DOE/SC-0135

Grand Challenges for Biological and Environmental Research: A Long-Term Vision

A Report from the Biological and Environmental Research Advisory Committee

Suggested citation for this report: BERAC. 2010. *Grand Challenges for Biological and Environmental Research: A Long-Term Vision; A Report from the Biological and Environmental Research Advisory Committee March 2010 Workshop*, DOE/SC-0135, BERAC Steering Committee on Grand Research Challenges for Biological and Environmental Research (www.science.doe.gov/ober/berac/BER_LTVreport.pdf).

Cover

Images on the cover represent a broad range of natural systems that drive the science supported by the Office of Biological and Environmental Research within the U.S. Department of Energy Office of Science. These systems are not only structurally and spatially complex with many different interacting parts spanning molecular to global scales, but they also are dynamically complex, encompassing processes that occur over time scales ranging from nanoseconds to centuries.

Image credits: Ball-and-stick representation of atoms within a molecule from Pacific Northwest National Laboratory. Ribosome molecular complex image from Ditlev Brodersen, Aarhus University. Microscopic image of fungal hyphae on a root surface, copyright Merton Brown, Visuals Unlimited. Forest ecosystem image from the U.S. National Park Service. Mountain and Earth images from iStockphoto. Cover developed at Oak Ridge National Laboratory.

DOE/SC-0135

Grand Challenges for Biological and Environmental Research: A Long-Term Vision

A Report from the
Biological and Environmental Research Advisory Committee
March 2010 Workshop

Chair
Gary Stacey (University of Missouri)

U.S. Department of Energy

December 2010

Prepared by the BERAC Steering Committee on
Grand Research Challenges for Biological and Environmental Research

Biological and Environmental Research Advisory Committee

Chair

Gary Stacey
University of Missouri

Janet Braam
Rice University

Robert E. Dickinson
University of Texas at Austin

James R. Ehleringer
University of Utah

Joanna S. Fowler
Brookhaven National Laboratory

Paul Gilna
Oak Ridge National Laboratory

Susan Hubbard
Lawrence Berkeley National Laboratory

Andrzej Joachimiak
Argonne National Laboratory

Ruby Leung
Pacific Northwest National Laboratory

Gerald (Jay) Mace
University of Utah

Stephen R. Padgette
Monsanto Company

Joyce E. Penner
University of Michigan

Gregory Petsko
Brandeis University

David A. Randall
Colorado State University

Karin Remington
National Institute of General Medical Sciences

Margaret A. Riley
University of Massachusetts

G. Philip Robertson
Michigan State University

Gary Sayler
University of Tennessee

Herman Shugart
University of Virginia

James Tiedje
Michigan State University

Judy D. Wall
University of Missouri

Warren M. Washington
National Center for Atmospheric Research

Raymond E. Wildung
Pacific Northwest National Laboratory

Mavrik Zavarin
Lawrence Livermore National Laboratory

Minghua Zhang
State University of New York at Stony Brook

Designated Federal Officer

David Thomassen
U.S. Department of Energy Office of Biological
and Environmental Research

Workshop Steering Committee

Gary Stacey
Chair, University of Missouri

James R. Ehleringer
University of Utah

Joyce E. Penner
University of Michigan

Gregory Petsko
Brandeis University

David A. Randall
Colorado State University

Margaret A. Riley
University of Massachusetts

Gary Sayler
University of Tennessee

Mavrik Zavarin
Lawrence Livermore National Laboratory

Contributing Authors

Chapter 2: Grand Challenges in Biological Systems

Adam Arkin, Nitin Baliga, Janet Braam, George Church, Jim Collins, Robert Cottingham, Joseph Ecker, Mark Gerstein, Paul Gilna, Jean Greenberg, Peter Greenberg, Jo Handelsman, Susan Hubbard, Andrzej Joachimiak, James Liao, Loren Looger, Elliot Meyerowitz, Eric Mjolness, Greg Petsko, Gary Sayler, Michael Simpson, Gary Stacey, Michael Sussman, and James Tiedje

