Report from the Subcommittee on the Evaluation of the Long Term Goal of Demonstrating Progress Toward Developing Computational Models of a Microbe and a Simple Microbial Community (Genomics:GTL Program)

Subcommittee Member: Virginia Torczon

Charge: The subcommittee was asked to evaluate progress toward the long term goal:

> *By 2015, demonstrated progress toward developing through the Genomes to Life partnership with the Biological and Environmental Research program, the computational science capability to model a complete microbe and simple microbial community.*

Outline:

1. Recap of the ASCR PART Long Term Measures.
2. Summary of ASCR activities in support of this goal:
    - Genomics:GTL awards through ASCR,
    - SciDAC Life Sciences, and
    - Institutes for the Advancement of Computational Biology Research & Education.
3. BERAC draft report and proposed changes to both the goal and the PART Long Term Measures.

ASCR PART Long Term Measures for Genomics:GTL

- *Definition of "Excellent"—In partnership with BER, develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated experimentally by the use or reengineering of that community based on model predictions.*
- *Definition of "Good"—In partnership with BER, develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated by its consistency with available data.*
- *Definition of "Fair"—In partnership with BER, develop a number of the components of a computational model that could accurately describe the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen.*
- *Definition of "Poor"—In partnership with BER, produce a modest output of computational research that could lead to the development of components of a model to describe the potential of microbial communities to clean up waste, sequester carbon, or produce hydrogen.*

Proposals funded by ASCR through the GTL Program

- Carbon Sequestration in Synechococcus: From Molecular Machines to Hierarchical Modeling
  Principal Investigators:
  G. Heffelfinger, Sandia National Laboratories
  A. Geist, Oak Ridge National Laboratory
- Development of Computational Tools for Analyzing and Redesigning Metabolic Networks
  Principal Investigators:
  C. Maranas, Penn State University
  A. Burgard, Genomatica, Inc.
- Intercellular Genomics of Subsurface Microbial Colonies
  Principal Investigators:
  P. Ortoleva, Indiana University
  K. Tuncay, Indiana University
  D. Gannon, Indiana University
  C. Meile, University of Georgia

SciDAC Life Sciences

Putting Microbes to Work: Developing microbial communities to generate clean fuel sources and clean up environmental contaminants

DOE Program Managers

John Houghton DOE Office of Biological & Environmental Research	Anil Deane DOE Office of Advanced Scientific Computing Research

SciDAC Life Sciences: Putting Microbes to Work

The Department's missions in energy and environment have focused life science efforts on microbes and microbial communities that have the potential to generate hydrogen or ethanol or to sequester carbon dioxide or environmental contaminants.

This field is new to SciDAC and will focus on developing new methods for modeling complex biological systems, including molecular complexes, metabolic and signaling pathways, individual cells and, ultimately, interacting organisms and ecosystems.

Such systems act on time scales ranging from microseconds to thousands of years and the systems must couple to huge databases created by an ever-increasing number of high-throughput experiments.

SciDAC Life Sciences Research Projects Announced in September 2006

Predicting the Function of Proteins for Newly Sequenced Organisms

Algorithms and engineering for gene function annotation for Joint Genome Institute genomes

Principal Investigator: Steven E. Brenner
Lawrence Berkeley National Laboratory

Green Energy: Advancing Bio-hydrogen

Developing a model of metabolism linked to H2 production in green algae

Principal Investigator: Michael Seibert
National Renewable Energy Laboratory

DOE Institutes for the Advancement of Computational Biology Research & Education

The Office of Advanced Scientific Computing Research of the Office of Science, U.S. Department of Energy, has created three institutes for the advancement of computational biology research and education, in support of the ASCR computational biology program, the ASCR-BER DOE Genomics:GTL program, and the broader Office of Science research programs.

DOE Institutes for the Advancement of Computational Biology Research & Education

The institutes support the advancement of computational biology research as an intellectual pursuit and provide innovative approaches to educating biologists as computational scientists.

The institutes focus on advancing computational biology research and education as counterbalancing and complementary activities to experimental biology.

The institutes utilize interdisciplinary teams of researchers, drawn from the physical and life sciences, computational mathematics and computer science. Two of the institutes are collaboratory efforts between universities and national laboratories.

DOE Institutes for the Advancement of Computational Biology Research & Education

The three institutes are

- Institute for Multi-Scale Modeling of Biological Interactions
  Johns Hopkins University
- Center for Computational Biology
  University of California, Merced
- BACTER (Bringing Advanced Computational Resources To Environmental Research) Institute
  University of Wisconsin, Madison

Institute for Multi-Scale Modeling of Biological Interactions

Johns Hopkins University in conjunction with the University of Delaware and Los Alamos National Laboratory

Focus:

- To study biological systems across multiple scales of time and length, ranging from protein interactions at the molecular level to the behavior of complex biochemical networks in entire organisms.

