

Computing for Bioenergy

138

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

Charles E. Lawrence¹ (charles_lawrence@brown.edu), Lee Newberg,^{2,3} William Thompson,¹ and Lee Ann McCue^{4*}

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

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 – $\text{KNO}_3 - (\text{H}_2\text{O})_n$

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

¹The Fritz Haber Center for Molecular Dynamics, Chemistry Institute, the Hebrew University, Jerusalem, Israel and ²Dept. 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 microhydration.

Molecular dynamics (MD) simulations are carried out for the complex of glucose with KNO_3 and for complexes of the type glucose – $\text{KNO}_3 - (\text{H}_2\text{O})_n$, 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.

submitted post-press

A Multi-Scale Approach to the Simulation of Lignocellulosic Biomass

Goundla Srinivas,¹ Yana Kholod,² Loukas Petridis,¹ James W. Evans,² Monica Lamm,² Theresa L. Windus,² Xiaolin Cheng^{1,3*} (chengx@ornl.gov), **Mark S. Gordon**,² and **Jeremy C. Smith**^{1,3}

¹Oak Ridge National Laboratory, Oak Ridge, Tenn.;

²Ames Laboratory, Iowa State University, Ames; and

³University of Tennessee, Knoxville

Project Goals: In concert with the imminent increase in the Department of Energy's (DOE) leadership supercomputing power to the petaflop range, the objective of this project is to develop multiscale methods for extending the time- and length-scales accessible to biomolecular simulation on massively parallel supercomputers. This project also aims to apply the developed multiscale approaches to obtain an understanding of the structure, dynamics and degradation pathways of extended cellulosic and lignocellulosic materials. Information from multiscale simulation, when closely integrated with experiment, will provide fundamental understanding needed to overcome biomass recalcitrance to hydrolysis.

The research involves the development of multiscale simulation methods and their application to solve critical problems needed for understanding the bottleneck in cellulosic ethanol production: the recalcitrance to hydrolysis of lignocellulosic biomass. The multiscale methodologies span from accurate quantum-chemical techniques, needed to understand critical local interactions in biomass, to atomistic and coarse-grained simulations, needed to approach systems-level phenomena. The codes developed will be parallelized for efficient use on petascale supercomputers. The physical simulation models of lignocellulosic biomass derived using the multiscale approaches will serve as a basis for interpreting an array of biophysical experiments, and when closely integrated with experiments, will eventually lead to a description of the physicochemical mechanisms of biomass recalcitrance to hydrolysis, and thus aiding in developing a strategy to overcome the recalcitrance. Our progress presented here mainly focus on: adapting the quantum mechanical fragment molecular orbital (FMO) method to the study of cellulosic/lignocellulosic biomass, developing an adaptive fast multipole based Poisson Boltzmann (PB) electrostatic solvation model for nano-scale biomolecular systems and self-consistently coarse-graining of cellulose/lignocellulose force fields.

This research is funded by the Genomic Science Research Program, Office of Biological and Environmental Research (BER), and the Scientific Discovery through Advanced Computing (SciDAC) program, U. S. Department of Energy (DOE), currently under FWP ERKJE84.

submitted post-press

Processivity of Cellobiohydrolase, Cellulose Structure, and Advanced Methods for Petascale Molecular Dynamics

James F. Matthews,¹ Xiaolin Cheng,² Malin Bergenstrahle,³ Eric H. Knoll,¹ Mark Williamson,⁴ Edward C. Uberbacher,² John W. Brady,³ Charles L. Brooks III,⁵ Linghao Zhong,⁶ Mark R. Nimlos,¹ Michael E. Himmel,¹ and **Michael F. Crowley**^{1*} (Michael.Crowley@nrel.gov)

¹National Renewable Energy Laboratory, Golden, Colo.; ²Oak Ridge National Laboratory, Oak Ridge, Tenn.; ³Cornell University, Ithaca, N.Y.; ⁴San Diego Supercomputer Center, San Diego, Calif.; ⁵University of Michigan, Ann Arbor; and ⁶Pennsylvania State University, Mont Alto