Chapter 3: Grand Challenges in Climate Research

David Bader, Paola Cessi, William Collins, Scott Denning, Robert Dickinson, David Easterling, Jae Edmonds, Johannes Feddema, Christopher Field, Ann Fridlind, Inez Fung, Isaac Held, Robert Jackson, Anthony Janetos, William Large, Margaret Leinen, Ruby Leung, Stephen Long, Gerald Mace, Carrie Masiello, Gerald Meehl, Donald Ort, Bette Otto-Bliesner, Joyce Penner, Michael Prather, David Randall, Philip Rasch, Edwin Schneider, Herman Shugart, Peter Thornton, Warren Washington, Raymond Wildung, Warren Wiscombe, Donald Zak, and Minghua Zhang

Chapter 4: Grand Challenges in Energy Sustainability

Jeffrey Bielicki, Marilyn Buford, Elsa Cleland, Virginia Dale, Cliff Duke, James Ehleringer, Alan Hecht, Daniel Kammen, Ruby Leung, Gregg Marland, Diane Pataki, Margaret Riley, and Philip Robertson

Chapter 5: Grand Challenges in Computing for Biological and Environmental Research

Robert Cottingham, Paul Gilna, Susan Hubbard, Joyce Penner, and David Randall

Chapter 6: Grand Challenges in Education and Workforce Training

Jo Handelsman and Margaret Riley

Table of Contents

Executive Summary

The interactions and feedbacks among plants, animals, microbes, humans, and the environment ultimately form the world in which we live. This world is now facing challenges from a growing and increasingly affluent human population whose numbers and lifestyles are driving ever greater energy demand and impacting climate. These and other contributing factors will make energy and climate sustainability extremely difficult to achieve over the 20-year time horizon that is the focus of this report. Despite these severe challenges, there is optimism that deeper understanding of our environment will enable us to mitigate detrimental effects, while also harnessing biological and climate systems to ensure a sustainable energy future.

This effort is advanced by scientific inquiries in the fields of atmospheric chemistry and physics, biology, ecology, and subsurface science—all made possible by computing. The Office of Biological and Environmental Research (BER) within the Department of Energy's (DOE) Office of Science has a long history of bringing together researchers from different disciplines to address critical national needs in determining the biological and environmental impacts of energy production and use, characterizing the interplay of climate and energy, and collaborating with other agencies and DOE programs to improve the world's most powerful climate models. BER science focuses on three distinct areas:

- What are the roles of Earth system components (atmosphere, land, oceans, sea ice, and the biosphere) in determining climate?
- How is the information stored in a genome translated into microbial, plant, and ecosystem processes that influence biofuel production, climate feedbacks, and the natural cycling of carbon?
- What are the biological, geochemical, and physical forces that govern the behavior of Earth's subsurface environment?

Ultimately, the goal of BER science is to support experimentation and modeling that can reliably predict the outcomes and behaviors of complex biological and environmental systems, leading to robust solutions for DOE missions and strategic goals.

In March 2010, the Biological and Environmental Research Advisory Committee held the *Grand Challenges for Biological and Environmental Research: A Long-Term Vision* workshop to identify scientific opportunities and grand challenges for BER science in the coming decades and to develop an overall strategy for drafting a long-term vision for BER. Key workshop goals included:

- Identifying the greatest scientific challenges in biology, climate, and the environment that DOE will face over a 20-year time horizon.
- Describing how BER should be positioned to address those challenges.
- Determining the new and innovative tools needed to advance BER science.
- Suggesting how the workforce of the future should be trained in integrative system science.

This report lays out grand research challenges for BER—in biological systems, climate, energy sustainability, computing, and education and workforce training—that can put society on a path to achieve the scientific evidence and predictive understanding needed to inform decision making and planning to address future energy needs, climate change, water availability, and land use.

Cross-Cutting Science Themes

Common science themes arose across the workshop's breakout sessions. These themes include complex systems science across scales, multidisciplinary research, computing and mathematics, and human impacts.

Complex Systems Science Across Scales

Spanning BER science is the need to understand complex biological and environmental systems over many spatial and temporal scales, from molecular to global and nanoseconds to centuries. Although capabilities for measuring physical, chemical, and biological parameters have greatly advanced, science still struggles to make meaning from these data, to integrate the parts into an understandable whole, and to move efficiently to predictions of biological and environmental outcomes. The biological and climate sciences are moving from piecemeal studies of individual components to a point where a whole system–level of understanding is possible and necessary to address energy and climate challenges.

