DOE/SC-0097

U.S. Department of Energy's

Genomics:GTL Bioenergy Research Centers

Office of Science

GTL Bioenergy Research Centers

White Paper

DOE GENOMICS:GTL
SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT

OFFICE OF SCIENCE
U.S. DEPARTMENT OF ENERGY

August 2006

Contacts

DOE Office of Biological and Environmental Research (OBER)

Chief Scientist: David Thomassen
david.thomassen@science.doe.gov, 301-903-9817

Staff: www.sc.doe.gov/ober/staff.html

Web sites

Genomics:GTL Program
genomicsgtl.energy.gov

GTL Bioenergy Research Centers
genomicsgtl.energy.gov/centers/

GTL Bioenergy
genomicsgtl.energy.gov/biofuels/b2bworkshop.shtml

GTL Roadmap
genomicsgtl.energy.gov/roadmap/

GTL Image Gallery
genomics.energy.gov/gallery/

DOE Office of Biological and Environmental Research
www.science.doe.gov/Program_Offices/BER.htm

DOE Office of Science
www.science.doe.gov

DOE/SC-0097

U.S. Department of Energy's
Genomics:GTL Bioenergy Research Centers
White Paper

In his Advanced Energy Initiative announced in January 2006, President George W. Bush committed the nation to new efforts to develop alternative sources of energy to replace imported oil and fossil fuels. Developing cost-effective and energy-efficient methods of producing renewable alternative fuels such as cellulosic ethanol from biomass and solar-derived biofuels will require transformational breakthroughs in science and technology. Incremental improvements in current bioenergy production methods will not suffice.

The Genomics:GTL Bioenergy Research Centers will be dedicated to fundamental research on microbe and plant systems with the goal of developing knowledge that will advance biotechnology-based strategies for biofuels production. The aim is to spur substantial progress toward cost-effective production of biologically based renewable energy sources. This document describes the rationale for the establishment of the centers and their objectives in light of the U.S. Department of Energy's mission and goals.

U.S. Department of Energy
Office of Science

Office of Biological and Environmental Research
Genomics:GTL Program

August 2006

Contents

Introduction

Since 2000, the Office of Biological and Environmental Research (BER) in the Department of Energy's (DOE) Office of Science (SC) has defined Genomics:GTL (GTL) as its highest-priority program. GTL builds on the successes of the Human Genome Project (HGP), initiated by DOE in 1986. A systems biology research program focusing on microbes and plants, GTL has the mission goal of tapping the powerful and diverse capabilities of nature—including microbes, microbial communities, and plants—to provide breakthrough biotechnologies for renewable energy production, carbon sequestration, and environmental remediation. While GTL has supported substantial research relevant to all three missions over the past 6 years (genomicsgtl.energy.gov), BER recognizes that developing scientific infrastructures to accelerate GTL's research toward the goals of urgent national needs (e.g., bioenergy) is vitally important.

In August 2005 BER, in collaboration with SC's Advanced Scientific Computing Research program, published *DOE Genomics:GTL Roadmap: Systems Biology for Energy and Environment* (U.S. DOE 2005). The document provided a comprehensive overview of the GTL program and set forth a plan to develop and deploy four major scientific facilities for GTL. These four facilities were to be organized by function, with one each for production and characterization of proteins and molecular tags, characterization and imaging of molecular complexes, whole-proteome analysis, and analysis and modeling of cellular systems.

As a detailed description of GTL's scientific challenges, approaches to confront these challenges, required technologies, and DOE needs and applications, the GTL Roadmap remains a valuable guide. In early 2006, however, DOE SC began to rethink its original plan and ultimately decided to revise the facilities described in the GTL Roadmap. The decision came in response to two developments.

First, in his 2006 State of the Union address, President George W. Bush announced the Advanced Energy Initiative (AEI), with a national goal of substantially reducing U.S. dependence on imported oil in the years ahead (AEI 2006). A major focus of AEI is on accelerating research for developing cost-effective fuels from biomass, particularly ethanol from cellulose.

Second, in February 2006, a committee of the National Research Council (NRC) of the National Academies issued *Review of the Department of Energy's Genomics: GTL Program* (NRC 2006). The NRC review strongly endorsed the importance of systems biology to DOE missions in energy, environment, and climate protection and praised the GTL program and its research achievements to date, recommending that the Department and the nation give special importance to GTL research. The panel, however, suggested a reorientation of the facilities plan, which it argued would enable the GTL program to achieve its research and mission goals more expeditiously. Rather than a succession of facilities organized around function, the NRC review recommended four research institutes. Each institute would incorporate, under a single roof, many or all technological capabilities planned for the original four facilities in a "vertically integrated" fashion as needed to achieve the institute's research goals. Each institute would be organized not around a function such as protein production but rather around a particular science theme or application such as bioenergy, carbon sequestration, or environmental remediation. NRC panelists argued that such a configuration—integrating necessary capabilities to study and manipulate biological processes at the molecular, cellular, and community system levels—also would enable GTL research programs to benefit more immediately from recent technological advances in such functions as producing, detecting, and imaging proteins and observing metabolites, cellular components, and whole cells.

The timing of the NRC report was propitious in light of the new urgency that the President's Advanced Energy Initiative had given to the Department's mission to develop alternative, renewable sources of energy. After detailed consultations with NRC panel members and others, DOE SC concluded that the recommendations were both sound and timely. DOE SC began developing new plans to deploy up to

two new Bioenergy Research Centers (hereafter Centers) devoted to research on microbes and plants, with the purpose of making substantial progress toward cost-effectively producing biologically based renewable energy sources. SC expects to issue a Funding Opportunity Announcement in 2006 for establishing up to two Bioenergy Research Centers. Universities, national laboratories, nonprofit agencies, and private firms, as well as consortia of two or more such institutions, will be eligible for funding to establish and operate a Center.

Basic Concept

Developing energy-efficient and cost-effective methods of producing alternative fuels such as cellulosic ethanol from biomass will require transformational breakthroughs in science and technology. Consequently, a major initial focus of research will be on addressing scientific and technical roadblocks to producing ethanol from cellulose cost-effectively and in quantities that can contribute substantially to the nation's transportation fuel supply. The Department also is interested in supporting well-directed research on other biofuels from biomass, including biodiesel, biofuels for aviation, and biologically produced or inspired hydrogen and other fuels from sunlight. The purpose of the new Bioenergy Research Centers will be to give greater impetus and focus to developing usable knowledge that will advance biotechnology-based strategies for biofuel production and ultimately lead to technologies deployable in the nation's energy economy.

As envisioned by the Department, the Centers are to be led by internationally recognized scientists and staffed by multidisciplinary teams incorporating a range of skills that might include genomics, microbial and plant biology, genetics, proteomics, physiology, biochemistry, structural and computational biology, bioinformatics, engineering, and others. All Centers are to be equipped with a full suite of instrumentation and tools essential to modern biology and biotechnology research. In addition, they will have access to the extremely powerful instruments and facilities available to the scientific community throughout DOE's national laboratories, including the DOE Joint Genome Institute's capabilities for high-throughput, whole-genome sequencing; advanced high-intensity light sources such as the Advanced Photon Source at Argonne National Laboratory; the Advanced Light Source at Lawrence Berkeley National Laboratory; the National Synchrotron Light Source at Brookhaven National Laboratory; the SPEAR III and Linac Coherent Light Source, the latter now under construction at the Stanford Linear Accelerator Center; the Spallation Neutron Source at Oak Ridge National Laboratory; and the Department's cutting-edge supercomputer capabilities, in which DOE now leads the civilian world. The Centers also will develop new technologies as needed for research progress, for example, production of proteins that are particularly difficult to isolate but critically important to the success of their research.

A Revised Methodology

Revision of the GTL facilities plan represents a shift in the Department's thinking on the scientific methodology underpinning the overall GTL program. The research paradigm set forth in the GTL Roadmap is essentially one of developing and using large-scale facilities and high-throughput technologies to generate masses of data on the components of multiple organisms. The Roadmap referred to this process as "discovery-driven science," contrasting it with traditional "hypothesis-driven science." In the Human Genome Project, DOE modeled this pioneering new approach to biology and to science in general. Rather than pursuing discovery stepwise at the laboratory bench via a series of hypothesis-driven experiments, DOE sought to radically accelerate the pace of discovery through the creation of facilities and high-throughput instrumentation that would rapidly generate large amounts of data. HGP's success in sequencing the human genome ahead of schedule and under budget was a direct consequence of this approach.

