesian-Based Model Discrimination

Andrea Knorr and **Ranjan Srivastava*** (srivasta@engr.uconn.edu)

University of Connecticut, Storrs, CT

Metabolic flux analysis has proven to be a powerful tool for analyzing, understanding, and engineering a variety of different organisms. However, when carrying out flux analysis, the metabolic reaction network is often underdetermined due to a lack of information. Under such circumstances, optimization theory may be used to determine the fluxes across the metabolic pathways. To carry out the optimization process, an appropriate objective function must be identified. Ideally the objective function should represent a biological process which, when optimized, would prove more beneficial for the organism than the optimization of any other process. Several such objective functions have been proposed, including maximization of growth and optimization of energy efficiency. In this work, we have adapted a Bayesian-based model discrimination method to allow us to determine which objective function is most probable for use with flux analysis.

The model system analyzed was the central metabolism of *E. coli* growing on succinate. Based on biological plausibility, two objective functions were compared. The first was maximization of growth, while the second was minimization of redox potential. Flux balance analysis and linear programming coupled with experimental data were used to predict metabolic fluxes using each objec-

tive function. Our adapted Bayesian model discrimination approach was then employed to determine which of the two objective functions was more likely. Growth maximization was shown to be the most probable objective function for use with flux analysis to determine metabolite distribution. It should be noted that the technique employed here is generic enough to be used to determine the most probable objective function for the metabolic analysis of any organism.

101

In silico Analysis of *Escherichia coli* Metabolism to Optimize Electron Production

David Byrne* (dbyrne@bu.edu), Daniel Segre, and **Timothy Gardner**

Boston University, Boston, MA

In metabolic engineering, it is desirable to manipulate microbes to overproduce specific by-products. In order to attain this bioengineering objective it would be ideal to be able to use informed computational predictions to effectively direct subsequent experimental microbial biotransformations. We are specifically interested in developing methods to predict environmental conditions or mutant microbial strains that optimize the electron production for potential application in microbial fuel cells and bioremediation.

To achieve this objective, the methods that we are developing incorporate simulations of genome-scale metabolic models, initially that of the *Escherichia coli* K-12 MG1655 consisting of 931 unique reactions and 625 metabolites. An algorithm searches these simulations for optimal environmental conditions and genetic modulations that maximize electron yield while minimizing competing cellular processes.

This problem may be mathematically posed as a bi-level optimization problem such that the inner optimization problem performs a linear programming optimization for maximum biomass yield based on stoichiometric and nutrient constraints while the outer problem maximizes electron production by modulating specific reactions available to the inner problem. A computational optimum may then be attained by simultaneously solving both the inner and outer problems using multiple-integer linear programming.

This computational framework may be used to address the following questions of biological significance: How does electron production relate to energy (e.g. ATP synthesis) and growth yield? How does electron production relate to changes in a microbe's environment? How does it relate to the general conditions requiring aerobic respiration, anaerobic respiration, or fermentation? How does electron production relate to specific changes in carbon, nitrogen, electron-acceptor, or other media substrate sources? How does electron production change when different types of genetic perturbations are imposed?

* Presenting author

102

Reverse-Engineering the Central-Metabolism Network of *Shewanella oneidensis* MR-1

J.F. Penders and **T.S. Gardner*** (tgardner@bu.edu)

Boston University, Boston, MA

With the recent availability of genome annotations for many microbial organisms, metabolic engineering has seen the development of genome-scale metabolic models. This new engineering tool seems promising for optimizing microbes for bioremediation and microbial fuel cell applications. Towards this metabolic engineering objective, the process of developing the most accurate possible model is essential. Current methods are based on intuition and consist in adding or removing reactions to improve model agreement with experimental data.

Shewanella exhibits an uncommon and complex metabolism. A facultative anaerobe, it has versatile respiratory pathways, and is able to use a broad spectrum of electron acceptors, including metals such as iron, manganese and uranium. In preliminary work, we have applied current methods to develop a metabolic model for the central metabolism of *Shewanella oneidensis* MR-1. We created a Pathway-Genome Database, containing 206 pathways, 1015 reactions and 807 enzymes using existing genome annotations coupled with manual curation. We used this metabolic map to build a model for the central-metabolism of *Shewanella*, but our model predictions diverged from known experimental observations kindly shared with us by the *Shewanella* Federation.