Increasingly clear is that the components of our environment are interconnected. Just as animals, plants, and communities of organisms are made up of many highly interactive individual cells, the Earth is composed of numerous interacting processes—from biological systems to atmospheric chemistries to oceans to global water cycles—that are governed by complex feedforward and feedback pathways. Our ability to model and predict the impacts of environmental and Earth system change is critical to our eventual understanding of this complexity and the development of potential solutions. For example: How do changes at the molecular level influence the ability of a cell or multicellular organism to respond to and impact its environment? How does the environment respond to the altered behavior of biological systems? How do these environmental changes interact with climate to mitigate or accelerate climate change? These questions become even more daunting when we try to extrapolate from the ancient record or to predict impacts into the future over large time scales. Yet these efforts are necessary for understanding the potential severity and impacts of climate change.

Multidisciplinary Research

Addressing the issues of complexity and scalability necessitates the intense interaction and integration of data and knowledge emanating from a broad array of scientific disciplines. A tighter coupling of theory, observations, experiments, models, and simulations developed across disciplinary interfaces is key to the deeper systems understanding required. As more sophisticated models are developed, we will be able to test their consistency with laboratory experiments and forecast that which we are unable to do experimentally. New hypotheses will be drawn from model results, enabling the design of new rounds of experiments that refine concepts to enable the predictive capabilities required for solutions to DOE mission challenges.

Computing and Mathematics

Computing and mathematics underlie the science required to address the grand challenges identified in this report (summarized on the following pages). As common models of scientific understanding are developed from the growing body of data, these approaches and new knowledge will be captured by computational technologies. However, the explosion in data, emanating from diverse methods, presents severe challenges for storage, integration, analysis, and, most important, understanding. New paradigms clearly will be required. Similarly, as the science progresses, there is an equally pressing need for the education and training of scientists to keep pace.

Human Impacts

The roles that humans play are important determinants of and vital contributors to global sustainability. By 2100, the world's population is projected to increase 40% over today's level, and this growing population will simultaneously be seeking a higher standard of living. The rapid pace of anthropogenic change requires that human impacts be factored into assessments if we are to have an accurate understanding of the Earth system 20 years from now. Likewise, ecosystem changes lead to changes in patterns of human behavior and activity. Therefore, any comprehensive Earth system model (ESM) must incorporate human activities, which by necessity include an analysis of socioeconomic factors.

Summary of Grand Challenge Research Recommendations

The vision and research challenges outlined as follows establish an initial framework of ideas that are expected to be further refined at follow-up workshops engaging the scientific community.

Grand Challenges in Biological Systems

As modern experimental and computational tools rapidly transform biology from an observational to an informational, data-intensive science, one manifestation is the expansion of the complementary fields of systems and synthetic biology. While systems biology provides the approaches needed to address biological complexity and achieve a systems-level understanding, synthetic biology tests understanding to see if a system functions as expected. Then such validated understanding potentially could be used to design and construct novel biological systems for purposes relevant to DOE missions.

Systems biology is a comprehensive, quantitative analysis of the manner in which components of a biological system interact functionally over time and space. By exploring how all components work together as a functional system, systems biology enables the discovery of the organizing principles and emergent properties that can be seen only via a systemic view of the entire biological process. An added complexity is understanding microbial and plant systems within the context of dynamic, real-world environments.

High-throughput technologies used in systems biology research are generating immense datasets previously unheard of in biology. These advances demand improvements in the management of large bioinformatic knowledgebases and the development of a new generation of computational tools that will support spatial and temporal modeling.

Synthetic biology is essentially a tool chest of methods and approaches for understanding natural systems using simplified or controllable analogs such as "minimal cell architectures," genetic circuits, metabolic networks, and constructed metabolic pathways. Coupled with computational modeling analogous to electronic circuit design, synthetic biology can enable the design and construction of new bioinspired materials, organisms, pathways, or information-processing systems. Key issues involve characterizing what nature already has assembled and determining how to access, harness, and improve biological components or processes to meet future challenges or understand system responses to change.