Under the GTL Roadmap paradigm, GTL was envisioned as continuing to pioneer the discovery-driven scientific methodology. In this concept, high-throughput instrumentation would generate data on genes,

proteins, and other components of thousands of organisms. The data would be gathered in the GTL knowledgebase to be mined and analyzed by means of supercomputers. Such algorithm-based analysis would provide insights into component function and, in turn, generate new hypotheses to form the basis of GTL-funded research carried out by a large and diverse group of scientists. In the words of the GTL Roadmap, the paradigm would combine "(1) discovery science as we navigate huge unexplored data sets that can reveal unforeseen properties and phenomena and (2) computationally driven hypothesis science to derive insights into previously unfathomable complexity."

Although the Department continues to believe in the importance of building the GTL Knowledgebase of systems biology data and analyzing and coordinating the data via high-end computation, it agrees with the NRC panel that this approach alone would be unlikely to yield results relevant to the Department's energy mission–related goals within a reasonable time. The GTL program will continue to support discovery-driven scientific research, but the Department agrees that, to address its energy mission effectively and in a timely manner, the GTL Centers will need to employ "a more problem-oriented approach."

The Centers' task will be to carry out fundamental research to identify key bottlenecks and roadblocks on the path to cost-effective production of fuels from biomass and to bring to bear an aggressive, creative, and multidisciplinary approach in overcoming these barriers. The focus on fundamental research is critical since cost-effective biological production of fuels, as well as the other key BER missions, presents significant challenges, with many unanswered questions that require solutions. DOE and the Office of Science have extensive experience with this kind of challenge, stemming from (1) many years of successfully supporting and managing large, interdisciplinary teams of researchers on major, mission-oriented scientific projects; (2) a world-leading suite of instruments and facilities for high-throughput genome sequencing, observation of cellular components and chemical reactions at the molecular level and at ultrafast time scales, and advanced computer modeling and simulation; and (3) a history of developing state-of-the-art technologies and tools to tackle difficult scientific challenges. Along with benefiting from the recent rapid advances in biotechnology instrumentation, largely driven by universities and the private sector, the Bioenergy Research Centers will be in a position to take advantage of DOE's advanced resources.

Objectives and Requirements

As stated earlier, developing energy-efficient and cost-effective methods of producing alternative fuels such as cellulosic ethanol from biomass will require transformational breakthroughs in science and technology. Incremental improvements in current bioenergy-production methods will not suffice. The focus on microbes (for cellular mechanisms) and plants (for source biomass) fundamentally exploits capabilities well known to exist in the microbial world. Thus "proof of concept" is not required, but considerable basic research into these capabilities remains an urgent priority.

Several developments have converged in recent years to suggest that systems biology research into microbes and plants promises solutions that will overcome critical roadblocks on the path to cost-effective, large-scale production of cellulosic ethanol and other renewable energy from biomass. The ability to rapidly sequence the DNA of any organism is a critical part of these new capabilities, but it is only a first step. Other advances include the growing number of high-throughput techniques for protein production and characterization; a range of new instrumentation for observing proteins and other cell constituents; the rapid growth of commercially available reagents for protein production; a new generation of high-intensity light sources that provide precision imaging on the nanoscale and allow observation of molecular interactions in ultrafast time intervals; major advances in computational capability; and the continually increasing numbers of these instruments and technologies within the national laboratory infrastructure, at universities, and in private industry. All these developments expand our ability to elucidate mechanisms present in living cells, but much more remains to be done.

The Centers are designed to accomplish GTL program objectives more rapidly, more effectively, and at reduced cost by concentrating appropriate technologies and scientific expertise, from genome sequence to an integrated systems understanding of the pathways and internal structures of microbes and plants most relevant to developing bioenergy compounds. The Centers will seek to understand the principles underlying the structural and functional design of selected microbial, plant, and molecular systems. This will be accomplished by building technological pathways linking the genome-determined components in an organism with bioenergy-relevant cellular systems that can be characterized sufficiently to generate realistic options for biofuel development. In addition, especially in addressing what are believed to be nearer-term approaches to renewable energy (e.g., producing cellulosic ethanol cost-effectively and energy-efficiently), the Center research team must understand in depth the current industrial-level roadblocks and bottlenecks (see section, GTL's Vision for Biological Energy Alternatives, below). For the Centers, and indeed the entire BER effort, to be successful, Center research must be integrated with individual investigator research, and coordination of activities, from DNA sequencing to high-throughput protein development and characterization.

The Centers will have long- and intermediate-term visions in bioenergy research, but they must be sufficiently flexible to allow adjustments in research directions in response to promising developments. The Centers are envisioned to serve as catalysts for bioenergy-related research supported by the broader GTL program, coordinate with other GTL-funded projects and, in the long term, with other programs within DOE. Results of fundamental research and technology development within each Center will be integrated with other current and future GTL Research Centers and across the entire GTL program.

Research and technology development at the Centers will lie at the frontier of basic science. Although research will maintain its focus on bioenergy applications, the Centers are intended to maintain the spirit of basic research, pursuing alternative avenues and a range of high-risk, high-return approaches to finding solutions. DOE realizes that scientific problems to be addressed by the Centers are inherently interdisciplinary. The Centers will require personnel with varied skill sets including, for example, expertise in genomics, microbial and plant biology, genetics, proteomics, physiology, biochemistry, structural and computational biology, bioinformatics, engineering, and, very likely, other areas as well. To this end, the Centers will include (or coordinate access to) needed technical capabilities such as instrumentation for producing and characterizing proteins and other required materials; analytical chemistry, including specialties such as proteomics; instrumentation for characterizing microbial cells, microbial communities, and plants; and bioinformatics and computation (see Core Technological Capability Needs for GTL Research Centers, below).

To carry out their research programs, the Centers will be expected to develop core capabilities in or have access to the analytical, imaging, structural, and computational capabilities requisite for the systems biology approach to biofuel production. The Centers also are encouraged to devote resources to developing new technological capabilities for overcoming challenges that cannot be addressed with currently available technologies and instrumentation.

The Office of Science has learned through long experience that effective management of scientific programs, projects, and facilities is critical to the success of research programs. In common with other major programs supported by SC, the Centers will be subject to periodic and rigorous outside review of their scientific program and their management structure, policies, and practices, including reviews by the Office of Science Office of Project Assessment.

Role of Biology and Biotechnology in America's Energy Future

Biological processes played a key role in producing the fossil fuels so critical to meeting today's world energy demand. Fossil fuels are derived from once-living biomaterials that were trapped beneath the sediments of ancient seas by a series of geological events. Over millions of years, the right mix of heat, pressure, and other factors transformed the biomaterials into fossil fuels. With biotechnological innovations, biology once again can play an important role in producing high-energy fuels. Plants and photosynthetic microorganisms are masters at harvesting chemical energy from sunlight—a virtually inexhaustible supply of energy. By harnessing their photosynthetic and metabolic capabilities, biological systems can be used to satisfy a greater portion of energy demand.

America's energy challenge is to meet projected increases in energy demand while decreasing dependence on foreign sources of energy. From 2004 to 2030, U.S. demand is projected to grow 34%, much more than the expected increase in domestic production (EIA 2006). Making up this shortfall without importing more oil and gas will require investments in science and technologies that will improve conservation and efficiency and expand the domestic energy supply system. A primary goal of the national energy policy is not only to enlarge the domestic supply but also to broaden our range of options in ways that will reduce vulnerabilities to supply disruptions and protect the environment (NEPDG 2001).

Another key factor in America's energy challenge is rising carbon dioxide (CO_2) emissions. Carbon dioxide is the most abundant greenhouse gas (GHG) in the atmosphere, and, based on projected energy use between 2004 and 2030, U.S. emissions could increase almost 40% (EIA 2006). With accelerated growth in fossil-fuel consumption projected for developing regions of the world, by 2025 annual global CO_2 emissions could be 55% higher than in 2002 (EIA 2005). In 2002, global energy use emitted about 7 gigatons of CO_2 (GtC) into the atmosphere. Several long-term projections estimate that these emissions could more than quadruple by 2100 (Nakicenovic et al. 2000). Stabilizing carbon dioxide concentrations at any level requires that global emissions derived from human activity (i.e., anthropogenic) must be controlled, begin a long-term decline, and ultimately fall to virtually zero. To this end, a variety of breakthrough energy technologies will be needed to significantly reduce GHG emissions, an area where biology can play a significant role.