Shewanella's complex respiratory metabolism thus complicates intuitive approaches about which reactions should be added to or removed from the model. Therefore, although current methods based on manual tuning are efficient for most organisms, they risk becoming cumbersome and are inappropriate for organisms exhibiting non-intuitive metabolism. In such cases, an automatic way to learn the metabolic network from experimental data is needed.

We have thus developed a novel computational method tool to reverse-engineer metabolic networks in *Shewanella.* Metabolic networks have unique properties such as high connectivity of certain nodes (such as co-factors) and non-pairwise connectivity of edges, and they require multiple-edge perturbations for effective network inference approaches. Our algorithm semi-greedily searches all biochemically feasible metabolic networks, and learns an optimal network by iterative comparison of experimental data with computed metabolite producibility and flux balance analysis.

We hope that the application of our novel method to *Shewanella* will aid investigators in unraveling its central metabolism and provide new insights into its extraordinary reducing capabilities.

103

Genome-Scale Metabolic Model of *Shewanella oneidensis* MR1

Brian Gates[1]* (bgates@genomatica.com), Grigoriy E. Pinchuk[2], Christophe Schilling[1], and Jim K. Fredrickson[2]

[1]Genomatica, Inc., San Diego, CA and [2]Pacific Northwest National Laboratory, Richland, WA

A genome-scale metabolic model of *Shewanella oneidensis* MR1 was created using the constraints-based approach. The development of the model was accelerated by an early stage prototype of an automated network reconstruction process that leveraged information contained in high-quality models of other organisms to predict a large percentage of the *Shewanella* metabolic network based on gene sequence homology. Subsequent manual network reconstruction efforts proceeded through review of the genome annotation and testing for the presence of certain essential pathways. The model was then refined in an iterative procedure in which model predictions were compared to experimental data.

The current model version includes 742 of the 5102 genes in *S. oneidensis*, which are translated to 626 proteins that catalyze 702 model reactions. An additional 33 reactions are present in the model without genetic support, referred to as non-gene associated reactions. An aggregate reaction for biomass formation was developed that included as substrates the major molecular components of biomass in the stoichiometry necessary to produce 1 g dry cell weight as well as sufficient ATP to meet the energetic demands of growth. Because detailed composition data for MR1 is not yet available, all simulations were performed using the biomass composition of *E. coli*. The model is currently able to simulate several aspects of metabolism in *S. oneidensis*, including substrate utilization profiles, byproduct secretion under certain conditions, and the use of several terminal electron acceptors. Consistent with experimental observations, the model predicts that growth on lactate under oxygen limited conditions will produce near-equimolar acetate secretion and no formate production. The model also predicts the known ability of MR1 to grow using iron, fumarate, or nitrate as a terminal electron acceptor.

As part of the on going efforts to continue refining the model to make it more consistent with the known physiology of the organism, we analyzed phenotypic respiration data generated using the Biolog Phenotype MicroArray™ platform for *S. oneidensis*. The array data was used to test the model's ability to predict aerobic respiration on 139 different carbon sources. Simulation matched experimental predictions for 29% (7/24) of the compounds that produced respiration, and 84% (97/115) of the compounds that did not. Further review of the genome led to the assignment of 6 additional metabolic reactions to genes, improving the respiration prediction rate to 42% (10/24).

The metabolic network of the *Shewanella* model was compared to the network of the *E. coli* genome scale model, the closest homolog used in the initial automated reconstruction. Genome-scale deletion analyses were performed with both models to determine essential genes for aerobic growth on lactate. 171 genes linked to 219 reactions were found to be essential in MR1, while 176 genes linked to 216 reactions were found in *E. coli* under the same conditions. However, there were 14 reactions that were essential only in MR1, and 10 in *E. coli*. Amino acid metabolism had more unique genes in *E. coli* than in MR1, due to the incorporation of serine glyoxylate aminotransferase as an alternate pathway for serine utilization in MR1.