Center for Computational Biology

University of California, Merced; Rensselaer Polytechnic Institute; and Lawrence Livermore National Laboratory

Focus:

- To sponsor multidisciplinary scientific projects in which biological understanding is guided by computational modeling.
- To facilitate the development and dissemination of undergraduate and graduate materials based on the latest computational biology research.

BACTER Institute

University of Wisconsin, Madison

Focus:

- To find the genes, transcribe the proteins, predict the folds, perform pairwise docking of the model structures, deduce pathways, and ultimately construct whole cell models of Rhodobacter sphaeroides and Shewanella oneidensis.
- To engage students in project-oriented computational biology research through three interdependent research tracks: Genomics and Biological Pathway Analysis; Protein Dynamics, Folding, and Docking; and Macrobiological Modeling.
- To train students to uncover biological mechanisms and pathways within these microbial organisms through the use of computational biology and synergistic collaborations with experimental groups.

Summary of informal Institute reviews, Summer 2006

Overall the institutes were doing a good job at developing curricula and engaging new and young scientists and students.

Need more interaction amongst each other and stronger, more up-to-date research components.

Conclusions:

- ASCR needs a biology sub-community that is conversant with high end computation.
- Fully engage BER in the creation of this sub-community.

Summary of ASCR activities in support of the Genomics:GTL Program

- Cooperation with BER.
- Joint review of projects for funding under the Genomics:GTL program.
- Institutes in support of computational biology research and education.
- Efforts to better integrate ASCR activities in computational biology with BER activities in the life sciences.

In general, ASCR is heeding the recommendation contained in the 2005 National Research Council (NCR) of the National Academies report *Mathematics and 21st Century Biology* from the Committee on Mathematical Sciences Research for DOE's Computational Biology.

From *Mathematics and 21st Century Biology*:

Recommendation: Funding agencies supporting mathematical research related to the life sciences should place increased emphasis on funding mechanisms and novel approaches to the organization of interdisciplinary research. The goal should be to foster effective collaboration between mathematical scientists and bioscientists by working to eliminate barriers posed by inadequate communication, disparate timescales for achieving research objectives, inequitable recognition of contributors to interdisciplinary projects, and cultural divisions within universities, research institutes, and national laboratories.

ASCAC concern:

- Does adequate communication exist with the biology research community?
- Are the timescales reasonable for achieving the program goal of developing the computational science capability to model a complete microbe and simple microbial community?

Cooperation between ASCR and BER program managers is not enough.

To achieve this goal by 2015, the biology and the computational science research communities must set a joint timescale that ensures the desired outcome is feasible.

BERAC Review of the Life Science PART Measures

The BERAC draft report recommends modifying both the charge and the PART Long Term Measures.

The BERAC recommendations are prompted by both

- the 2006 National Research Council of the National Academies report *Review of the Department of Energy's Genomics:GTL Program* and
- changes in Administration policy regarding liquid transportation fuels.

Specifically:

From the NRC *Review of the Department of Energy's Genomics:GTL Program*:

> *...plant biology research be included in the program, where appropriate, because plants represent a major pathway to the production of bioenergy, play an important role in carbon sequestration and global nutrient cycles, and are potential sources of bioremediation.*

From the BERAC draft report:

> *...when this PART Measure was developed, the Administration had made hydrogen a priority but had not yet added a focus on liquid transportation fuels such as cellulosic ethanol. BERAC recommends that the Life Science PART measure be modified to include biofuels such as cellulosic ethanol.*

ASCR and BER Life Sciences PART Long Term Measures

ASCR Original: *By 2015, demonstrated progress toward developing through the Genomes to Life partnership with the Biological and Environmental Research program, the computational science capability to model a complete microbe and simple microbial community.*

BER Original: *By 2015, characterize the multi protein complexes (or lack thereof) involving a scientifically significant fraction of a microbe's protein's. Develop computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen.*

BER Life Sciences PART Long Term Measure

BER Original:

By 2015, characterize the multi protein complexes (or lack thereof) involving a scientifically significant fraction of a microbe's protein's. Develop computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen.

Replacement proposed by BERAC:

By 2015, provide sufficient scientific understanding of plants and microbes to develop robust new strategies to produce biofuels, clean up waste, or sequester carbon. This includes research that supports the development of computational models to direct the use and design of improved organisms carrying out these processes.

