

An Overview of the Sandia National Laboratories Genomes to Life Project, “Carbon Sequestration in *Synechococcus* Sp.: From Molecular Machines to Hierarchical Modeling”

Grant S. Heffelfinger
Sandia National Laboratories
Albuquerque, NM

Abstract

The Sandia-led Genomes to Life (GTL) project: “Carbon Sequestration in *Synechococcus* Sp.: From Molecular Machines to Hierarchical Modeling,” is focused on developing, prototyping, and applying new computational tools and methods to elucidate the biochemical mechanisms of the carbon sequestration of *Synechococcus* Sp., an abundant marine cyanobacteria known to play an important role in the global carbon cycle. Our effort includes five subprojects: an experimental investigation, three computational biology efforts, and a fifth which deals with addressing computational infrastructure challenges of relevance to this project and the Genomes to Life program as a whole. Our experimental effort is designed to provide biology and data to drive the computational efforts and includes significant investment in developing new experimental methods for uncovering protein partners, characterizing protein complexes, identifying new binding domains. Our computational efforts include coupling molecular simulation methods with knowledge discovery from diverse biological data sets for high-throughput discovery

and characterization of protein-protein complexes and developing a set of novel capabilities for inference of regulatory pathways in microbial genomes across multiple sources of information through the integration of computational and experimental technologies. We are also investigating methods for combining experimental and computational results with visualization and natural language tools to accelerate discovery of regulatory pathways and developing set of computational tools for capturing the carbon fixation behavior of complex of *Synechococcus* at different levels of resolution. Finally, because the explosion of data being produced by high-throughput experiments requires data analysis and models which are more computationally complex, more heterogeneous, and require coupling to ever increasing amounts of experimentally obtained data in varying formats, we have also established a companion computational infrastructure to support this effort. This element of our project will be discussed in the larger GTL program context as well. More information about our project and partners can be found at www.genomes-to-life.org