Report of the Biological and Environmental Research Advisory Committee (BERAC)

Review of the Life Sciences PART Measure: (Performance Assessment and Rating Tool)

Progress and Recommendation for Modification

October 16, 2006

At its July 10, 2006, meeting, the Biological and Environmental Research Advisory Committee (BERAC) was asked by the Office of Biological and Environmental Research (BER) to assess progress on the long term research goals of the Office, as defined in the Office of Science Strategic Plan and through the Federal Program Assessment Rating Tool (PART). A subcommittee of BERAC, chaired by Dr. Christopher Somerville, undertook the assessment of the progress of the Life Sciences Program in meeting its long term goals. The subcommittee membership is found in Appendix A. The three tasks that were assigned to this subcommittee are:

Task 1 – Assess BER's progress toward its current long term Life Sciences PART measure.
Task 2 – Assess the appropriateness of BER's current Life Sciences PART measure.
Task 3 – Assess how well BER is positioned to make progress toward the proposed new long term Life Sciences PART measure.

The subcommittee responses to these tasks are as follows:

Task 1 – Assess BER's progress toward its current long term Life Sciences PART measure.

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

BERAC Assessment: In February 2006, the National Research Council (NRC) of the National Academies issued its "Review of the Department of Energy's Genomics: GTL Program." The NRC Committee was asked and responded to the following question:

> ***Question 1:*** *Is the Genomics: GTL program, as currently designed, scientifically and technically well tailored to the challenges faced by the DOE in energy technology development and environmental remediation?*
>
> ***Answer 1:*** *Yes, the use of systems and synthetic biology approaches in the Genomics: GTL program to address some of the most pressing issues in microbial genomics relevant to DOE's mission in energy security, environmental remediation, and carbon cycling and sequestration is not only appropriate but necessary. The study of individual components only does not provide knowledge on systems integration at the level of pathways, organisms, and microbial consortia—for example, on the effects of introducing new metabolites or new or engineered organisms to a community or on organism or community responses. Systems biology research is needed to develop models for predicting the behavior of complex biological systems, to engineer microorganisms for bioremediation and energy-related needs, and to understand carbon cycling.*

Current and planned research of the Genomics: GTL program promises to provide the predictive understanding of microorganisms needed to:

- *Develop affordable and reliable carbon-neutral energy alternatives from plants and microorganisms.*
- *Develop biological solutions to the many recalcitrant problems of legacy wastes.*
- *Increase understanding of the role of microbial communities in global carbon cycling to enable the development of carbon-sequestration techniques for addressing climate change.*

The committee endorses DOE's use of a systems approach to achieve its mission goals through Genomics: GTL and supports its plan to enlarge funding of the program to $200 million per year for basic research. The committee suggests that plant biology research be included in the Genomics: GTL 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. The committee's suggestion is consistent with the Energy Basic and Applied Sciences Act of 2005, which calls for an emphasis on both plants and microorganisms in the program. Systems biology research on plants and microorganisms is not likely to be conducted on a large scale without DOE's visionary thinking. Because the productive applications of new technologies to advance science will be hampered by the lack of appropriate tools, the committee finds that the concept of infrastructure for research and technology development offers a logical and even necessary pathway for achieving DOE's research goals.

Recommendation 1: The committee recommends that DOE and the nation give high priority to genomics research aimed at achieving DOE's mission goals.

The NRC Committee went on to identify issues to be addressed by the Genomics: GTL program.

A variety of issues will need to be addressed in the course of achieving the long-term goals of Genomics: GTL. Among these is the need to improve and implement genomics-enabled, high-throughput studies of genetic diversity in Genomics: GTL environments. The resulting information would contribute greatly to understanding aspects of ecosystem-level population biology, evolution, and function that are currently lacking and are critical to the mission of the Genomics: GTL program. The following are insights:

- *Description and then development of predictive models for how complex microbial consortia respond to natural and imposed selection.*
- *Identification of the genomic diversity best suited to manipulation of Genomics: GTL target processes, for example, remediation of specific contaminants in unique environments.*

- *Characterization of genotypes and the genes and proteins that most strongly influence system function.*
- *Understanding how human intervention may alter community structure and function, and identifying and quantifying related risk factors, if any. In particular, if genetically modified organisms are to be released into open field settings for bioremediation, DOE should make strong efforts to gain public acceptance for such release.*

Central to the Genomics: GTL mission is the need to identify the molecular machines that underlie target processes. The challenge is not simple, in that what we conceive of as distinct molecular entities may exist on any of a number of scales, from coherent protein complexes, to physically unrelated complexes in a single cell, to proteins present in unrelated taxa, but where complementary activities yield a desired outcome. Major challenges include the following:

- *Identification and functional characterization of the proteins and the complexes that underlie Genomics: GTL target processes.*
- *Formulation of models that predict the function of these key cellular or organismal components in situ.*
- *Development of strategies to improve the efficiency of these "molecular machines."*
- *Improving methods for analysis and interpretation of gene and protein function in heterologous systems, including both computational and experimental approaches.*

Much of the progress envisioned under Genomics: GTL will require derivation and application of novel technologies, principles, and computational approaches that permit biologists and engineers to understand and manipulate the Genomics: GTL ecosystems. Key milestones toward this broad goal are:

- *Improved technologies for surveying taxonomic and genetic diversity in target environments, including the development of tools for both culture-dependent and culture-independent methods and strategies to deal with ultrarare genomes.*
- *Development of experimental tools, concepts, and mathematical methods that can model transient and stable states and identify the control points for particular system parameters.*
- *Establishment of predictive models of microbial behavior during discrete phases of development and in response to external biotic and abiotic stimuli.*
- *Development of new methods and instrumentation to measure key biological parameters that may be relevant to system function, including metabolite flow and protein function in vivo and in situ.*
- *Establishment of methods to reproduce native ecologies in the laboratory or to analyze them in situ.*
- *Understanding of the consequences and frequency of events that may alter population function, such as horizontal gene transfer, alterations of physical-chemical environments, and introduction of nonnative species.*

> *Broadly stated, the goal of systems biology is to uncover properties of organisms and communities that would not be made apparent by analysis of their components in isolation. Few would argue that our current understanding and methods are adequate to develop a quantitative model of even one bacterium, much less a collection of genotypes in a single species, and even less an entire ecosystem. Systems biology suffers from a dearth of general principles that can guide further study.*

This discussion by the NRC committee supports the overall strategy being taken by the Genomics: GTL program to develop computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen. To date, the Genomics: GTL program has supported a wide variety of research projects focused on developing models for various components of microbial systems, tools and strategies needed for system modeling, biological resources underpinning the development of these models and the development and use of models to better predict and understand biological systems. These projects represent a robust portfolio and appropriate combination of experimentation, modeling, and tool development for modeling. Examples include:

- Arkin, Lawrence Berkeley National Laboratory, Rapid Detection of Stress Response Pathways in Metal/Radionuclide Reducing Bacteria
- Church, Harvard, Microbial Ecology, Proteogenomics, and Computational Optima
- Lovley, University of Massachussetts at Amherst, Genome-Based Models to Optimize In Situ Bioremediation of Uranium and Harvesting Electrical Energy from Waste Organic Matter
- Church, Gene Network Sciences, Computation Hypothesis Testing: Integrating Heterogeneous Data and Large-Scale Simulation to Generate Pathway Hypotheses
- Collins, Boston University, Rapid Reverse Engineering of Genetic Networks via Systematic Transcriptional Perturbations
- Hood, Institute for Systems Biology, Development of Advanced Tools for Data Management, Integration, Analysis and Visualization Through a Comprehensive Systems Analysis of the Halophilic Archaeon
- Lawrence, Brown University, Development of Bioinformatics and Experimental Technologies for Identification of Prokaryotic Regulatory Networks
- Sauro, Keck Graduate Institute, Computational Resources for GTL
- Colvin, University of California at Merced, A Center for Computational Biology
- Mitchell, University of Wisconsin at Madison, BACTER: Bringing Advanced Computational Techniques to Environmental Research
- Paulaitis, The Johns Hopkins University, Institute for Multiscale Modeling and Analysis of Complex Interactions in Biology
- Fredrickson, Pacific Northwest National Laboratory, The Shewanella Federation
- Davidson, California Institute of Technology, Animal Gene Regulatory Networks
- Kumar, Defense Advanced Research Project Agency Projects at Virginia Polytechnic Institute (Jigcell Simulator), Lawrence Berkeley National Laboratory (Spatial Modeling Tasks), SRI International (Database Warehouse For Bio-Spice)

- McAdams, Stanford University, Global Characterization of Genetic Regulatory Circuitry Controlling Adaptive Metabolic Pathways

As noted in the NRC report, there are tremendous challenges remaining on all fronts before the Genomics: GTL program can efficiently use "computational models to direct the use and design of microbial communities to clean up waste, sequester carbon, or produce hydrogen." These include the ability to develop more robust, predictive models of individual microbes and then to extend that capability to microbial communities. In addition to the modeling research noted above, Genomic: GTL's investments in metagenomic research are laying a sound foundation for future modeling of microbial communities. BERAC believes that the program is on target to achieve a long term grade of Excellent for this measure as it is currently defined.