All of the model development and simulation was enabled by the SimPheny™ software platform. The model and the SimPheny server dedicated to *S. oneidensis* is now being accessed by multiple

* Presenting author

distributed research groups in the *Shewanella* Federation as part of a beta-test for coordinated high quality remote access to SimPheny. Through the availability of a first version model and remote access to SimPheny we are beginning to enable model-driven discovery research for *S. oneidensis* in a collaborative, multi-institutional research setting. We expect that these efforts will accelerate the pace of discovery and our overall understanding of metabolic physiology in *Shewanella* species.

104

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

Sean Conlan[1]* (sconlan@wadsworth.org), William Thompson[4], Lee Newberg[1,2], Lee Ann McCue[3], and **Charles Lawrence**[4]

[1]The Wadsworth Center, Albany, NY; [2]Rensselaer Polytechnic Institute, Troy, NY; [3]Pacific Northwest National Laboratory, Richland, WA; and [4]Brown University, Providence, RI

A thorough knowledge of an organism's transcription regulatory network is a critical component of understanding the biology of the organism as a whole. The foundation of a prokaryotic regulatory network is the *cis* (transcription factor binding sites) and trans (transcription factors) elements which constitute the molecular wiring diagram. Comparative genomics, which makes inferences about the properties of an organism through analysis of related genomes, offers a powerful set of tools for deciphering this network. The field of comparative genomics, particularly with regards to microbial genomics, is entering an unprecedented time, with high-throughput sequencing facilities enabling the sequencing of many closely related bacterial strains and isolates. By comparing closely-related genomes, there is the potential to discover the minor sequence changes responsible for important phenotypic variations observed between related bacterial strains. Using closely related strains for comparative studies of transcription regulation is attractive because these species are most likely to share common transcription factors, and therefore common *cis*-regulatory elements. Unfortunately, the recent speciation of closely related genomes results in correlation among the sequences that complicates the detection of functionally conserved motifs. To facilitate high-resolution comparative genomics studies, that are able to leverage the power of both closely- and distantly- related genomes, we have developed a phylogenetically-rigorous Gibbs recursive sampler, orthoGibbs. Phylogeny is incorporated through the use of an evolutionary model and sequence weights. OrthoGibbs was used to detect known transcription factor binding sites upstream of a study set of genes from *Escherichia coli* and 7 other gamma-proteobacterial genomes. We were able to demonstrate improved specificity and positive predictive value for orthoGibbs when compared to the non-phylogenetic Gibbs sampler. In addition, orthoGibbs identified sites known to be bound by Fur, upstream of iron-regulated genes in 11 sequenced *Shewanella* species.

105

Integrated Computational and Experimental Approaches to Facilitated Model Development

R. Mahadevan* (rmahadevan@genomatica.com), T. Fahland, Y. Kwan, J.D. Trawick, I. Famili, and C.H. Schilling

Genomatica, Inc., San Diego, CA

Advances in high-throughput technologies have lead to the sequencing of several microbial genomes with important applications in bioremediation, CO_2 sequestration, alternative energy sources and industrial biotechnology. The availability of such sequence information has enabled the development of metabolic models, which are genomically and biochemically structured databases and are valuable for data interpretation and computational analysis. The models will also enable the rational design of strategies for improving the efficiency of bioremediation and CO_2 sequestration and other processes related to the Department of Energy's core missions. Genome-scale constraint-based modeling has been shown to be successful in predicting physiology under varied conditions. The manual development of such genome-scale models is labor intensive and time consuming. Thus, there is a critical need to develop approaches for the rapid development of systems level cellular models. In this project, we have developed an automated approach to develop genome scale models based on a combination of genome sequence and high-throughput phenotyping, that can be used to deliver a model-driven approach to biological discovery in the newly sequenced microorganisms. The initial phase of this project focuses on the design and prototype development of approaches for automated metabolic reconstruction based on both the genome sequence and through comparison with the existing high quality metabolic models in our database. We are also developing methods for automatically identifying candidate reaction sets to close gaps in the metabolic pathways associated with the synthesis of essential biomass components utilizing high-throughput experimental data from growth phenotyping experiments. Initial results from the proposed approaches will be shown for a newly developed genome-scale model of *Bacillus subtilis*.