DOE's strategic energy goal is to "protect our national and economic security by promoting a diverse supply and the delivery of reliable, affordable, and environmentally sound energy" (U.S. DOE 2003). GTL supports this goal by providing the fundamental scientific knowledge needed to develop biological technologies that produce clean, renewable, carbon-free, and carbon-neutral alternatives to fossil fuels. These biotechnical advances will expand and improve our current portfolio of options and ultimately help lead to a secure, sustainable energy future for the United States and the world (U.S. DOE 2005). GTL's mission science goals are to (1) understand the principles underlying the structural and functional design of microbial and molecular systems and relevant plant systems and develop the ability to model, predict, and engineer optimized enzymes, microorganisms, and plants for the production of biologically based renewable energy sources; and (2) analyze thousands of natural and modified variants of such processes as cellulose degradation, fermentative production of ethanol or other fuels, and biophotolytic hydrogen and fuel production.

Applying biology to build a new U.S. bioenergy industry can benefit this nation's energy security, economy, and environment in many different ways. Biofuels, especially ethanol from plant material (biomass), have the potential to reduce our dependency on foreign oil in the transportation sector and diversify our energy-technology portfolio. As renewable alternatives that can be harvested on a recurring basis, bioenergy crops (e.g., poplar trees and switchgrass) and agricultural residues (e.g., corn stover and wheat straw) can provide

American farmers with important new sources of revenue. Consumption of biofuel produces no net CO_2 emissions, releases no sulfur, and has much lower particulate and toxic emissions than fossil fuels (Greene et al. 2004). In addition to ethanol, other biobased energy alternatives include biodiesel, butanol, methanol, hydrogen, and methane.

Biomass is used to meet only 3% of current U.S. energy consumption (EIA 2006). In 2005, the United States produced 4 billion gallons of ethanol from corn grain, enough to meet about 2% of the nation's gasoline consumption (RFA 2006; EIA 2006). Ethanol from biomass has promise for meeting a significantly larger portion of gasoline demand, but higher production costs, technical difficulties, inefficiencies in biomass conversion, and uneven distribution across the continental United States currently prevent ethanol from being cost-competitive with gasoline.

Another concern has been the uncertainty in determining how much land must be dedicated to bioenergy crops to make a real difference in transportation fuel demand and how this would impact agricultural and forestry practices. About 60 billion gallons of ethanol would be needed to displace 30% of the nation's current gasoline consumption. Producing this much ethanol would require an annual supply of roughly a billion dry tons of biomass feedstocks. A recent report prepared for the U.S. Department of Agriculture and DOE has projected that relatively modest changes in the use of farmlands and forests could produce more than 1.3 billion dry tons of biomass per year (Perlack et al. 2005). As research improves efficiencies in vehicles, agricultural production, and biomass conversion, the availability of land and sunlight in the United States should be sufficient to produce a significant portion of the biofuels needed to meet domestic transportation-related demand without disrupting agricultural land use for food and fiber crops.

In addition to reducing our dependence on oil, biofuels also have great potential for decreasing GHG emissions associated with fossil-fuel collection and consumption. A transition to such large-scale use of biofuels and biotechnologies could create a new bioenergy industry potentially worth trillions of dollars over the 21st Century.

Before biomass and biotechnologies can compete successfully with established energy sources for market share, much basic research is needed for a more complete understanding of the biological processes underlying biofuel production. Applying this understanding in innovative ways will enable the development of breakthrough technologies.

Genomics and Systems Biology: Microbes

DOE's Office of Science, as noted, has played a major role in inspiring, supporting, and guiding the biotechnology revolution of the last 25 years. DOE's Office of Health and Environmental Research, precursor to BER, was the first federal agency to provide directed funding to the Human Genome Project and was one of three major participants in sequencing the human genome. The BER Microbial Genome Program supported fundamental environmental microbiology research and the genome sequencing of more than 200 microbes, many with direct relevance to the energy mission. BER's GTL program, with its mission of systems biology research for renewable energy production, carbon sequestration, and environmental remediation, is the direct heir to the human and microbial genome programs. Since 2000, GTL has supported a mix of fundamental biological research and the development of novel technological capabilities at institutions across the nation. This research has focused on the diverse biochemical capabilities of microbes and microbial communities and plants as they relate to potential biological solutions to DOE mission challenges.

Every organism's genome encodes its ability to reproduce, metabolize, and sustain life. Genome data provide the foundation for studying biological processes rather than examining isolated components. The longstanding successful approach to biological research—variously described as "single gene," "reductionist," or "linear"—is piecemeal and, while productive, is not particularly efficient at addressing questions of biological complexity and integration.

Microbes can provide the basis for a biotechnology revolution in energy and environmental applications. These untapped natural treasures are the foundation of the biosphere and sustain all life on earth. Extreme genetic diversity and the ability to function in complex communities give microbes extraordinary biochemical capabilities and adaptability. The single-celled organisms are masters at living in almost every environment and harvesting energy in almost any form, from solar radiation to mineral chemistry, and transforming it into chemical compounds that power life. Furthermore, the overwhelming majority of microbes do not cause disease; rather, they enable life to exist. By understanding how microbes function in their many environments and by continuing to explore their diversity, we can reveal their contributions to earth ecosystems and gain access to an extraordinarily vast living library of genetic potential. We also can understand how microbes can provide the basis for environmental remediation and for creating new sources of renewable, less-polluting energy sources and new generations of processes for industrial application.

The new approach for exploration—systems biology—will allow us to envision the microbe as a complete set of intersecting processes and to create models for simulating microbial operation and response. This is a major first step toward illuminating the most fundamental principles of living cells and achieving a predictive understanding of the scale and complexity of natural systems. A comprehensive approach to understanding biology encompasses many cellular components. Although a genome is a fixed catalogue of information, a "parts list," it dynamically determines the machinery of a cell in response to changing environments. Thousands of genes encode even greater numbers of proteins that mediate biology in a "just-in-time" strategy by associating in myriad ways (protein "machines") to form intricate metabolic pathways within a cell.

Demonstrating the power of these finely-tuned systems, microbes rapidly respond to environmental cues by adjusting their entire cellular operation (O'Toole 2003). Microbes, in their adaptability, also have the best of both the single- and multicellular worlds. Using mechanisms we are only beginning to understand, microbes carry on a dialogue that establishes community and environmental awareness and enables individual microbes to function together as multicellular organisms (e.g., "biofilms") in complex geochemical environments. Functions include sensing the environment; assembling appropriate cells or communities of cells as environmental conditions change; regulating and carrying out cellular functions, including critical energy capture and manipulation; and providing for reproduction, sporulation, or senescence as conditions dictate (Check 2002).

Microbes are not merely reactive to environmental changes; they can actively cause (and in the past have caused) enormous changes to earth's environment. The most dramatic example is the introduction of oxygen by cyanobacteria some 2.3 billion years ago, thus enabling oxygen-using (aerobic) forms of life that led to the complex multicellular plants and animals seen today, including ourselves!

We can now rapidly and accurately decode the genomes of microbes and microbial communities in complex natural ecosystems (metagenomes), and early efforts are under way to characterize the diversity of the accompanying protein sets. While more than 99% of microbes historically have been hidden from study because they could not be cultured (Handelsman et al. 1998), genomics allows the assessment of these microbial systems to determine their makeup and some of their functionalities. Other emerging technologies such as proteomics and imaging will allow us to track molecules and cells in complex living systems to add a functional perspective without the need for classic culturing.

Genomics and Systems Biology: Plants and Biomass

Optimizing plant biomass for more efficient processing requires a better understanding of plant cell-wall structure and function. Plant cell walls contain molecules (cellulose, lignin, and hemicellulose) composed of long chains of sugars (polysaccharides) that can be converted to transportation fuels such as ethanol. This process involves using enzymes to break down (hydrolyze) the polysaccharides into their component

sugars for fermentation by microbes to ethanol. Significant challenges for efficient conversion are both the large number of enzymes required to hydrolyze diverse sugar linkages and the physical inaccessibility of these compounds due to the presence of other cell-wall components. Several thousand genes are estimated to participate in the synthesis, deposition, and function of cell walls, but very few have been identified, and little is known about their corresponding enzymes. Many questions remain, for example: how polymers such as cellulose and lignin are synthesized, how cell-wall composition is regulated, and how composition relates to the biological functions of cell walls. To answer these questions, we need to discover the functions of many hundreds of enzymes, where proteins are located within cells, whether or not they are in complexes, where and when the corresponding genes are expressed, and what genes control the expression and activities of the proteins involved. Application of new or improved biological, physical, analytical, and mathematical tools will facilitate a detailed mechanistic understanding of cell walls. That knowledge will permit optimization of various processes involved in producing biomass and converting it to fuels.