Significant advances in analytical methodology, particularly mass spectrometry, now provide capacity for near high-throughput environmental analysis of biological materials, including proteins and small information molecules, giving rise to the rapidly expanding field of infochemicals. New biologically inspired research in this area can take advantage of leadership-class DOE computational and analytical characterization capabilities.

Integrating systems and synthetic biology approaches with conceptual and numerical models is helping to build connections from genome- and molecular-level research to investigations at physiological and ecosystem levels. Research needs range from systems-level understanding at the cellular level to challenges associated with investigating systems of increasing biological complexity.

Workshop participants identified the following grand challenge research recommendations for biological systems:

- Enabling predictive biology.
  - Develop a simulation model of a single cell for accurately predicting phenotype from genotype-environment interactions.
  - Establish new model organisms for relevant ecological process understanding.
  - Use robust biochemical, functional, and experimental evidence to enhance genome and metagenome annotation.
  - Develop the biological understanding for generalizing and applying models from simple systems to more complex systems, working toward an ultimate goal of modeling the evolution and dynamics of a complex biological system under environmental stress conditions.
  - Identify general design principles used in natural systems, articulate these, and then use them to distill our understanding of biology.
- Measuring and analyzing biological systems.
  - Apply advanced computational and analytical capabilities to characterize the information molecules and network interactions used by biological systems.
  - Improve capabilities for imaging a single cell at a resolution of one molecule per cell.
  - Measure microbial processes and interactions in the real world and in experimental simulations.
  - Provide standardizable, reproducible sampling protocols and observations enabling functional characterizations of systems and synthetic design.
  - Define the range of analytes, ligands, and fluxes of materials to be measured.
  - Define scalability requirements for measurement technologies and develop capabilities for *in situ* and *in vitro* sensors.
- Exploring ecosystem function and elemental cycling.
  - Understand, predict, and manipulate the types and rates of ecosystem responses and feedbacks that result from and influence climate change.
  - Deploy synthetic (or nonsynthetic) biology to understand and manipulate ecosystem function.
  - Determine the molecular basis of robustness, fitness, and selection.
  - Develop a complete understanding of the biogeochemical cycles important to regulating carbon flux through biological systems.
  - Apply functional metagenomics to enable mass balance closure for biogeochemical cycles and transfer this information to biosystem design.
  - Manage plant and microbe stress response to control carbon biosequestration and remediation of metals and radionuclides.
  - Define the fundamental microbial basis for permafrost carbon-methane transformations.
  - Develop designs for optimizing carbon flow for biomass production, carbon allocation, and biosequestration to reduce rates of atmospheric CO_2 accumulation and to increase terrestrial carbon storage by 50% in 20 years.
  - Determine carbon and nutrient dynamics in natural systems.

Grand Challenges in Climate Research

Twenty years from now, observation and modeling systems will produce much more detailed analyses and predictions of both weather and climate. Included will be additional regional-scale diagnostics such as temperature, precipitation, and extreme weather event statistics. ESMs will also become more effective at finer scales and for shorter time periods.

These dramatic increases in model resolution—enabled in part by exponential increases in computing power—will result in approximately 1-km global grid spacing for the atmosphere and 1- to 5-km resolution for oceans. Such resolution is comparable to the scale at which many observational measurements are made. Forthcoming observational technologies and tools will describe Earth system processes in greater detail, enabling more robust models of their effects on climate, as well as the degree to which change in the climate system affects these processes. These processes include ice sheet and glacier dynamics; clouds, aerosols, and precipitation; global biogeochemical cycling of carbon, water, and other natural compounds; soils and terrestrial vegetation, including human decision making about crop choices and management; groundwater resources; biological interactions and feedbacks; anthropogenic greenhouse gas emissions; and interactions among all of these processes.

Comprehensive comparisons of ESM predictions with observations will be possible, as well as quantification of model errors. However, breakthrough developments are needed to determine how to translate information about disparate, coupled processes and phenomena across scales—from the molecular and pore scale to the local, regional, or global scale of the prediction. ESMs are also complex in terms of their conceptual basis, embodiment as computer programs, and sheer numbers of variables and parameters. New approaches are needed to deal with this complexity, especially for analyzing and understanding model results.