106

New Technologies for Metabolomics

Jay D. Keasling* (jdkeasling@lbl.gov), Carolyn Bertozzi, Julie Leary, Michael Marletta, and David Wemmer

Lawrence Berkeley National Laboratory, Berkeley, CA

Microorganisms have evolved complex metabolic pathways that enable them to mobilize nutrients from their local environment and detoxify those substances that are detrimental to their survival. Metals and actinides, both of which are toxic to microorganisms and are frequent contaminants at a number of DOE sites, can be immobilized and therefore detoxified by precipitation with cellular metabolites or by reduction using cellular respiration, both of which are highly dependent on cellular metabolism. Improvements in metal/actinide precipitation or reduction require a thorough understanding of cellular metabolism to identify limitations in metabolic pathways. Since the locations of bottlenecks in metabolism may not be intuitively evident, it is important to have as

* Presenting author

complete a survey of cellular metabolism as possible. Unlike recent developments in transcript and protein profiling, there are no methods widely available to survey large numbers of cellular metabolites and their turnover rates simultaneously. The system-wide analysis of an organism's metabolite profile, also known as "metabolomics", is therefore an important goal for understanding how organisms respond to environmental stress and evolve to survive in new situations, in determining the fate of metals and actinides in the environment, and in engineering or stimulating microorganisms to immobilize these contaminants.

The goals of this project are to develop methods for profiling metabolites and metabolic fluxes in microorganisms and to develop strategies for perturbing metabolite levels and fluxes in order to study the influence of changes in metabolism on cellular function. We will focus our efforts on two microorganisms of interest to DOE, *Shewanella oneidensis* and *Geobacter metallireducens*, and the effect of various electron acceptors on growth and metabolism. Specifically, we will (1) develop new methods and use established methods to identify as many intracellular metabolites as possible and measure their levels in the presence of various electron acceptors; (2) develop new methods and use established methods to quantify fluxes through key metabolic pathways in the presence of various electron acceptors and in response to changes in electron acceptors; (3) perturb central metabolism by deleting key genes involved in respiration and control of metabolism or by the addition of polyamides to specifically inhibit expression of metabolic genes and then measure the effect on metabolite levels and fluxes using the methods developed above; and (4) integrate the metabolite and metabolic flux data with information from the annotated genome in order to better predict the effects environmental changes on metal and actinide reduction.

Recently, microorganisms have been explored for metal and actinide precipitation by secretion of cellular metabolites that will form strong complexes or by reduction of the metal/actinide. A complete survey of metabolism in organisms responsible for metal and actinide remediation, parallel to efforts currently underway to characterize the transcript and protein profiles in these microorganisms, would allow one to identify rate limiting steps and overcome bottlenecks that limit the rate of precipitation/reduction.

Not only will these methods be useful for bioremediation, they will also be useful for improving the conversion of plentiful renewable resources to fossil fuel replacements, a key DOE mission. For example, the conversion of cellulosic material to ethanol is limited by inefficient use of carbohydrates by the ethanol producer. Identification of limitations in cellulose metabolism and in products other than ethanol that are produced during carbohydrate oxidation could lead to more efficient organisms or routes for ethanol production – metabolomics is the key profile to identify these rate-limiting steps.

107

Metabolomic Functional Analysis of Bacterial Genomes

Clifford J. Unkefer[1]* (cju@lanl.gov), Pat J. Unkefer[1], Munehiro Teshima[1], Kwasi Godwin Mawuenyega[1], Norma H. Pawley[1], Rodolfo A. Martinez[1], Marc A. Alvarez[1], Daniel J. Arp[2], Luis Sayavedra-Soto[2], Xueming Wei[2], and Norman Hommes[2]

[1]Los Alamos National Laboratory, Los Alamos, NM and [2]Oregon State University, Corvallis, OR

Achieving the GTL goal of obtaining a complete understanding of cellular function requires integrated experimental and computational analysis of genome, transcriptome, proteome as well as the metabolome. Metabolite concentration is a product of cellular regulatory processes, and thus the metabolome provides a clear window into the functioning of the genome and proteome. Like the proteome, metabolic flux and metabolite concentrations change with the physiological state of the cell. Because metabolite flux and concentration are correlated with the physiological state, they can be used to probe regulatory networks. The power of metabolome analysis is greatly enhanced by stable isotope labeling. By combining stable isotope labeling and NMR/Mass spectral analysis, one can assess not only the metabolite concentration, but the flux through metabolic pathways as well. This approach is essential to establishing precursor product relationships, and to test if putative pathways identified from analysis of the genome are operational. Our initial results on the metabolome of *Nitrosomonas europaea* are reported here. *N. europaea* is a chemolithotroph, which derives its energy for growth from the oxidation of NH_3 to NO_2^- and fixes CO_2 as a carbon source.