Project Goals: The SciDAC project is focused on molecular modeling of molecular and macromolecular systems that are key to the understanding and designing of cellulose degrading technologies acting on biomass. The project is enabling scientific investigation of cellulose structure and properties and enzymatic mechanism (processivity) for degrading cellulose to sugars as a raw material for biofuels. Speed, new sampling methods, and simulation size (scaling) are the aims of code development and performance work with the additional creation of validation suites for code integrity. The main thrust is to enable the specific needs of modeling to produce the methods necessary for thermodynamic, kinetic, and mechanistic insight into cellulose conversion processes at a speed which will make this research possible, mainly through the use of high-performance computers at the petascale level and advanced sampling methods.

The mechanism by which Cel7A cellobiohydrolase degrades cellulose is not currently known nor is the structure and molecular properties of cellulose microfibrils as found in biomass. The degrading of cellulosic biomass to sugars is primarily accomplished both in nature and in biorefineries by enzymes such as Cel7A; the understanding of both the structure of cellulose and of the mechanism of cellulose-degrading enzymes is critical to improving the technology of producing biofuels from biomass. There have been several barriers to studying these molecular systems with modeling, primarily the lack of reliable force fields for cellulose and the lack of highly scalable molecular dynamics programs that can treat these systems and have the particular thermodynamic sampling abilities and force field features to answer the questions which are unique to this difficult problem. Our progress presented here is fivefold: cellulose force field evaluation and characterization, molecular dynamics code validation, enabling highly parallel programs to utilize cellulose force fields, solving parallel bottlenecks in existing codes, and designing highly parallel implementations of sampling methods for studying enzymatic mechanisms and cellulose morphology and decrystallization.

The SciDAC project “Understanding Processivity of Cellobiohydrolases” is supported by the Office of Biological and Environmental Research and OASCR in the DOE Office of Science.

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

¹Dept. of Genetics, Harvard Medical School, Boston, Mass. and ²Dept. 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 β -1,4-glucanase in glycoside hydrolase family 9 (GH9). Cellulolytic *Clostridia* usually have numerous genes for GH9 proteins: the *C. thermocellum* ATCC 27405 genome has 16 GH9 genes, *C. cellulolyticum* H10 has 13 GH9 genes, and *C. cellulovorans* has 5 GH9 genes. In contrast, *C. phytofermentans* has only a single GH9-encoding gene, *cphy3367*. The *C. phytofermentans* strain with an intron insertion in *cphy3367* (strain AT02-1) grows normally on some carbon sources such as glucose, cellobiose, and hemicellulose, but has lost the ability to degrade cellulose (Fig 1). Although *C. phytofermentans* up-regulates the expression of numerous enzymes to breakdown cellulose, this process thus relies upon a single, key hydrolase, Cphy3367. Generally,

these results show that targeted gene inactivation can be used to identify key enzymes for the breakdown of biomass by *C. phytofermentans*. Future genetic studies of in *C. phytofermentans* will untangle the roles of additional hydrolases in cellulose degradation.

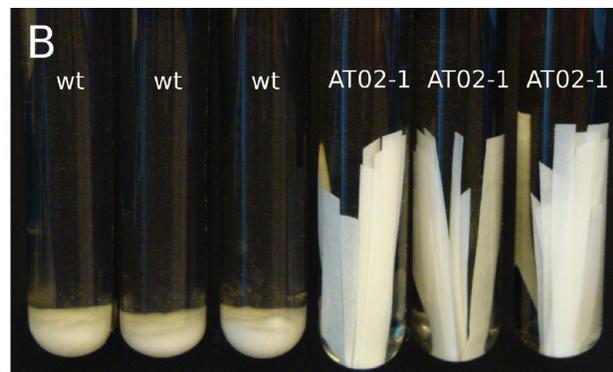
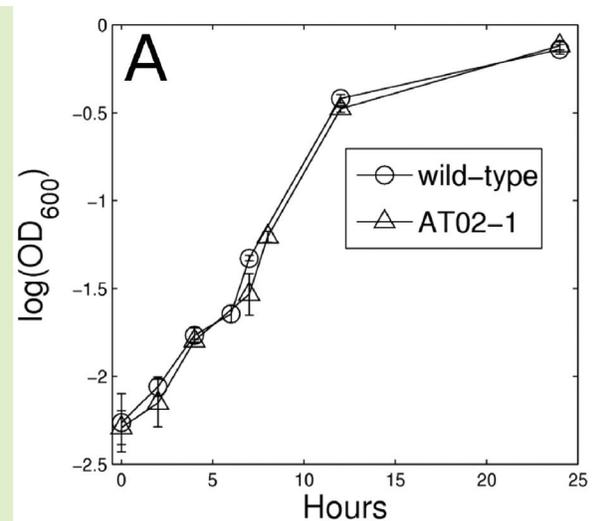