Major opportunities exist to increase productivity and conversion-process efficiencies by altering fundamental aspects of plant growth, development, and responses to biotic and abiotic stress. Altering cell-wall composition to increase the relative amount of cellulose and decrease lignin, for example, could have significant effects. Eventually a comprehensive physiological cell-wall model incorporating biophysical aspects with structural properties and knowledge about the proteins involved will aid in rational development of highly productive feedstock species whose cell walls are optimized for conversion.

A systems-level approach to understanding and manipulating plants and microorganisms central to processing biomass into liquid fuels depends on obtaining and using detailed chemical and biochemical information on organism states and structures to build functional models. A systems-level understanding of model plants will facilitate improvement of plant cell-wall composition in crops dedicated to conversion into biofuels. New approaches and tools will be necessary to characterize the detailed structures of principal types of plant cellulose and their relative energies and interrelationships with such other structural components as lignins and noncellulosic polysaccharides.

GTL's Vision for Biological Energy Alternatives

A national vision for bioenergy and biobased products was defined by the Biomass R&D Technical Advisory Committee (called BTAC): "By 2030, a well-established, economically viable bioenergy and biobased products industry will create new economic opportunities for rural America, protect and enhance our environment, strengthen U.S. energy independence, provide economic security, and deliver improved products to consumers" (BTAC 2002). GTL supports this national vision by providing a detailed understanding of the biological processes that mediate biofuel production. In addition to multiple workshops sponsored by BER over the past several years, in December 2005 BER and DOE's Office of the Biomass Program of the Office of Energy Efficiency and Renewable Energy convened a joint workshop to define barriers and challenges to rapid expansion of cellulosic ethanol production and ways to speed solutions through concerted application of modern biology tools (U.S. DOE 2006). Although the focus was ethanol, the science applies to additional fuels that include biodiesel and other bioproducts having critical roles in any such plan. Our limited understanding of many necessary microbial and plant processes presents fundamental scientific challenges that must be overcome before we can develop and use successful bioenergy technologies. This section discusses some of those challenges with respect to two areas: Cellulosic ethanol and biohydrogen production.

Ethanol from Biomass: Cellulose Degradation and Conversion

Understanding the conversion of biomass to ethanol begins with the structural and chemical complexity of the three primary polymers that make up plant cell walls: Cellulose, hemicellulose, and lignin. Depending on plant species and cell type, the dry weight of a cell wall typically consists of about 35 to

50% cellulose, 20 to 35% hemicellulose, and 10 to 25% lignin (Saha 2004). Cellulose is the most abundant biomaterial on earth. Each cellulose molecule is a linear polymer of glucose residues. Depending on the degree of hydrogen bonding within and between cellulose molecules, this polysaccharide is found in crystalline or paracrystalline (amorphous) forms. Cellulose exists within a matrix of other polymers, primarily hemicellulose and lignin. Hemicellulose is a branched sugar polymer composed mostly of pentoses (five-carbon sugars) and some hexoses (six-carbon sugars). Lignin is a complex, highly cross-linked aromatic polymer covalently linked to hemicellulose, thus stabilizing the mature cell wall. These polymers provide plant cell walls with strength and resistance to degradation, which also makes them a challenge to use as substrates for biofuel production. Enzymes, such as cellulases, hemicellulases, and other glycosyl hydrolases synthesized by fungi and bacteria, work together synergistically to degrade the structural polysaccharides in biomass. These enzyme systems, however, are as complex as the plant cell-wall substrates they attack. For example, commercial cellulase preparations are mixtures of several types of glycosyl hydrolases, each with distinctly different functions (exocellulases, endocellulases, exoxylanases, endoxylanases, cellobiases, and many others). Optimizing these enzymes will require a more detailed understanding of their regulation and activity as a tightly controlled, highly organized system; tailoring them to the plant biomass being targeted for conversion may be possible.

Today, the biochemical conversion of biomass to ethanol involves three basic steps: (1) thermochemical treatments of raw lignocellulosic biomass to make the complex polymers more accessible to enzymatic breakdown; (2) production and application of special enzyme preparations (cellulases and hemicellulases) that hydrolyze plant cell-wall polysaccharides to a mixture of simple sugars; and (3) fermentation, mediated by bacteria or yeast, to convert these sugars to ethanol. A more complete understanding of enzymes and microbes involved in biomass conversion to ethanol is needed to overcome many current inefficiencies in the production process.

A more detailed discussion of approaches to improving cellulase systems is in the GTL Roadmap. Briefly, one aim of GTL is to provide resources for screening thousands of natural and modified enzyme variants, enabling high-throughput production and functional analysis of these enzymes, elucidating regulatory controls and essential molecular interactions, and developing models for analyzing the structure and activity of natural and engineered enzyme systems. Another GTL aim is to enable integrated bioprocessing (i.e., the conversion of biomass to ethanol in a single step). Accomplishing this will require genetically modified, multifunctional organisms or stable, mixed cultures capable of carrying out all necessary biologically mediated transformations.

Ethanol from Biomass: Gaps in Scientific Understanding

To hope for any progress on these daunting challenges, we must increase our understanding of plant growth, biomass production, and microbial processes essential to bioethanol production; and develop and improve technologies based on this understanding. Scientific questions in need of further investigation (and this is a partial list at best) include the following.

What controls the synthesis and architecture of plant cell walls? The cellulosic biomass in plant cell walls consists of a complex and dynamic assemblage of celluloses, hemicelluloses, pectins, lignins, and proteoglycans. The organization and interactions among polymers of the cell wall—constructed for strength and resistance to biological, physical, and chemical attack—constitute a barrier to depolymerizing enzymes and limit subsequent bioconversion efficiency. How are cellulose microfibrils assembled and cross-linked to hemicellulose and lignin? How does the growing cellulose microfibril interact with the membrane and cytoskeleton, and what factors control cellulose fibril orientation, length, and composition? How is lignin synthesized, and can we manipulate its composition or levels?

Can we identify key traits affecting biomass yield and conversion efficiency and target them for selection and improvement? The traditional goal in plant breeding is first to identify useful genetic variations for traits of interest (e.g., disease resistance or biomass yield) by screening natural or mutagenized populations of individuals. Those with useful variations are intercrossed to produce progeny with new combinations of the desired trait. Thus, multiple rounds of breeding for a plant having a number of trait improvements become time consuming and expensive. Developing the ability to more efficiently correlate desired traits with DNA polymorphisms or markers will facilitate this process. It will allow breeders to monitor plants for a trait that can be difficult to recognize due to tissue-specific or developmental-stage–specific expression. Improved marker-assisted breeding in slow-maturing tree species, for example, could eliminate many decades of expensive breeding steps to develop more highly productive plants amenable for processing to biofuels.

What factors determine soil ecosystem function and productivity? Soils are complex ecosystems composed of unknown numbers of organisms, with the interactions among microbes, fungi, and plant roots controlling nutrient uptake. The effects of crop-residue quantity and composition on soil ecosystems are largely unknown. Systems biology analyses of highly productive soils, including their ecosystem-scale genomic characterization and an improved understanding of the plant-microbe associations, are required to ensure sustainable biomass-harvesting practices.

What is the extent of natural diversity among biomass-degrading and ethanologenic organisms? Evidence suggests that the naturally occurring diversity of microbes (and therefore of enzymes that may be useful in addressing GTL goals) is vastly greater than previously suspected (Venter 2004). Over the last 30 years, most research devoted to ethanol production from cellulose has focused on fungal systems (primarily *Trichoderma reesei*) for breakdown of cellulose into sugars, coupled with the sugar fermentation processes of yeast (*Saccharomyces cerevisiae*) (Demain et al. 2005). A deeper understanding of a greater variety of cellulolytic and ethanologenic systems is critically needed. Bacterial species in diverse physiological groups (e.g., those having various tolerance levels for oxygen, temperature, and salt concentrations) are known to hydrolyze cellulose; thus, a wide range of natural habitats need to be explored for novel cellulolytic activities in bacteria.