Before sequencing its genome, *N. europaea* is only known to grow when provided with ammonia as an energy source, O_2 as the terminal electron acceptor, and CO_2 as the carbon source. However, *N. europaea* is not abstinent with regard to other energy sources, electron acceptors, and carbon sources. Based on homology analysis of the *N. europaea* genome genes encoding for a PTS fructose/mannose transporter, a complete glycolytic pathway, a complete TCA cycle, and a complete electron transport pathway. Based on the analysis of the genome, we have demonstrated that while using ammonia as an energy source, *N. europaea* can grow using a number of heterotrophic carbon sources. We cultured the organism in medium containing [1,2-$^{13}C_2$]pyruvate, and analyzed the ^{13}C-isotopomers of metabolites by MS and NMR spectroscopy. The results demonstrate that despite the fact that their genome encodes for a complete TCA cycle, *N. europaea* expresses only a branched TCA cycle in the presence of pyruvate.

Chemostat development - Essential to the comparative metabolomics, metabolic flux analysis and metabolic regulation studies are steady state culture conditions. During the ammonia oxidation process, the *N. europaea* acidifies its medium and builds up nitrate rapidly during growth of batch cultures. These conditions, coupled with the increasing cell density in exponential batch cultures make impossible to attribute the effect of a particular variable on the metabolite profiles or metabolic flux. To minimize culture variability, we have developed a chemostat specifically for growth of *Nitrosamonas europaea*. Our chemostat tightly regulates rate of addition of O_2, NH_3 and CO_2 using computer controlled mass flow controllers. In addition, the chemostat monitors and adjusts pH, [O_2], stirring rate, and the rate of nutrient addition. Our initial chemostat cultures of *N. europaea* were carried out at a 0.026 h^{-1} dilution. This culture was feed CO_2 and a nutrient solution at a constant rate. In this culture, the addition of ammonia was used to maintain the pH of the cultures. After two days, the culture reached steady state indicated by a constant OD_{600nm} = 0.45. In addition, the accumulated concentration of nitrate also plateaued indicating that the rate of production of nitrate by the culture was constant. As *Nitrosomonas europaea* grows it oxidizes ammonia. The linear rate of ammonia addi-

* Presenting author

tion is also an indication of steady state growth. We routinely achieve steady state cultures with an optical density (600nm) of 0.55, which is 3-5 times greater then that reported for any batch culture conditions. Initial metabolite profiles of these steady state cultures will be reported.

108

Linking Nano-Scale Patterns of ^{15}N and ^{13}C Metabolism to Bacterial Morphology