Decadal projections will be an important outcome of these emerging models. These forecasts will capture El Niños and decadal modes of climate variability and their influences on extreme weather. They also will quantify regional time-evolving climate change to inform decision makers balancing options about mitigation of and adaptation to climate change with other societal needs. Emerging ESMs will be used for predicting not only climate change, but also the consequences of particular societal choices on ecosystems. Needed developments include tools that predict energy consumption and the consequences of energy use, as well as tools that support energy infrastructure planning.

Workshop participants identified the following grand challenge research recommendations for climate science:

- Develop higher-resolution models to integrate many more relevant processes than current models and to describe climate change over much longer time scales.
- Improve parameterizations and basic knowledge about aerosols that affect clouds and factors that control their number and concentration for microphysics, radiative transfer, and turbulence processes to quantify indirect aerosol forcing and resulting precipitation changes.
- Develop ecosystem-observing systems to monitor biogeochemical cycles, estimate critical process parameters, and provide model tests in ocean and terrestrial biospheres, including subsurface soils.
- Advance understanding of important biological interactions and feedbacks to identify potential tipping points and possible mitigation strategies such as carbon biosequestration.
- Improve integration of anthropogenic climate forcings into ESMs and develop new techniques to evaluate these coupled models at both global and regional scales.
- Establish new observational technologies and use them to comprehensively compare ESM predictions with observations and to quantify model errors.

Grand Challenges in Energy Sustainability

Sustainability and its sibling, resilience, have emerged in this century as the new discipline of sustainability science and technology. It can be thought of as neither basic nor applied research, but as a field defined by the problems it addresses rather than by the disciplines it employs. The emerging research field of sustainability science deals with the broad challenges and interactions between natural and social science to ensure that natural resources, ecosystem services, and economic opportunities are globally available to future generations. Over the coming decades, drivers such as climate and land-use change; food, water, and energy security; and life-cycle assessment in industrial production and agriculture will intersect internationally. There is an overwhelming need for a research agenda that is responsive to these interconnections in order to define, quantify, and ameliorate (where possible) changes in the underlying fabric supporting sustainability.

Endeavors to achieve energy sustainability suggest the need to understand the consequences and processes necessary for mitigating the problems and externalities resulting from how the energy and industrial sectors are organized. Research must address problems caused by present activities (e.g., CO_2 emissions), develop future alternatives, and anticipate potential problems and consequences of these alternatives so that they can be minimized.

Research also needs to address not only technical and scientific issues and constraints, but also the degree to which solutions are implemented and the subsequent consequences. Varied ethical, legal, and societal implications are inherent in the science and technology of energy systems, and the degree to which solutions are adopted affects social systems.

Sustainability, by its very nature, calls for multidisciplinary research that spans spatial and temporal scales. Compartmentalized research, development, and associated agendas can inform understanding of system components, but sustainability is an area of study that must address the overlap and interrelationships of systems.

Workshop participants identified the following grand challenge research recommendations for energy sustainability:

- Analyze and compare potential approaches to organize land use, water use, and energy systems in ways that achieve sustainable energy, food, biodiversity, and ecosystem functioning.
    - Develop the understanding needed to produce enough energy to support more people at a higher standard of living in ways that sustain growth and minimize negative environmental impacts.
    - Determine the means to double, over 20 years, the share of energy needs met by bioenergy in environmentally and economically sustainable ways.
    - Develop affordable and competitive options for energy supply and conversion that minimize negative direct, indirect, and life-cycle impacts on climatic, environmental, and ecological systems.
- Identify and characterize potential Earth system drivers, feedbacks, and vulnerabilities to state changes so that their consequences and triggers might be avoided.
    - Characterize the spatial and temporal variabilities of biological systems throughout Earth and be able to understand, predict, and manipulate their rates of change.
    - Develop technologies to remotely monitor ecosystem services in real time and at landscape scales.
- Develop unifying models and frameworks capable of testing and evaluating the significance of potential global change issues, including energy, land use, and water. The impacts these issues have on both society and the environment also must be tested and assessed.
    - Develop the scientific basis for life-cycle analysis and full-cost accounting of ecosystem services for extant and next-generation energy technologies.