How do soluble enzymes act on an insoluble crystalline substrate? Hydrolysis of crystalline cellulose is the rate-limiting step in biomass conversion to ethanol because aqueous solutions of enzymes have difficulty acting on this insoluble, highly ordered structure. Cellulose molecules in their crystalline form are packed so tightly that enzymes and even small molecules such as water are unable to permeate the structure. How do different biomass-degrading enzymes work together as a synergistic system? Cellulases and hemicellulases are secreted from cells as free enzymes or as large, extracellular complexes known as cellulosomes. The collective activity of these enzyme systems is much more efficient than the individual activity of any isolated enzyme; therefore, to truly understand how enzymes function, they must be studied as systems rather than individually or a few at a time. In addition, these systems eventually must be analyzed under laboratory conditions more representative of real-world environments.

Why are ethanologenic organisms less efficient at using certain sugar substrates? A varied mix of hexoses (e.g., glucose, mannose), pentoses (e.g., xylose, arabinose), and oligosaccharides are released from the hydrolysis of lignocellulosic materials, and no microorganism is capable of efficiently fermenting all these sugars. The most widely studied ethanologenic microbes (e.g., yeast) prefer to use glucose as a substrate. Even when yeast cells are modified genetically to use xylose, they ferment all available glucose before switching to the much slower fermentation of xylose. Conversion rates can vary greatly, depending on such factors as the type of sugar substrate being fermented, environmental conditions (e.g., pH, temperature), and concentrations of certain products from other metabolic pathways.

How effective are sugar transporters at translocating different sugars across the cell membrane? Sugar transporters are membrane-bound proteins that take up sugars from the environment and deliver them to the metabolic pathways inside cells. The inefficient transport of different sugar substrates by microbes can result in low product yield and is a major obstacle to the efficient conversion of biomass to ethanol. Our limited understanding of sugar transporters (as well as many other energy-relevant, membrane-associated proteins) is due to a lack of adequate techniques for producing membrane proteins and studying their structure and function.

Why do different enzymatic and microbial processes operate optimally at different temperatures? Cellulases operate optimally at temperatures higher (>40°C) than those tolerated by current ethanologenic organisms, so these two processes cannot be consolidated into a single process step. Thermophily (tolerance of high temperatures) improves the robustness of enzymes or microbes needed for industrial-scale processes and reduces the likelihood of culture contamination. The basis by which enzymes, pathways, and entire microbes are made thermophilic is poorly understood, and methods for inserting cellulolytic or fermentative pathways into thermophilic organisms are not well developed.

What are the requirements for producing and maintaining stable mixed cultures? We currently do not understand in sufficient detail the dynamics of microbial consortia that carry out stable mixed processes such as aerobic and anaerobic digestion. At a minimum, cultures used in bioethanol-production systems will need to be resistant or stable despite contamination by outside microbes or other potentially toxic materials or life forms. Some microbial consortia clearly are capable of biological processes not possible for a single constituent organism, but the principles of association that allow these consortia to form and function are not understood. Without this understanding, we will not be able to design or engineer such systems.

How can we improve systems for genetically engineering microorganisms or plants involved in bioethanol production? Although many studies have expressed genes from cellulolytic organisms in *Escherichia coli* or other mesophilic organisms, systems for expressing foreign genes in cellulolytic or thermophilic organisms need further development. Our limited understanding of microbial regulation prevents successful engineering of a microbe capable of versatile expression of lignocellulolytic enzymes, utilization of multiple sugars, glycolysis, and fermentation. High-throughput capacity for transforming diverse bioenergy crops is needed to enable forward and reverse genetic approaches for introducing new genetic capabilities.

Biohydrogen Production

Hydrogen is a promising energy carrier of the future: It can be derived from a variety of energy sources and used in fuel cells with high efficiency; combustion of hydrogen produces only water as a by-product, making it a nonpolluting, carbon-free energy alternative. The most common industrial methods for producing hydrogen include steam reformation of natural gas, coal gasification, and water splitting with electricity typically generated from fossil fuels. However, these industrial processes often require more energy than the resulting hydrogen yields and release carbon dioxide, other GHGs, and pollutants as by-products. Some microorganisms produce hydrogen naturally, and biotechnologies based on these microbial systems could lead to clean, renewable sources of hydrogen. In a recent report on the hydrogen economy, however, NRC noted that "substantial, fundamental research needs to be undertaken before photobiological methods for large-scale hydrogen production are considered" (Hydrogen Economy 2004).

Several reviews have examined the potential of photobiological hydrogen production (Madamwar, Garg, and Shah 2000; Ghirardi et al. 2000; Melis and Happe 2001; Tamagnini et al. 2002; Levin, Pitt, and Love 2004; Nath and Das 2004; Prince and Kheshgi 2005). Although microorganisms produce hydrogen by different mechanisms, the step can be represented by the simple chemical reaction $2H+ + 2e- <—> H_2$. This reaction is known to be catalyzed by either nitrogenase or hydrogenase enzymes. Although alternative

biological hydrogen production-pathways exist, each with its own set of advantages and disadvantages, the following discussion will focus on the challenges that must be overcome to improve one type of biological hydrogen production known as biophotolysis.

Under certain conditions, green algae and cyanobacteria can use water-splitting photosynthetic processes to generate molecular hydrogen (H_2) rather than fix carbon, the normal function of oxygenic photosynthesis. Bidirectional hydrogenases in these organisms use electrons from the photosynthetic electron-transport chain to reduce protons to yield H_2. Biophotolysis holds potential for the scale of hydrogen production necessary to meet future energy demand. This approach is promising because water is the source of electrons or reducing power required to generate hydrogen. Water is a clean, renewable, carbon-free substrate available in virtually inexhaustible quantities. Another advantage of biophotolysis is the more efficient conversion of solar energy to hydrogen. Reengineering microbial systems for direct production of hydrogen from water eliminates inefficiencies associated with carbon fixation and biomass formation. Theoretically, the maximal energetic efficiency for direct biophotolysis is about 40% (Prince and Kheshgi 2005), compared with a maximum of about 1% for hydrogen production from biomass (Hydrogen Economy 2004).

Hydrogen, although promising enormous benefits in the future, entails additional challenges beyond the science. There are economic challenges (getting more value out than the investment in making it) and larger societal issues that include developing the necessary infrastructure, transportation, and storage. Most experts do not view hydrogen as a near-term contributor of significance to U.S. energy needs. Continued investment is worthy and prudent, however, because a technological breakthrough could alter hydrogen's status in the energy economy, and, as noted above, hydrogen is the ultimate nonpolluting and potentially inexhaustible source of energy. Some scientific GTL challenges include engineering oxygen-tolerant, efficient hydrogenases and designing microbes that can use them. Hydrogenases known to tolerate oxygen generally are not very efficient hydrogen producers, so better hydrogenases with sufficient activity and oxygen tolerance will need to be engineered. The engineered hydrogenases could be used in bioinspired nanostructures that maintain optimal conditions for hydrogen production. In addition, if successfully engineered, photosynthetic microbes genetically modified to take advantage of these modified hydrogenases could produce hydrogen at higher rates and efficiency from the biophotolysis of water. They also might be grown in extensive farms of sealed enclosures (photobioreactors). Hydrogen would be harvested for use in energy applications, with oxygen released as a by-product.

Biohydrogen Production: Gaps in Scientific Understanding

Understanding biophotolysis well enough to model hydrogenase structure and function, regulatory and metabolic networks, and eventually entire organisms will stimulate the kind of biotechnological innovations needed to engineer the ideal organism for use in hydrogen bioreactors or the ideal enzyme-catalyst for use in bioinspired nanostructures for hydrogen production. Achieving this level of understanding, however, will require basic research that investigates a greater range of hydrogen-producing enzymes and organisms, mechanisms of hydrogenase assembly, oxygen sensitivity of hydrogenase, electron-transfer rate limitations, and regulatory and metabolic processes that influence hydrogen production. Some specific issues relevant to these basic research needs follow.