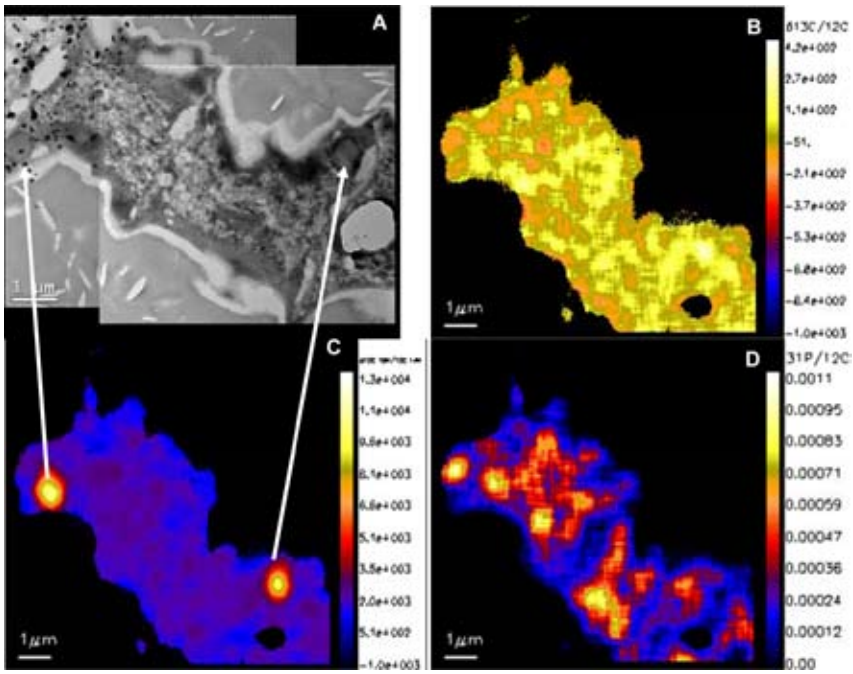
Jennifer Pett-Ridge[1]* (pettridge2@llnl.gov), Stewart J. Fallon[1], Juliette Finzi[2], Doug Capone[2], Radu Popa[2], Ken Nealson[1], Ian D. Hutcheon[1], and **Peter K. Weber**[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA and [2]University of Southern California, Los Angeles, CA

The technical challenges of tracing isotopes within individual bacteria and nanoparticles are overcome with high resolution Nano-Secondary Ion Mass Spectrometry (NanoSIMS). Samples are sputtered with an energetic primary beam, liberating secondary ions that are separated by a mass spectrometer and detected in a suite of electron multipliers. Using this technique, five isotopic species may be analyzed concurrently with spatial resolution as fine as 50nm and isotope ratio precision of ± 1.5‰. A high sensitivity isotope ratio 'map' can then be generated for the analyzed area.

We used this technique to quantitatively describe ^{13}C and ^{15}N uptake and transport in two marine cyanobacteria grown on $NaH^{13}CO_3$ and $^{15}N_2$. These diazotrophic bacteria are faced with the

Figure 1. Corresponding NanoSIMS and TEM images of two Trichodesmium cells after 8 hrs of incubation with ^{13}C-HCO_3^- and ^{15}N-N_2: (a) TEM, (b) C isotope ratio and (c) N isotope ratio (d) P: ^{12}C ratio. The filament was embedded in epoxy, ultramicrotomed into 200 nm thick sections, stained to reveal ultrastructure in TEM, and then analyzed in the NanoSIMS. The nitrogen image allows the TEM and NanoSIMS images to be correlated. The carbon and nitrogen isotope ratio data are shown as deviations from standard values in parts per thousand, as indicated in the legends ($\delta^{13}C$ and $\delta^{15}N$). ^{15}N enrichment (c) shows the localization of the newly fixed nitrogen. This fixation is correlated with N storage structures (cyanophycin) indicated in the TEM image.



challenge of isolating regions of N-fixation (O_2 inhibited) and photosynthesis (O_2 producing). In *Anabaena oscillarioides*, we found that specialized N-fixing heterocyst cells are depleted in ^{15}N relative to neighboring vegetative cells, due to high rates of N-export relative to N-fixation. Elevated $\delta^{15}N$ was also observed in intracellular zones within vegetative cells prior to septation; these ^{15}N-rich walls are attributed to newly formed proteins delineating the zone of physical separation of daughter cells. *Trichodesmium* IMS-101 are also capable of fixing both CO_2 and N_2 concurrently throughout the day, yet this species does not contain heterocysts. Using sequentially harvested bacteria, we measured alternating $^{15}N/^{13}C$ temporal enrichment patterns in this species that suggest tightly regulated changes in fixation kinetics. Spatial enrichment features indicate how ^{15}N and ^{13}C "hotspots" are dispersed throughout individual cells, and indicate isolated locations of increased N_2 fixation, sites of amino acid/protein synthesis, and cyanophycin storage granules (Figure 1). Regions of Mo and Fe accumulation suggest heightened N-fixation activity in adjacent groups of cells; both metals are co-factors for the nitrogenase enzyme. This combination of NanoSIMS analysis and high resolution microscopy allows isotopic analysis to be linked to morphological features and holds great promise for fine-scale studies of bacterial metabolism and environmental function.