- Advance the science to detect, understand, and mitigate the environmental and security issues associated with the potential resurgence of nuclear power.
- Advance the discipline of ecological informatics and incorporate social and policy elements into science-based sustainability research.
- Develop the science perspective necessary to set target resolutions for the spatial and temporal dimensions underlying systems of interest and be able to convert global change estimations into model visualizations.
- Determine the degree to which future sustainable energy facilities should be centralized.

Grand Challenges in Computing for Biological and Environmental Research

As experimental systems become increasingly digital, more accurate, and efficient, the resultant datasets continue to grow exponentially and encompass an expanding diversity of data types. These data represent not only fundamental measurements but all of the surrounding conditions and factors that influence biological function and relate to a more accurate and detailed description of the surrounding environment. This increase in scientific data has been faster than anticipated over recent decades because of improvements in electronics, computing, and experimental systems. Growth over the next 20 years is expected to be significantly greater, given the goals of improved models and the expected further advancements in experimental techniques.

A theme shared across BER science is that as the research community develops common "models" of scientific understanding, the approaches and new knowledge are captured in computational systems. Substantial increases in data—along with possible limits in the growth of computer processing speed—will demand new approaches such as parallel computational methods. Similarly, experimental protocols and the reported data should accurately describe current scientific understanding and experimental results and provide the supporting ontological and semantic information to enable machine-readable support for subsequent searching, analysis, and reuse. Ultimately, these models will be captured entirely in computational systems that then will provide an automated basis for data storage, integration, query, and retrieval. These systems will enable computational simulations and projections that allow further assessment and understanding, resulting in new directions for future experiments and model improvements.

Working toward these new approaches is becoming critical because of the increasing scale and complexity of data. For these data to be useful, computing capabilities are required that can consistently and coherently manage and integrate the large sets of data and information.

Computational modeling at multiple scales is required for understanding biological systems within a bacterium or in the context of environmental and ecological interactions that impact the carbon cycle and climate. Models could interact at each scale, but today there is no continuity between models of microbial systems and those that predict global climate. As models become refined and increasingly precise at each scale, learning how a model can interact meaningfully with those at adjacent scales will present opportunities to gain further insight.

Workshop participants identified the following grand challenge research recommendations for computing:

- Establish a new data management paradigm for data-intensive science with ontologies as a basis for semantic data representations; standards for experimental protocols and data exchange; and an open-access, open-development data management infrastructure.
- Create a new publishing paradigm that credits and rewards researchers for publishing peer-reviewed

datasets or analytical methods in addition to conventional peer-reviewed journal articles.

- Develop new computing paradigms capable of meeting the enormous parallel processing and data-intensive analysis needs now emerging for biological, climate, and environmental data.
- Standardize experimental and computational protocols and methods to increase data integration, data usability, and system interoperability to improve research productivity.
- Improve data usability and model accuracy by ensuring that appropriate data quality standards are created and stored with the accompanying data.
- Design and build software solutions that provide researchers with better access to increasingly large, complex, and interrelated datasets.
- Develop virtual laboratories and tools to more fully engage human cognitive faculties and provide richer opportunities for scientific collaborations.

Grand Challenges in Education and Workforce Training

BER-supported research clearly is shifting from a focus on the parts to a quest for the whole, developing approaches for studying processes over a hierarchy of scales ranging from the subnanometer to kilometers. There is an imperative need to educate our future workforce to think about properties of whole systems, which rarely can be explained simply as an accumulation of parts. This new research framework necessitates reorienting education from teaching the known to exploring the unknown. Education research dictates that the teaching of science should capture the nature of inquiry, presenting science as methods of scientific investigation rather than as extant bodies of knowledge. Inquiry-driven teaching impels instructors to use problem-based learning, open-ended laboratory exercises, and collaborative discussion—all techniques shown to foster higher-order thinking skills as well as retention of information. BER must develop a clear, unifying vision for education and workforce development, emphasizing the integration of scientific disciplines and tackling complexity through the use of multidisciplinary teams.