What is the extent of natural diversity among hydrogenases and hydrogen-producing organisms? Research conducted by the J. Craig Venter Institute has established the existence of vast numbers of bacteriorhodopsin-like genes previously unknown to science in marine environments (Venter 2004). A number of organisms containing hydrogenases have been identified and sequenced, and, while they cannot be cultured in the laboratory using current procedures, they establish the existence of a wealth of potential targets for future explorations. Studying hydrogenase enzymes involved in nonbiophotolytic pathways could provide structural or functional insights to guide the engineering of biophotolytic systems.

How are hydrogenases assembled, and how are metals incorporated into the active site? Two major types of hydrogenases are defined by their biologically unique metallocenters: Nickel-iron (NiFe) and iron only (Fe). NiFe hydrogenases are found in many bacteria and some cyanobacteria. Fe hydrogenases are found in some bacteria and green algae. In green algae, hydrogenases are bidirectional (capable of catalyzing hydrogen oxidation or proton reduction to produce H_2); in cyanobacteria, hydrogenases are either bidirectional or are uptake enzymes. Although turnover is much higher for Fe hydrogenases, NiFe hydrogenases are more oxygen tolerant. The metallocenters of both NiFe and Fe hydrogenases form complexes with such unusual inorganic cofactors as carbon monoxide or cyanide. Little is known about the assembly or maintenance of an active hydrogenase, and several genes may be involved in the synthesis of cofactors required for activity. A better understanding of hydrogenase assembly will enable the engineering of enzymes with improved function.

How do we overcome hydrogenase oxygen sensitivity? The bidirectional Fe hydrogenases that catalyze the hydrogen-evolution reaction in biophotolytic systems are highly sensitive to oxygen, a product of the water-splitting reaction in the photosynthetic pathway's first step. Oxygen sensitivity also makes hydrogenase isolation from cells and its subsequent analysis a challenge that will be met by new technologies.

What potential electron-transfer rate limitations are associated with each step of the biophotolytic hydrogen-production pathway? Key factors that can impact the partitioning of electrons between hydrogenase and competing pathways include the buildup of a pH gradient across the photosynthetic membrane and variations in concentrations of critical electron-transport carriers. Understanding how electron fluxes in an organism are regulated will aid the development of mechanisms for directing more electrons toward proton reduction and hydrogen production.

What regulatory and metabolic pathways influence H_2 production? A thorough examination of hydrogen metabolism in green algae and several different strains of cyanobacteria from diverse habitats will provide new insights into the control of hydrogen-production pathways. By understanding how an organism sustains and regulates hydrogen production, we will be able to determine which metabolic pathways contribute, how eliminating hydrogen-consuming reactions affects hydrogen metabolism and other cellular processes, and how organisms can be adapted to increase hydrogen yields.

Thus, the aim of the GTL program and its planned Bioenergy Research Centers is to answer these questions related to cellulose-to-ethanol and biohydrogen and questions underlying other bioenergy-related challenges. The full force of biotechnology and genomics will be applied to ascertain and utilize skills that biological organisms have possessed for billions of years.

Technological Capabilities for Systems Biology Research on Microbes and Plants

A broad spectrum of analytical tools in the GTL Centers will be needed to rapidly advance our current understanding of fundamental scientific issues impeding biofuel production. Potential economic and environmental advantages of developing bioenergy options cannot be realized, however, without pursuing answers to key fundamental scientific questions underlying critical R&D breakthroughs. The following section lists scientific and technological capabilities needed above and beyond today's reality. Since most of them will be equally applicable and relevant to both the cellulose-to-ethanol and biohydrogen missions, these needs are combined below, with specific applications to one or the other mission noted.

Cellulose-to-Ethanol and Biohydrogen: Needed Scientific and Technological Capabilities

Improving current understanding of bioethanol production will require a variety of new technologies and technological capabilities, including approaches for surveying enzyme diversity; visualizing enzyme systems; efficiently producing functional enzyme systems and membrane proteins; cultivating microbial

consortia; integrating transcriptomics, proteomics, and metabolomics; and genetically engineering microorganisms for integrated bioprocessing. Some of these technologies exist today but are in low-throughput form. Much more efficient and high-throughput technologies will always be a high priority. Again, more detail on these technologies was laid out in the GTL Roadmap, but some specific needs are noted below. Key capabilities to address many gaps in current understanding of biophotolytic hydrogen production include developing microbial hosts to produce hydrogenase enzymes, screening of large numbers of enzymes for desired functionalities, large-scale molecular profiling to provide a global-view of hydrogen production, *in vivo* visualization of hydrogenase structure and activity, modeling of regulatory and metabolic networks, and metabolic engineering. Given the shared science and technology needed for these two missions, the Centers' value clearly would extend beyond a particular bioenergy strategy. Needed capabilities include the following.

Large-scale production of cellulose synthases, cellulase enzyme systems, sugar transporters, hydrogenases, natural and engineered variants of all of these critical enzymes, and other proteins. This is a key need because of the difficulty in working with these systems without sufficient amounts of purified and characterized proteins for experiments. It is especially important with proteins about which we know little or nothing (around 40% of the total) that will require improved methods for protein production and characterization. High-throughput techniques and expression systems for efficiently producing membrane proteins, sets of different enzymes that work together, and enzyme complexes such as cellulosomes are in need of development. With so much variability among natural hydrogenases and engineered variants, high-throughput capabilities for producing large numbers of enzymes (perhaps hundreds of thousands to millions) to screen for O_2 tolerance, H_2-production activity, spectroscopic examination, and structural analysis could accelerate the discovery of enzymes best suited for biotechnological applications.

Suites of microbial hosts to produce cellulases, fermentation enzymes, and hydrogenases from many different organisms. Potentially thousands of enzymes from many different organisms will need to be produced and analyzed. Other requirements include methods for producing eukaryotic enzymes in simpler prokaryotic systems, designing host organisms that can provide the intracellular environment for proper protein assembly and folding, and screening proteins from these host organisms.

Methods to grow stable mixed cultures. Improved experimental and modeling tools are needed for producing a mixed microbial culture or even a plant-microbial culture. The goal is to enable stable performance for each population carrying out one part of overall ethanol (or hydrogen) production or nutrient uptake.

Molecular profiling to provide a global view of cellular activity. Improvements in computational capabilities and large-scale molecular profiling techniques (transcriptomics, proteomics, metabolomics, and interactomics) are needed to obtain a global view of ethanol or microbial hydrogen production and plant responses to programmed developmental or environmental signals. Systems-level analyses could guide experimental investigations by defining gene regulatory networks controlling the expression of genes involved in ethanol or hydrogen production or cofactor synthesis. Such analyses could identify pathways activated or deactivated during biofuel production in multiple organisms under varying conditions, especially for complex communities comprising soil ecosystems.

Methods to integrate transcriptomic, proteomic, and metabolomic information. Techniques that integrate data gathered from these global molecular measurements are essential in determining which genes are expressed and functionally active during cellulose synthesis and utilization, ethanol fermentation, or hydrogen production, and which metabolites influence the activity of enzymes involved in these pathways. As an insoluble substrate, cellulose cannot enter cells and induce the expression of genes involved in cellulose hydrolysis. Metabolic profiling could be used to identify which substrates or metabolites activate or repress expression of key cellulolytic genes and at what quantities. In addition to illuminating regulatory strategies

for cellulases and other coexpressed enzymes such as ligninases, these integrated "omic" approaches could be used to build regulatory and metabolic maps to guide genetic engineering. For example, these maps could be used to identify the best potential gene knockouts that redirect carbon flux from a particular sugar substrate toward ethanol fermentation and bypass competing pathways that produce other organic end products. Similarly, such maps could enable the best approaches to optimizing hydrogen production or modifying lignin composition.

Methods to genetically engineer organisms currently refractory to transformations. To take only one example, *Clostridium thermocellum* is an anaerobic bacterium capable of both hydrolyzing cellulose and fermenting sugars to ethanol, but its yields are poor and conversion is slow. Improved methods for genetically modifying this and other cellulolytic microbes are needed. To achieve the ambitious goal of developing an organism capable of integrated bioprocessing, we must focus on better ways of genetically transforming "nontraditional" microbes (i.e., other than *E. coli*) to understand how microbial systems function and how their interacting pathways influence each other. *Rhodopseudomonas palustris* is capable of multiple relevant processes, including hydrogen generation (Larimer 2004), and improved transformation methods could help to optimize this attribute. Complex polyploid eukaryotes such as perennial grasses or trees present similar challenges for the rapid and stable introduction of multigene cassettes.