FIG 1 *C. phytofermentans* strain with disruption of *cphy3367* (AT02-1) had similar growth rates as wild-type on glucose A, but had lost the ability to degrade filter paper cellulose B. Growth curves are means of triplicate cultures. Error bars show one standard deviation and are smaller than the symbols where not apparent.

141

Microbioreactor Technology for Obligate Anaerobes

Harry Lee (harrylee@pharyx.com) and Paolo Boccazzi* (boccazzi@pharyx.com)

Pharyx, Inc., Boston, Mass.

<http://www.pharyx.com>

Project goals are 1) determine microbioreactor designs that will support anaerobic inoculation and fermentation, 2) identify optimal materials for fabricating anaerobic bioreactors, 3) determine the range of process parameters where microbioreactor data corresponds to serum tubes and stirred tank fermentors, 4) monitor enzyme activity on-line.

Anaerobic microorganisms have evolved biochemical pathways that can be exploited for industrial applications. These include the ability to breakdown environmental pollutants for bioremediation, the breakdown of cellulose into simple sugars for biofuels, and the production of specialty chemicals. However, there remains a tremendous challenge to the scale-up of bioenzymatic activities to industrial processes. While systems biology approaches and metabolic engineering promise to contribute to our understanding of these systems, a key bottleneck is in conducting controlled experiments to ground these approaches with high quality data. Thus far, experiments are frustrated by the laborious set-up and operation of stirred tank bioreactor systems, which for anaerobic microbiology is further encumbered by the requirement of an anaerobic environment. The absence of easy to use systems also holds back more traditional microbiology approaches such as mutagenesis and screening and directed evolution.

We are developing a parallel bioreactor system, based on microfluidic integration technology and disposable microbioreactor modules, with application specific customizations for anaerobic fermentation. These customizations are aimed to enable up to 32 simultaneous anaerobic fermentations under controlled conditions, with online monitoring of growth kinetics and other phenotypes such as enzyme activity. A unique feature of this system is the ability to operate it in ambient air through careful inoculation port and reactor and control module design, or to operate it within an anaerobic bag, taking advantage of its compact size.

Preliminary anaerobic fermentations in microbioreactor devices are shown in Figure 1. Cell growth in the three bioreactor chambers compared to the uninoculated medium control is clear, and exponential growth can be seen for $t < 5$ h, indicating obligate anaerobes can be cultured in plastic bioreactors fabricated using our microfluidic integration platform. However, for the *Butyrivibrio fibrisolvens* and *Clostridium acetobutylicum* fermentations, optical density measurements were confounded by the gas bubbles generated by the microorganisms. This was confirmed visually during the fermentation and evidenced in the photographs,

shown in Figure 2, of the devices at the end of the fermentation.