109

Animal Gene Regulatory Networks

R. Andrew Cameron* (acameron@caltech.edu) and **Eric H. Davidson**

California Institute of Technology, Pasadena, CA

Recent progress in this Genomes to Life Project falls in three areas: (1), Authenticating the sea urchin endomesoderm gene regulatory network at the DNA sequence level; (2), Extending the network to the whole embryo; (3) Advances in gene regulatory network theory. Following is a brief discussion of each area.

Authentication of the Network

The endomesoderm network is now the most extensive and experimentally the best supported of all gene regulatory networks for developmental processes. These are a special class of networks because of the enormous information processing requirements of development; in the history of life they were the last to evolve, and they underlie the extremely complex process of spatial (as well as temporal) patterning of gene expression. Because the endomesoderm network has become the paradigmatic example, and its progress represents the growing edge of the field, it is essential to be able to base its predictions and regulatory logic interactions directly in the genomic regulatory code. This means recovering and subjecting to experimental analysis at the DNA sequence level the *cis*-regulatory modules at the key nodes of the network, the process we term "authenticating the network" (Davidson and Levine, 2005). An imposingly difficult task on the face of it, due to the advantages of the sea urchin system for high throughput *cis*-regulatory research, and the technologies developed in this lab for such research, it has become feasible to carry out cis-regulatory tests across the whole network. At present in this lab there are *cis*-regulatory authentication projects supported by this Project underway on the following genes at major network nodes (see Figure): ***gatae***, ***tbrain***, ***foxa***, ***brn1/2/4***, ***cyclof ilin***, ***brachyury***, ***blimp1/krox***; of these, ***cyclof ilin*** is complete and a report is In Press. Studies were published last year on ***otx*** (**Yuh et al, 2004**); on ***delta*** (**Revilla and Davidson, 2004**) and on ***wnt8*** (**Minokawa et al, 2005**). In addition there were published functional studies on a new gene added into the network, ***brn1/2/4*** (**Yuh et al, 2005**), and on ***gatae*** (**Lee and Davidson, 2005**). The

* Presenting author

key circuitry of the endomesoderm network is increasingly able to be expressed in the "hardwired" terms of the A's, C's, G's, and T's of the genomic *cis*-regulatory sequence. It is encouraging that current evidence, published and otherwise, is so far almost entirely corroborating the network linkages that were predicted by the perturbation analyses on the basis of which the network was constructed.

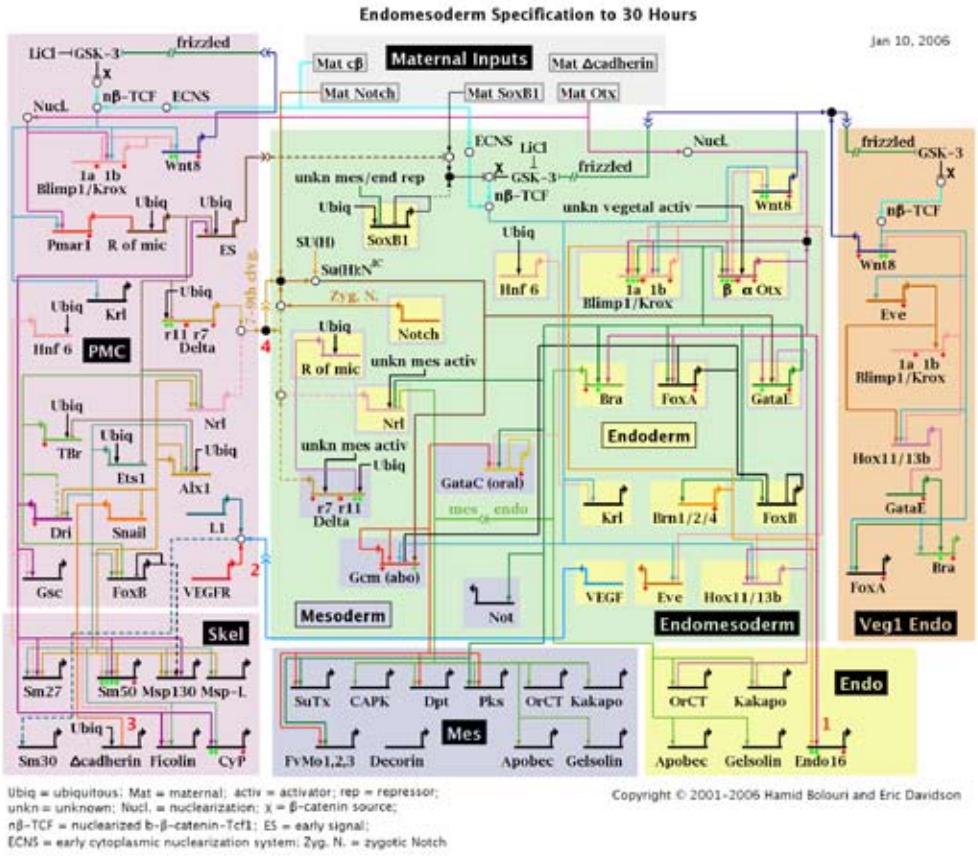
Extending the network to the rest of the embryo