Methods to perform *in vivo* visualization and characterization of molecular machines. Advanced imaging techniques will provide new insights into how growing cellulose microfibrils interact with the cytoskeleton, and how cellulases interact with crystalline cellulose and overcome current barriers to efficient cellulose hydrolysis (e.g., substrate accessibility, product or substrate inhibition, low product yield). Structural information and imaging from X-ray, nuclear magnetic resonance spectroscopy, scanning transmission electron microscopy, and other techniques will be needed to identify additional interactions among cellulases and other molecules needed for efficient function. Although crystal structures of some hydrogenases have been determined, this information provides only snapshots of enzyme structure. Advanced techniques for visualizing different stages of hydrogenase assembly or dynamic and spatial monitoring of hydrogenase activity in living cells in response to extracellular conditions will be critical to building predictive models for engineering hydrogenases optimized for biotechnological applications. Advanced imaging techniques to define and monitor plant cell-wall synthesis also will be needed for optimal redesign of physical properties or enzyme accessibility.

Metabolic engineering. Metabolic engineering involves genetically modifying organisms to target and manipulate enzymatic, regulatory, or transport pathways that impact particular biological processes such as carbon assimilation and ethanol or hydrogen production. Models could guide metabolic engineering, for example, by identifying control points for manipulating the flow of electrons to hydrogenase or by predicting how cellular activity and hydrogen yields may be impacted by a variety of conditions. These conditions include the elimination of a particular metabolic pathway or the buildup of a pH gradient across the photosynthetic membrane. Metabolic engineering also might be used to redirect the flow of carbon-containing compounds to plant cell-wall synthesis or to alter the ratio of specific components within the plant cell wall to facilitate subsequent breakdown into biofuels.

Core Technological Capabilities Needed for GTL Bioenergy Research Centers

Research at the Centers will focus on the study of critical microbial and plant properties and processes on three systems levels—molecular, cellular, and community. Accordingly, the GTL Centers will incorporate, coordinate, or develop core capabilities to carry out their work, tailoring them to specific research needs. These may include capabilities in genomics, genetics, physiology, biochemistry, structural and computational biology, bioinformatics, synthetic genomics, nanoscience, and engineering. Many technologies relevant to production and characterization of macromolecules; analytical capabilities; proteomics and

metabolomics; imaging at the molecular, cellular, and systems levels; microbial culturing; and data management and computation are described in detail in the GTL Roadmap. Following is a summary of several technologies that might be essential enabling components of Center research. For economy and efficiency, some technologies sited at other locations could be connected virtually.

Sequencing. Genome sequencing will take place at the DOE Joint Genome Institute under separate support from the DOE BER and will be independent of the Bioenergy Research Centers. DOE JGI will be a front-end resource for GTL and its Centers.

Genomics. Building on genomic sequencing of relevant microbes and plants, subsequent technologies must include (but are not limited to) expression systems and microarrays to quantitatively assess actual use of genetic information; gene synthesis to generate, either in native or modified form, gene sequences that can be transferred into appropriate hosts to test gene-product function in the milieu of cell environment; gene cloning for production of protein and probes to interrogate gene and mRNA expression under varying circumstances; and modification or mutagenesis to test the effects of deliberate alterations in the function of genomic DNA *in situ*.

Proteomics and Metabolomics. The result of a gene sequence usually is a protein, and proteins form the structures in a cell and carry out its functions. Technologies needed for synthesis, isolation, analysis, characterization, and manipulation of proteins will include high-throughput generation of proteins, complexes, cell-wall structures, and metabolites; methods for production and biophysical characterization; expression systems for soluble and membrane-associated proteins; generation of recalcitrant proteins and complexes; improvements in scaling protein production and efficiency; high-throughput technologies for microfluidics, affinity assays, and chips to assess protein binding and interactions; and functional characterizations or high-throughput screens for protein activities. Analytical capabilities for measuring the presence and absolute quantity of the full range of molecules (proteins, RNAs, metabolites, and molecular machines) in microbial and plant systems will be needed. Such capabilities will require small sample volumes, measurement reproducibility and sensitivity, low cost, and high-throughput capacity compared to existing technologies. As noted earlier, Bioenergy Research Centers will focus on some of the more difficult and less-soluble proteins encoded by the genome because they comprise the subset most relevant to bioenergy generation. Also, proteins are post-translationally modified by phosphorylation and glycosylation, the addition of small groups that can dramatically affect many such properties as binding, interactions, and activities. Assessing post-translational modifications and their effects is challenging under the best of circumstances, but proteins the Centers will need to explore (e.g., membrane-associated, complexed with other proteins into molecular machines, and sometimes present either transiently or at best in small concentrations) will be among the most difficult to characterize. Potentially, such miniaturization technologies as microfluidics and microelectromechanical systems (called MEMS) with potential to provide new functionalities in detection, manipulation, and analysis of biological systems can be developed for both metabolomics and proteomics.

Microbial culturing. The vast preponderance of the microbial world has been largely unstudied because it comprises members that do not grow readily (or grow very slowly, if at all) in commonly used culture conditions. Many bacteria are fastidious, with preferred pH, temperatures, oxygen levels (varying from none to very high), and levels of various metals, ions, and nutrients. The potential range of culture conditions is large, and better technologies for determining optima for multiple microbes, both as individuals and as consortia, are badly needed.

Imaging. Technologies are necessary to image intracellular constituents and components, including individual proteins, protein complexes, metabolite distributions, and structural features (e.g., microfilaments). Activity rates must be measured in real time to build models eventually that can represent protein, pathway, or whole-cell functions and have sufficient predictive value to inform the models. Technologies to spatially

and temporally map multiple processes through the life cycles of microbial or plant systems need to be developed and validated. Technologies to "count" numbers of copies of proteins and protein complexes (as a function of cell cycle, time, and environmental perturbations) are needed to ground predictive models.

Biomass processing and plant cultivation. As Genomics:GTL science improves our understanding of the nature of plant biomass and how it can be converted to simpler sugars and then fermented to liquid and other fuels, technologies will be needed to improve the "front end" of this conceptual pipeline from plant matter to biofuel and build a systems understanding of it. If a chain is as strong as its weakest link, a pathway is as fast as its slowest step. If the slow step is at the front end (i.e., breakdown of plant material into a form that microbial or fungal enzymes can process), the faster and more controlled processing of biomass will be necessary. Plants are well known to grow at a very wide range of rates, from bamboo that can grow feet a day to trees that can take a decade or more to reach maturity. Technologies for plant cultivation are needed, both to generate the starting material for biomass processing and for test systems to carry out the Centers' science approaches.

Data-management, computational, and bioinformatic capabilities. The Bioenergy Research Centers, as well as the linked GTL science program, already are generating vast amounts of data (and this will only increase) that need to be preserved and used. The GTL Knowledgebase must include the full spectrum of capabilities, from a data repository incorporating standardized data structures and formats to accessible and user-friendly analytical tool sets, visualization tools, and communication protocols for ready access from any platform. All will be vital to building structural models of macromolecules and complexes, performing molecular dynamic simulations of their intermolecular interactions to elucidate testable predictions of their functions, and linking them in pathways to move biomass from solid form to fuel. Increased computational power will be required as the nature of biological computation becomes more complex. As the Bioenergy Research Centers continue to generate large volumes of data from their high-throughput methodologies and technologies, the challenges of data storage, access, and effective and efficient use will increase. Needed technologies include enhancements in genome annotation especially related to recognition of interacting proteins, regulatory signals and structures relevant to bioenergy activities in microbes and plants; infrastructure for large-scale genome analysis and databases for microbial genomes; new computational models and mathematical approaches for predictive modeling, simulating, and visualizing complex types of cellular processes and interactions and complex data sets; new computing hardware and networking infrastructure for analysis, modeling, and simulation from large data sets, including terascale communications; distributed approaches to capacity computing problems; and environments for petascale, numerically intensive, physics-based simulations.

Future GTL Research Centers for DOE's Environmental Missions

In addition to developing biofuels for a more secure energy source, DOE has two other mission challenges related to the GTL program: Developing biological solutions for intractable environmental problems and assessing options for carbon sequestration by understanding the relationships between climate change and the earth's microbial systems. For more detail, see the GTL Roadmap (U.S. DOE 2005). As with the Bioenergy Research Centers, future Centers will maintain the spirit of basic research while pursuing specific DOE mission goals. These Centers will be vertically integrated in research and technological capabilities, focusing on microbial properties and processes on three systems levels—molecular, cellular, and community.

