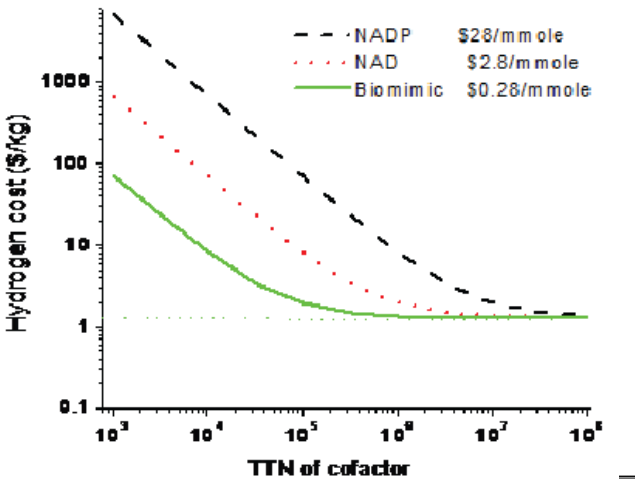

Fig. 4. The estimated hydrogen production costs in terms of turnover number of enzymes and co-enzymes. Carbohydrate prices are $0.18 per kg [2].

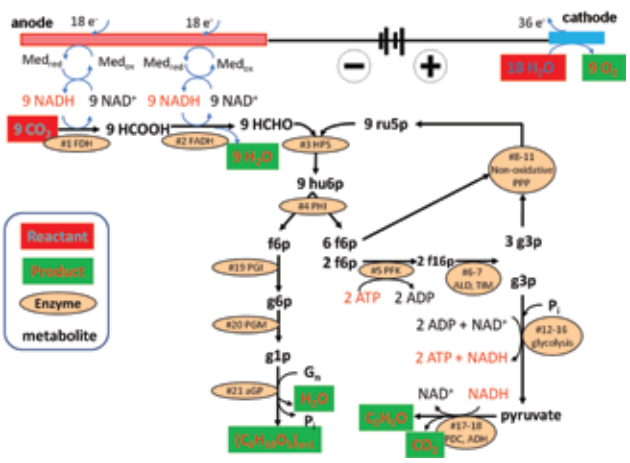

Fig. 5. Artificial photosynthesis that can fix CO_2 by using electricity for producing biofuels and food (under review).

Computing for Bioenergy

138

Bayesian Computational Approaches for Gene Regulation Studies of Bioethanol and Biohydrogen Production

Charles E. Lawrence[1] (charles_lawrence@brown.edu), Lee Newberg,[2,3] William Thompson,[1] and Lee Ann McCue[4]*

[1]Center for Computational Molecular Biology and the Division of Applied Mathematics, Brown University, Providence, R.I.; [2]The Wadsworth Center, New York State Dept. of Health, Albany; [3]Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, N.Y.; [4]Pacific Northwest National Laboratory, Richland, Wash.

http://www.brown.edu/Research/CCMB/

Project Goals: We are extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs and developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of alpha-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

Decreasing America's dependence on foreign energy sources and reducing the emission of greenhouse gases through the development of biofuels are important national priorities. These priorities have catalyzed research on cellulosic ethanol as a clean, renewable energy source to replace fossil fuels, and biohydrogen as a carbon-free energy carrier. Turning these biofuels into viable alternative energy sources requires further research into the degradation of cellulose and fermentation of the resulting sugars, and the metabolic and regulatory networks of biohydrogen production. The genomes of many of the microbial species capable of these processes have been sequenced by the GTL and other programs, and many more are expected soon. These sequence data provide a wealth of information to explore nature's solutions for the production of biofuels. In particular, among the over 170 α-proteobacterial species with genome sequence data available are several species with metabolic capabilities of interest, including efficient fermentation of sugars to ethanol and the ability to produce hydrogen. Understanding the regulatory mechanisms and complex interplay of metabolic processes in these species is key to realizing the promise of biofuels. Thus, our research goal is to identify the ensemble of solutions that have been explored by the α-proteobacteria to regulate the metabolic processes key to biofuel production.

* Presenting author

The solution space explored by these species spans three scales: molecular (genes/gene products), cellular (genomic), and communities (clades). On the finest scale, the catalytic steps of biofuel production (ethanol or hydrogen) are performed by individual enzymes that are the products of individual genes, each regulated by a set of *cis* and *trans* elements. At the cellular level, the expression of the gene products is often coordinated via a set of *trans* elements (transcription factors) that interact with all or most genes in these pathways to form a regulatory unit called a regulon. Furthermore, the collection of species that encode the genes form clades which have explored a catalytic and regulatory space on an evolutionary time scale. We are developing probabilistic models to represent these multiscale processes, Bayesian statistical inference procedures and computational methods to identify the posterior distributions of these parameters, efficient point estimates of their values, and Bayesian confidence limits for these estimates.