Workshop participants identified the following grand challenges for education and workforce training:

- Engage science educators.
- Develop a centralized education mission.
- Initiate interdisciplinary fellowships.
- Enhance career development programs.
- Support engineering education to address crucial shortages in engineering disciplines relevant to BER's grand research challenges.
- Install education experts at the national laboratories.
- Develop collaborative teaching programs.

1 Introduction

Strengthening the economy and national security by reducing our dependence on oil, developing sustainable energy alternatives, restoring contaminated environments, and understanding how carbon dioxide (CO_2) and other energy-related emissions impact Earth's climate and carbon cycle—all of these headline issues represent some of the greatest challenges facing the United States today and for decades to come. The science that can provide innovative solutions to these challenges and inform policy and decision making is what drives the research programs supported by the Office of Biological and Environmental Research (BER) within the Department of Energy (DOE) Office of Science (see sidebar, Drivers for BER Science, p. 2). BER's mission is to provide research opportunities and scientific user facilities that advance our understanding of complex biological and environmental systems important to DOE missions so that, ultimately, we can build predictive models of these systems.

History of BER Science

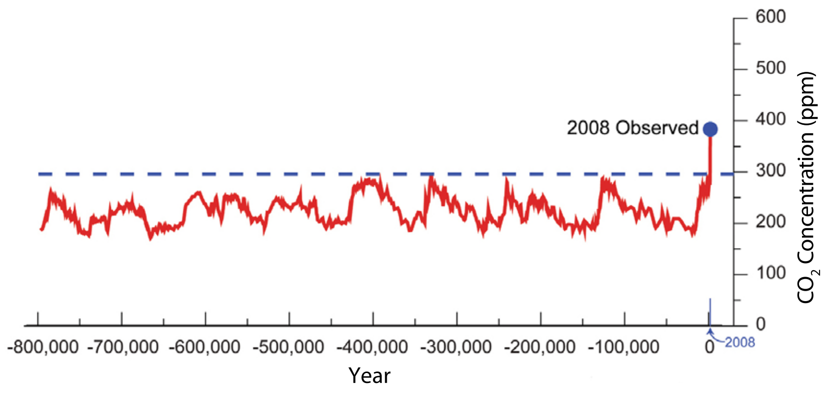
BER has a long history of bringing together researchers from different disciplines to address critical needs in determining the biological and environmental impacts of energy production and use, characterizing the interplay of climate and energy, and collaborating with other agencies and DOE programs to improve the world's most powerful climate models.

With origins in atomic energy research during the middle of the 20th century, BER's predecessor agencies were charged with examining the biological effects of exposure to radiation and energy by-products. Recognizing the need for a complete reference sequence of human DNA, BER initiated its Human Genome Program in 1986, which led to the international Human Genome Project and launched a new era of genome-enabled biology. Early DOE research to understand and predict the global distribution of radioactive particles from nuclear weapon tests led to development of the first general circulation models, forerunners to the global climate models currently supported by BER and other agencies. Furthermore, DOE supported the first ecological studies to examine the fate and effects of radioactive substances in the environment. Insights and tools resulting from this research have translated to current BER studies of contaminant transport and carbon cycling in terrestrial ecosystems.

Today, BER continues to advance research at the leading edge of science areas essential to achieving a sustainable future for energy, climate, and the environment. By revealing the physical, chemical, and biological drivers of climate change, BER is working to resolve some of the greatest uncertainties in projecting future climate. Using the power of genomics and systems biology, BER is exploring how microbes and plants can provide new options for energy production and carbon biosequestration. By collectively studying and modeling the interconnected microbiological, hydrological, and geochemical processes in the subsurface, BER is building a foundation of knowledge for assessing and improving proposed approaches to environmental remediation. In essence, BER programs and activities currently are driven by three science questions:

- What are the roles of Earth system components (atmosphere, land, oceans, sea ice, and the biosphere) in determining climate?
- How is the information stored in a genome translated into microbial, plant, and ecosystem processes that influence biofuel production, climate feedbacks, and the natural cycling of carbon?

Drivers for BER Science: Energy and Climate



Historical Trends in CO_2 Concentration. Analysis of air bubbles in an Antarctic ice core has shown that natural processes have kept the atmospheric CO_2 concentration within a range of about 170 to 300 parts per million (ppm) over at least the last 800,000 years. As a result of human activities, the present CO_2 concentration of about 385 ppm is about 30% above its highest level in hundreds of thousands years. [Figure adapted from: U.S. Global Change Research Program (www.globalchange.gov). Figure based on data from: Lüthi et al. 2008 and Tans 2008.]