Rationale for BERAC's Proposed Replacement:

BER's current Life Sciences PART Measure has two separate goals:

1. *characterize the multiprotein complexes (or lack thereof) involving a scientifically significant fraction of a microbe's proteins and*

2. *develop computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen.*

The second of these two goals is still highly relevant and critical to the long term success of the program. BERAC recommends that the modeling component be retained in concept but integrated with the proposed changed (see below) in the first element.

Rationale for BERAC's Proposed Replacement:

The first element of the Life Science PART Measure that originally focused on multiprotein complexes was developed under the assumption that the Genomics:GTL Program would develop technology-based user facilities, including high throughput user facilities for protein production and for the isolation and characterization of molecular machines.

Rationale for BERAC's Proposed Replacement:

Based on the recommendations of the NRC review of the Genomics: GTL program and the emphasis in the Administration's Advanced Energy Initiative on liquid biofuels such as cellulosic ethanol, the decision was made not to develop the originally planned GTL User facilities. Instead, the Program will develop a series of vertically integrated research centers to accelerate fundamental research that leads to breakthroughs in basic science to make biofuels a cost-effective alternative to fossil fuels.

..., this change in Genomics:GTL strategy from the development of technology-based user facilities to vertically integrated research centers necessitates a change in the Life Sciences PART measure.

BERAC's Proposed Replacement:

The original definitions of Poor, Fair, Good and Excellent as "grades" assigned to progress in achieving the PART goals were based on the two component PART measure. BERAC is now proposing vide infra, *that BER adopt a single component and that the original definitions be modified with removal of the references to multiprotein complexes.*

Recap of BER Life Sciences PART Long Term Measure

BER Original:

By 2015, characterize the multi protein complexes (or lack thereof) involving a scientifically significant fraction of a microbe's protein's. Develop computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen.

Replacement proposed by BERAC:

By 2015, provide sufficient scientific understanding of plants and microbes to develop robust new strategies to produce biofuels, clean up waste, or sequester carbon. This includes research that supports the development of computational models to direct the use and design of improved organisms carrying out these processes.

Life Sciences PART grading scale: Excellent

Original:

Definition of "Excellent"—develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated experimentally by the use or reengineering of that community based on model predictions.

Replacement proposed by BERAC:

Excellent: Systems biology understanding and computational models that accurately describe the capabilities and potential of key processes in microbes, microbial communities, or plants for production of biofuels, to clean up waste, or to sequester carbon are developed and validated experimentally by the use or reengineering of those microbes, microbial communities or plants based on model predictions.

Life Sciences PART grading scale: Good

Original:

Definition of "Good"—develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated by its consistency with available data.

Replacement proposed by BERAC:

Good: Systems biology understanding and computational models that accurately describe the potential of key microbes, microbial communities, or plants for production of biofuels, to clean up waste, or to sequester carbon are developed and validated by their consistency with available data.

Life Sciences PART grading scale: Fair

Original:

Definition of "Fair"—develop a number of the components of a computational model that could accurately describe the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen.

Replacement proposed by BERAC:

Fair: Systems biology understanding and computational models that describe the potential of key microbes, microbial communities or plants for production of biofuels, to clean up waste, or to sequester carbon are developed but are not yet validated.

Life Sciences PART grading scale: Poor

Original:

Definition of "Poor"—produce a modest output of computational research that could lead to the development of components of a model to describe the potential of microbial communities to clean up waste, sequester carbon, or produce hydrogen.

Replacement proposed by BERAC:

<u>Poor</u>: Systems biology understanding of the potential of key microbes, microbial communities or plants for production of biofuels, to clean up waste, or to sequester carbon is developed but robust computational models describing these systems are not developed.

Replacement metrics proposed by BERAC:

- Answers to scientific questions relevant to DOE mission areas are developed by funded research.
- New molecular components identified in genetic and biochemical pathways relevant to DOE mission.
- New engineered plant traits and microbial capabilities developed.
- Development of fundamental knowledge that results in strategies and methodologies that impact enhanced feedstock production and facilitation of conversion of biomass to biofuel.
- Improvements in potential yield from a particular feedstock.

- Potential improvements in the efficiency of conversion, as determined by identification and modification of key steps in the conversion.

- Decrease in energy costs required to process feedstock to fermentation substrates.

- Papers published on research supported by this program.

- Patents filed for research supported by this program.

Summary of ASCAC concern:

Developing a computational model that accurately describes the potential of a microbial community (or plant) to clean up waste, sequester carbon, or produce hydrogen (or other biofuels)—and then validating this computational model experimentally by the use or reengineering of that community (or plan) based on model predictions—is an ambitious goal to meet by 2015.

Will the addition of the charge "*System biology understanding*" split the focus and thus hinder progress toward the original goal?