As originally designed, the Genomics: GTL Program intended to construct four high – throughput facilities that would be focused on: (1) protein production; (2) molecular machines); (3) proteomics; and (4) cellular systems. The NRC report concluded that

> *…building four single-purpose facilities may not be the most effective way to meet the multiscale and multidisciplinary challenges of systems biology … The committee strongly encourages DOE to rethink its user-facility construction plans and to consider the creation of up to four integrated facilities. Each facility will combine the capabilities of the original planned facility types in a vertically integrated manner so that it can tackle all aspects of a problem or small set of problems in parallel and potentially achieve goals more quickly. The first vertically integrated facility would focus on one or two of DOE's mission goals, such as bioenergy …*

In this regard, BERAC notes that the Genomics: GTL Program has also made substantial progress toward its long term goal to "characterize the multi protein complexes (or the lack thereof) involving a scientifically significant fraction of a microbe's proteins." However, this component of the long term measure is no longer within the highest priority goals of Genomics: GTL, as based on the NRC recommendation not to develop high throughput user facilities for protein production and isolation and characterization of molecular machines. Because this component of the long term measure had been consonant with the original vision of the GTL facilities, it is appropriate to evaluate the initial progress toward this goal. It should also be noted that the understanding of multiprotein complexes will remain important to the overall goals of the program.

The Genomics: GTL program has invested in several large ($2-5 million per year), multi-institutional projects at both national laboratories and universities that are focused on developing experimental strategies and tools for isolating and characterizing multiprotein complexes. Examples include:

- McAdams, Stanford University, Dynamic Spatial Organization of Multi-Protein Complexes Controlling Microbial Polar Organization, Chromosome Replication, and Cytokinesis

- Biggin, Lawrence Berkeley National Laboratory, High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in Desulfovibrio Vulgaris
- Tainer, Lawrence Berkeley National Laboratory, Molecular Assemblies, Genes, and Genomics Integrated Efficiently (MAGGIE)
- Buchanan, Oak Ridge National Laboratory, Genomes to Life Center for Molecular and Cellular Systems

While three of these four projects are relatively new, the Buchanan project has developed a small demonstration pipeline that will have a capacity of at least 32 assays per week by the end of FY 2007 (approximately 1600 multi protein complex targets in FY 2007) with plans to double that capability in FY 2008. The Genomics: GTL program has made some excellent research investments to develop tools and strategies for characterizing large numbers of multiprotein complexes. If the program had continued with its plans to develop a high throughput user facility for the isolation and characterization of multiprotein complexes, it would likely have received a long term grade of Excellent for this original measure.

Task 2 – Assess the appropriateness of BER's current Life Sciences PART measure.

BER's current Life Sciences PART measure has two separate goals: (1) characterize the multiprotein complexes (or the 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.

As noted above and in the NRC report on the Genomics: GTL Program, 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 change (see below) in the first element. It should be noted that 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.

The first element of the Life Sciences 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. 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 relevant to DOE missions. The first two of these centers will focus on research directed towards helping to make biofuels cost-effective alternatives to fossil

fuels. In August 2006, the Office of Science issued a Funding Opportunity Announcement for two Bioenergy Research Centers that will be funded at approximately $25 million per year each for five years. Funding decisions for these Bioenergy Research Centers will be made in mid 2007 so details of the research foci will not be known until then. However, 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.

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.

BERAC proposes the following as a replacement for the Life Sciences PART measure:

> *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.*

Why is this measure important? The proposed microbial and plant research that focuses on basic, molecular-level processes in nature offers tremendous promise for enabling technology that will lead to a safer, stronger, healthier and more secure world. Microbes can thrive in the most inhospitable environments on earth, including highly contaminated waste sites, and they have almost every imaginable capability, including the ability to use toxic wastes as sources of food and fuel. In the case of organic wastes, environmental engineers are already using bacteria as chemical factories for cleanup. Plants are already used as feedstocks for the production of biofuels such as ethanol from cellulose; however, the processes used in this conversion are difficult, energy intensive, and not cost effective or competitive with gasoline from fossil fuels. Genomics: GTL brings the potential for designer plants whose tough lignin and cellulose are much more accessible and easier to break down to sugars than is the case in native species. Genomics: GTL also brings the potential for designer microbes or communities of microbes to accelerate the degradation process and to produce fuels other than ethanol. These advances would make cellulosic-derived fuels plentiful and affordable alternatives for a significant fraction of the gasoline we use today.