This initiative, one of our original ambitions in this Project, has now been launched. The tremendous reward that will be there for the taking, if we are successful, is that when the components of the whole embryo are included in a network analysis, the overall network will constitute a closed and complete system; after the earliest stages the inputs will all arise within the network. Such a closed, whole organism developmental system has never before been available for study at this level. The effort that has begun is aimed at solving the network for the whole embryo for the same time frame to which the endomesoderm network pertains, ~6-30 hrs after fertilization. The missing parts which are our objective are the oral and aboral ectoderm networks, and the neurogenic apical plate network. Several other labs have begun to work on the latter, and we hope they will be successful, so we won't have to do it also. The oral and aboral ectoderm networks are a joint project of the laboratory of our subcontractor in the GTL project, David R McClay at Duke, and ourselves. We have devised a new high throughput methodology to build these gene regulatory networks: (1) Determine all regulatory genes that encode sequence specific DNA binding factors predicted in the genomic sequence of this animal, measure their time course of expression, and for those which are expressed by 30 hrs, determine where they are expressed. This is all complete, except for some Zn Finger transcription factor genes, and they will be done soon. (2) For those expressed specifically in oral or aboral ectoderm or both, obtain and verify morpholino substituted antisense oligonucleotides; (3) Carry out a multiplexed perturbation analysis such that the effect of blocking translation of each gene on the expression of all other specific oral and aboral ectoderm regulatory genes is determined at once in a matrix analysis; (4) Utilize a very high density whole transcriptome genomic tiling array chip for the perturbation analysis. (5) Apply the soon to be released version of BioTapestry network building software, which has been built essentially for this purpose, to deconvolve the temporal, spatial, and perturbation data and generate the allowable network architectures. If this new approach works it will revolutionize the practice experimental network building. The network structure will be computationally determined; all expressed genes or any desired subset can be included so the issue of completeness will disappear; and the observations can be multiplexed so that the solution of the network will be vastly accelerated.

Network Theory

During this year the PI worked out a coherent body of new theory for gene regulatory networks which control spatial and temporal gene expression, and which function essentially to set up regulatory states of domains. This will appear in a monograph to be published next year by Elsevier, "**The Regulatory Genome: Gene Networks in Development and Evolution**". Developmental gene regulatory networks are here treated as networks of information processing nodes, i.e., the individual *cis*-regulatory modules of the network. The ground was laid in previous work undertaken as part of the GTL Project, in which the logic operations integrated within various *cis*-regulatory modules are treated explicitly (**Istrail and Davidson, 2005**). Information processing also emerges from the operations of the subcircuits of which the networks are composed. It is the architecture of these subcircuits which determines the biological functions of the developmental process. The network as a whole is thus to be considered as a large, genomically encoded, delocalized computational device which interprets regulatory information in order to program the dynamic process of development. This "computer" has the capacity to respond conditionally to every possible regulatory state the genome will encounter in each cell of the organism throughout the life cycle.

Figure 1. The Sea Urchin Endomesoderm Gene Regulatory Network



* Presenting author