Environmental Remediation

DOE has the challenge of remediating a large volume of legacy wastes consisting of a "witches' brew" of heavy metals, radionuclides, and various toxic organics that have been reacting together, sometimes for as long as 60 years. DOE is committed to cleaning up the large volumes of soil, sediment, and groundwater

contamination at diverse facilities and sites across the nation. Projected costs of restoring these sites and disposing of wastes is $142 billion (Closure Planning Guidance 2004).

Microbes found in the contaminated subsurface and other environments often have the metabolic capability to degrade or otherwise transform these contaminants to potentially less-harmful forms. Although cost and effectiveness comparisons of metal and radionuclide bioremediation to traditional methods are not available, cost savings for bioremediation of organics are estimated to range from 30 to 95%. In addition, *in situ* bioremediation that takes advantage of natural microbial populations in the subsurface has the potential to reduce costs and increase the efficiency of groundwater treatment, as compared to conventional pump-and-treat technology.

The GTL mission science goal regarding environmental remediation is to understand the processes by which microbes function in the earth's subsurface, microbial mechanisms that can impact contaminant fate and transport, scientific principles of bioremediation based on native microbial populations and their interactions with the environment, and methods to relate genome-based understanding of molecular processes to long-term conceptual and predictive models. Attaining this goal will require understanding the biogeochemical processes, from the fundamental-molecular to community levels, and describing contaminant transformation processes coinciding with simulated changes in microbial-community composition and structure.

Carbon Cycling and Sequestration

Atmospheric GHG concentrations have been increasing for about two centuries, mostly as a result of human activities, and are now higher than they have been for more than 400,000 years. Although the effects of increased CO_2 levels on global climate are uncertain, many agree that doubling atmospheric CO_2 concentrations, predicted for the middle of this century by the Intergovernmental Panel on Climate Change, could have a variety of serious environmental consequences (IPCC 2001). Global climate change is a long-term energy and environmental challenge requiring major investments in targeted research and development. Gaining a greater knowledge of carbon cycling through ecosystems is a critical element of the national strategy to understand climate change and changes that might occur due to anthropogenic GHGs and to solutions to reduce future increases in CO_2.

Oceans cover 71% of the earth's surface, and marine and terrestrial ecosystems play a major role in the global cycling of carbon. Microbes are essential to maintaining the planet's ability to sustain life, including recycling most of the earth's biomass and both assimilating and respiring large amounts of carbon dioxide. Small changes in these natural fluxes, induced by climate change or natural processes, could overwhelm any mitigation attempts we might make within global energy systems. The DOE mission of global carbon management requires a comprehensive understanding of terrestrial and marine microbial communities so we can learn the role played by these communities in carbon sequestration and cycling. The GTL program seeks to provide a systems-level understanding of microbial processes essential to carbon cycling. Knowledge revealed by GTL research will be incorporated into global climate models to provide a robust science base for evaluating potential impacts of proposed carbon-management strategies.

Cited References

AEI. 2006. *Advanced Energy Initiative*, The White House, National Economic Council (www.whitehouse.gov/stateoftheunion/2006/energy/energy_booklet.pdf).

BTAC. 2002. *Vision for Bioenergy and Biobased Products in the United States*, Biomass Technical Advisory Committee (www.climatevision.gov/sectors/electricpower/pdfs/bioenergy_vision.pdf).

Check, B. 2002. "AAM Sees Bright Prospects for Microbial Ecology Research in the Genomic Era," *ASM News* **68**(9), 427–31.

Closure Planning Guidance. 2004. Office of Environmental Management, U.S. Department of Energy (http://web.em.doe.gov/program.html).

Demain, A. L., M. Newcomb, and J. H. D. Wu. 2005. "Cellulase, *Clostridia*, and Ethanol," *Microbiol. Mol. Biol. Rev.* **69**, 124–54.

EIA. 2006. *Annual Energy Outlook 2006 with Projections to 2030*,

DOE/EIA-0383, Energy Information Administration, U.S. Department of Energy (www.eia.doe.gov/oiaf/aeo).

EIA. 2005. *International Energy Outlook 2005*, DOE/EIA-0484, Energy Information Administration, U.S. Department of Energy (www.eia.doe.gov/oiaf/ieo/).

Ghirardi, M. L., et al. 2000. "Microalgae: A Green Source of Renewable H_2," *Trends Biotechnol.* **18**, 506–11.

Greene, N., et al. 2004. *Growing Energy: How Biofuels Can Help End America's Oil Dependence*, Natural Resources Defense Council, New York (www.nrdc.org/air/energy/biofuels/biofuels.pdf).

Handelsman, J., et al. 1998. "Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products," *Chem. Biol.* **5**, R245–49.

The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs. 2004. National Research Council and National Academy of Engineering, National Academies Press, Washington, D.C.

IPCC. 2001. *Climate Change 2001: The Scientific Basis*, ed. J. T. Houghton et al., Cambridge University Press, Cambridge, U.K. (www.grida.no/climate/ipcc_tar/wg1/index.htm).

Larimer, F. W. et al. 2004. "Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium *Rhodopseudomonas palustris*," *Nat. Biotechnol.* **22**, 55–61.

Levin, D. B., L. Pitt, and M. Love. 2004. "Biohydrogen Production: Prospects and Limitations to Practical Application," *Int. J. Hydrogen Energy* **29**, 173–85.

Madamwar, D., N. Garg, and V. Shah. 2000. "Cyanobacterial Hydrogen Production," *World J. Microbiol. Biotechnol.* **16**, 757–67.

Melis, A., and T. Happe. 2001. "Hydrogen Production: Green Algae as a Source of Energy," *Plant Physiol.* **127**, 740–48.

Nakicenovic, N., et al. 2000. *Special Report on Emissions Scenarios*, ed. N. Nakicenovic and R. Swart, Cambridge University Press, New York (www.grida.no/climate/ipcc/emission/index.htm).

Nath, K., and D. Das. 2004. "Biohydrogen Production as a Potential Energy Resource: Present State-of-Art," *J. Sci.Ind. Res.* **63**, 729–38.

NEPDG. 2001. *National Energy Policy: Reliable, Affordable, and Environmentally Sound Energy for America's Future*, National Energy Policy Development Group (www.whitehouse.gov/energy/National-Energy-Policy.pdf).

NRC. 2006. *Review of the Department of Energy's Genomics:GTL Program*, Committee on Review of the Department of Energy's Genomics:GTL Program, National Research Council, National Academies Press, Washington, D.C. (genomicsgtl.energy.gov/pubs/NASreport22206.pdf)

O'Toole, G. A. 2003. "To Build a Biofilm," *J. Bacteriol.* **185**(9), 2687–89.

Perlack, R. D., et al. 2005. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*, DOE/GO-102005-2135, Oak Ridge National Laboratory, Oak Ridge, Tenn. (http://feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf).

Prince, R. C., and H. S. Kheshgi. 2005. "The Photobiological Production of Hydrogen: Potential Efficiency and Effectiveness as a Renewable Fuel," *Crit. Rev. Microbiol.* **31**, 19–31.

RFA. 2006. *From Niche to Nation: Ethanol Industry Outlook 2006*, Renewable Fuels Association (www.ethanolrfa.org/resource/outlook).

Saha, B. C. 2004. "Lignocellulose Biodegradation and Applications in Biotechnology," pp. 2–34 in *Lignocellulose Biodegradation*, American Chemical Society, Washington, D.C.

Tamagnini, P., et al. 2002. "Hydrogenase and Hydrogen Metabolism of Cyanobacteria," *Microbiol. Mol. Biol. Rev.* **66**, 1–20.

U.S. DOE. 2006. *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda*, DOE/SC-0095, U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy (genomicsgtl.energy.gov/biofuels/).

U.S. DOE. 2005. *Genomics:GTL Roadmap: Systems Biology for Energy and Environment*, U.S. Department of Energy Office of Science (genomicsgtl.energy.gov/roadmap/).

U.S. DOE. 2003. *The Department of Energy Strategic Plan*, DOE/ME-0030, Office of Program Analysis and Evaluation, Office of Management, Budget, and Evaluation, U.S. Department of Energy (http://strategicplan.doe.gov).

Venter, J. C., et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* **304**, 66–74.