This understanding may lead to:

- introduction of engineered traits in existing plant feedstocks
- development of new plant feedstocks for biofuel production or carbon sequestration
- improvements in feedstock deconstruction and decomposition
- introduction of engineered traits in organisms carrying out bioconversion, carbon capture, or bioremediation

- development of improved fermentation, bioconversion, or bioremediation organisms
- consolidation of biological processes for improved efficiency of biomass conversion, carbon capture, or bioremediation
- the development of new types of liquid fuels

BERAC proposes the following "grading" scale for the revised PART measure:

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.

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.

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.

Poor: 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.

Expert Review every three years will rate progress towards achieving the long term measure, using the excellent – poor rating scale. This review will rely on a variety of metrics, as identified below. It is critical to note that some of these metrics, especially those that deal with changes in feedstock and conversion rates, will only have meaning after basic research has been conducted over several years. Also, because of the relatively dramatic change in some aspects of the Genomics: GTL through the inclusion of the biofuels focus, these metrics may need to be modified as the program gains more experience with biofuels-related research. The metrics that are proposed at this time include:

- 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

Task 3 – Assess how well BER is positioned to make progress toward the proposed new long term Life Sciences PART measure.

For the following reasons, the BER program is well positioned to make progress toward the new, proposed PART measure:

First and foremost, the scientific contributions that will derive from the research conducted by the two new Bioenergy Research Centers will, by definition, contribute directly to the PART measure. While the specific scientific focus of these two research centers will not be known until sometime in mid-2007, each Center will have deliverables/benchmarks and ongoing scientific and management reviews that will help focus the objectives of the research to the proposed short, intermediate and long term goals. This management strategy, in addition to the very competitive process that will result in the selection of these Research Centers, make it highly likely that these Centers will make substantial scientific contributions that enable the BER program to make progress toward the proposed new long term Life Sciences PART measure.

Second, as described above in Task 1, the Genomics: GTL program has already made a good start by developing a robust research portfolio and appropriate combination of experimentation, modeling, and tool development for modeling.

Third, the fundamental systems biology research of the entire Genomics: GTL program, from the genomic sequencing of individual microbes, microbial communities and the poplar tree to the development of imaging tools to new strategies for studying biological systems provide a solid underpinning for all Genomics: GTL research, including progress toward the Life Science PART measure.

Finally, although the Genomics: GTL program has only recently expanded its vision to include the very essential plant research, BER has already begun to invest in the type of plant research that is critical to the long term success of the Life Sciences PART measure. The Joint Genome Institute has expanded its repertoire of sequencing targets to include the large, complex genomes of a variety of plant species, including sorghum and soybean. In late 2005, BER partnered with the U.S. Department of Agriculture to issue a solicitation – "Plant Feedstock Genomics for Bioenergy: A Joint Research Solicitation-USDA, DOE" – for genomics-based research that will lead to the improved use of biomass and plant feedstocks for the production of fuels such as ethanol or renewable chemical feedstocks. The initial portfolio of merit-based awards includes plant feedstocks ranging from poplar, alfalfa, sorghum, wheat and other grasses, with several

projects leveraging genome sequence information generated at the JGI. This solicitation will be reissued later in 2006, with awards expected in FY2007. This is an important first step in integrating plant research into the Genomics: GTL and BER portfolio.

Although it is too early to measure the success of BER's new Genomics: GTL initiatives, including the Bioenergy Research Centers and investments in plant feedstock genomics research, BERAC is confident that the Genomics: GTL Program is well poised to allow the program to earn a long term rating of excellent on the Life Sciences PART measure.

Appendix 1

BERAC Subcommittee Membership

Dr. Christopher R. Somerville (Chair)
Department of Plant Biology
Carnegie Institution
260 Panama Street
Stanford, CA 94305

Dr. Michelle S. Broido (BERAC Chair)
Associate Vice Chancellor for Basic Biomedical Research,
and Director, Office of Research, Health Sciences
University of Pittsburgh
Scaife Hall, Suite 401
3550 Terrace Street
Pittsburgh, PA 15261

Dr. John Pierce
DuPont Central Research
Experimental Station, E328/251
P.O. Box 80328
Wilmington, DE 19880

Dr. Margaret A. Riley
Morrill Science Center, South 306
University of Massachusetts, Amherst
Amherst, MA 01003

Dr. Melvin I. Simon
California Institute of Technology
M/C 147-75
1200 East California Blvd.
Pasadena, CA 91125