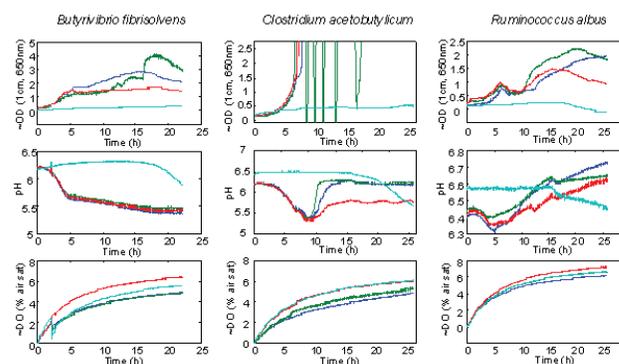


Figure 1. Online sensor measurements of dissolved oxygen (DO), pH and optical density (OD) for anaerobic fermentations of three obligate anaerobes. Growth studies were conducted in our anaerobic bioreactor modules which contained 4 growth chambers each. Light blue lines represent the uninoculated controls, other 3 lines represent 3 independent reactor replicates in each cassette. All strains were grown in modified peptone yeast extract glucose (MPYG) medium (Atlas RM, 2004. Handbook of microbiological media. CRC, Boca Raton). Measurements were taken every 60 seconds.

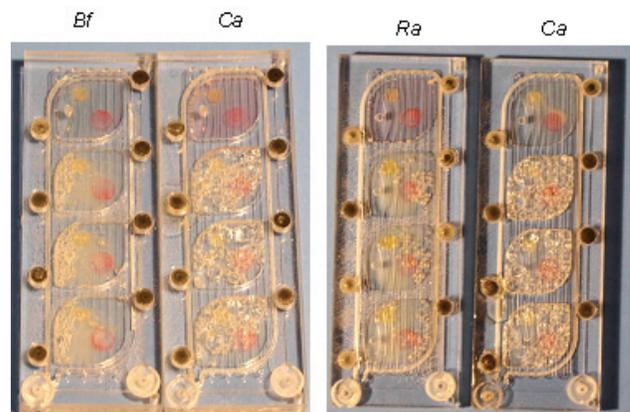


Figure 2. Photographs of integrated bioreactor devices. Pink circles are dissolved oxygen sensors, yellow circles are pH sensors, black plugs are butyl rubber injection ports. A stream of nitrogen flowed through the serpentine channels that surround the chambers. The chambers contain a medium with visible gas bubbles, particularly in the Bf (*Butyrivibrio fibrisolvens*), Ca (*Clostridium acetobutylicum*), Ra (*Ruminococcus albus*). First growth reactor from top of each cassette contained uninoculated medium.

142

A Genome-Wide Perspective on the Regulation of Plant Carbohydrate Conversion to Biofuels in *Clostridium phytofermentans*

Elsa Petit,¹ Tom Warnick,¹ Greg Latouf,^{1,6} James Hayes,^{2,5} Amy Biddle,^{1*} Zhiyi Sun,¹ Danny Schnell,^{2,4,5} **Sue Leschine**,^{1,4,5} and **Jeff Blanchard**^{1,4,5} (blanchard@microbio.umass.edu)

¹Dept. of Microbiology, ²Dept. of Biochemistry, ³Graduate Program in Molecular and Cellular Biology, ⁴The Massachusetts Institute for Biofuels Research, and ⁵Institute for Cellular Engineering, University of Massachusetts, Amherst and ⁶Qteros Inc.

Project Goals: Derive genetic models of the conversion of cellulosic biomass to alcohols in *C. phytofermentans* by microarray analysis of the fermentation of simple sugars contained in feedstocks relevant to the agricultural and forestry industries.

The economic costs of degrading cellulosic biomass currently hold back widespread use of plant biomass from agricultural and forestry wastes as a domestic renewable alternative to gasoline. *Clostridium phytofermentans* ferments all major components of the plant cell wall to ethanol and is an emerging model organism for understanding the direct conversion of plant biomass to fuels. The *C. phytofermentans* genome contains an abundance of carbohydrate degradation enzymes, carbohydrate transporters and transcriptional regulators. In order to build a foundation for engineering improvements on industrially relevant feedstocks, microarray experiments were carried out on a number of purified plant cell wall carbohydrates. The results demonstrate that *C. phytofermentans* regulates the stoichiometry of the plant degradative and assimilatory machinery in response to substrate availability. Because of the modularity of the sugar degradation and assimilation systems and the lack of a cellulosome, further improvement in biomass degradation through genetic manipulation is relatively straightforward.