Specifically, we are characterizing the gene/clade interface by extending our phylogenetic Gibbs sampling algorithms to reconstruct the joint posterior space of the ancestral states of regulatory motifs, and developing point estimates and confidence limits for these discrete high-dimensional objects. We are also applying our existing models and technologies, along with the above modifications, to clades of α-proteobacterial species, to identify regulatory mechanisms and reconstruct the ancestral states of the regulatory networks for the efficient fermentation of sugars to ethanol and the production of biohydrogen.

139

Sugar-Salt and Sugar-Salt-Water Complexes: Structure and Dynamics of Glucose – KNO_3 - $(H_2O)_n$

Madeleine Pincu[1,2]* (mpincu@uci.edu), Brina Brauer,[2] **Robert Benny Gerber**,[1,2] and Victoria Buch[2†]

[1]The Fritz Haber Center for Molecular Dynamics, Chemistry Institute, the Hebrew University, Jerusalem, Israel and [2]Dept. of Chemistry, University of California, Irvine
[†]Deceased

Project Goals: To investigate structure, dynamics, and energetics of saccharides and their interactions with ions and biological osmolytes in the presence of micro-hydration.

Molecular dynamics (MD) simulations are carried out for the complex of glucose with KNO_3 and for complexes of the type glucose – KNO_3 - $(H_2O)_n$, for $n \leq 11$. Structure and dynamic properties of the systems are explored. The MD simulations are carried out using primarily the DLPOLY/OPLS force field, and global and local minimum energy structures of some of the systems are compared with ab initio MP2 calculations. The main findings include: (1) Complexation with KNO_3 leads to an "inverse anomeric effect," with the β-glucose complex more stable than the α-glucose by ~ 1.74 Kcal/mol. (2) As temperature is increased to 600K, the KNO_3 remains undissociated in the 1:1 complex, with the K^+ hooked to the equilibrium site, and the NO_3^- bound to it, undergoing large-amplitude bending/torsional motions. (3) For $n \geq 3$ water molecules added to the system, charge separation into K^+ and NO_3^- ions takes place. (4) For n =11 water molecules all hydroxyl groups are hydrated with the glucose adopting a surface position, indicative of a surfactant property of the sugar. (5) Comparison of DLPOLY with MP2 structure predictions indicates that the empirical force field predicts global and local minimum structures reasonably well, but errs in giving the energy rankings of the different minima. Implications of the results to effects of salts on saccharides are discussed.

Small Business Innovation Research (SBIR)

140

Genetic Analysis of Cellulose Degradation by *Clostridium phytofermentans*

Andrew C. Tolonen[1]* (tolonen@genetics.med.harvard.edu), Amanda C. Chilaka,[1,2] and **George M. Church**[1]

[1]Dept. of Genetics, Harvard Medical School, Boston, Mass. and [2]Dept. of Biology, Northeastern University, Boston, Mass.

Project Goals: See below.

Microbial cellulose degradation is a central part of the global carbon cycle and has potential for the development of inexpensive, carbon neutral biofuels from non-food crops. The major roadblock to the use of cellulosic biomass as a biofuel feedstock is the recalcitrance of cellulosic fibers to breakdown into sugars. *Clostridium phytofermentans* grows on both of the two main components of plant biomass, cellulose and hemicellulose, by secreting enzymes to cleave these polysaccharides and then fermenting the resulting hexose and pentose sugars to ethanol. In order to breakdown cellulose biomass, *C. phytofermentans* has a repertoire of 161 carbohydrate-active enzymes (CAZy), which include 108 glycoside hydrolases spread across 39 families.

Broadly, our goal to understand the genetic mechanisms that permit to *C. phytofermentans* to efficiently convert cellulosic biomass to ethanol. To enable targeted gene inactivation in *C. phytofermentans,* we show that interspecific conjugation with *E. coli* can be used to transfer a plasmid into *C. phytofermentans* that has a resistance marker, an origin of replication that can be selectively lost, and a designed group II intron for efficient, targeted chromosomal insertions without selection. We applied these methods to inactivate Cphy3367, a ß-1,4-glucanase in glycoside hydrolase family 9 (GH9). Cellulolytic *Clostridia* usually have numerous