One of the greatest risks to U.S. energy security, economic growth, and environmental health is our dependence on oil. Each year the United States exports billions of dollars to import about 60% of the petroleum it consumes, primarily for transportation. Greater strain on the world's finite oil supplies is expected as developing countries become more industrialized and energy consumption increases, driving prices higher in a global energy market threatened by oil-induced economic turbulence. By investing in scientific innovation that can sustainably expand and diversify our domestic energy supply, we can begin to take control of the nation's energy future. This investment presents numerous opportunities for biological systems science supported by the Department of Energy Office of Biological and Environmental Research (BER) to help develop new biology-based energy alternatives and to understand and predict the effects that different energy options will have on biological and environmental services and resources.

Energy choices are inextricably linked to future climate. Global climate has been relatively stable for thousands of years; however, the warming trend of the last few decades continues to intensify, and the most recent decade is the warmest in the instrumental record. The increase in the global average surface temperature is faster and steadier than can be explained by natural variability alone, and multiple lines of strong scientific evidence show that the rapid accumulation of anthropogenic CO_2, primarily from fossil energy use, is one of the most significant factors driving this temperature change and influencing future climate. Human activities have pushed atmospheric CO_2 concentration to a level greater than any the Earth has experienced in hundreds of thousands of years (see figure). Unknown are the rate and magnitude of biological, environmental, and climatological responses to this unprecedented increase.

Understanding how the biosphere and Earth's biogeochemical cycles will change in a warmer, higher-CO_2 world is a key scientific challenge that must be addressed to improve climate forecasting and inform decisions about climate mitigation activities. Crossing certain temperature thresholds in biological and environmental systems could unleash massive amounts of stored carbon or initiate other abrupt, unexpected changes at local, regional, and global scales. For example, permafrost currently occupies 16% of global soil area, sequestering an estimated 1672 Pg (1.67×10^{12} metric tons) of carbon—more than double the amount in the atmosphere. Relative to the 10 Pg of carbon emitted annually from human activities, permafrost thawing and associated microbe-mediated conversion of permafrost carbon to greenhouse gases could add 100 to 750 Pg to the atmosphere by 2100, yet these processes are poorly understood and not represented in climate models. BER is improving the integration of research on biological processes that play critical roles in carbon cycling, research on the physical and geochemical processes influencing climate, and development of climate and Earth system modeling. This integrated approach uniquely positions BER to provide the foundational science that can improve climate projections and inform sustainable strategies for producing energy, sequestering carbon, and identifying and responding to climate risks.

- What are the biological, geochemical, and physical forces that govern the behavior of Earth's subsurface environment?

Advancing Complex Systems Science

Although the questions driving BER science cover a diverse set of mission priorities, a common theme across all of these research areas involves applying and improving approaches to understand and predict the behavior of complex systems. Defining a complex system depends on the context, but, in general, such a system consists of many heterogeneous, interdependent components that collectively interact, self-organize, and evolve to exhibit larger-scale, emergent properties or behaviors that no individual component can display in isolation. Whether studying the movement of radionuclides in a groundwater plume, the interface between a microbial cell and a plant cell wall, the response of a soil microbial community to elevated temperatures, or global climate simulations of a future scenario that doubles current atmospheric CO_2—BER science must deal with complexity.

To influence or manipulate a complex system, the understanding of it needs to be sufficient to enable scientists to anticipate how the system will respond under certain conditions and where and for how long this response will have an effect. This is why modeling and simulation play such significant roles in BER science—they underlie the ultimate goal of prediction. Biological and environmental systems, however, are especially challenging to model. These systems are not only structurally and spatially complex with many different interacting parts that span molecular to global scales, but they also are dynamically complex, encompassing processes that occur over time scales ranging from nanoseconds to centuries.

A reductionist research approach that completely deconstructs a system and characterizes its constituent parts, though important, cannot provide the level of understanding needed for biological and environmental systems. A holistic approach, which aims to define the organizing principles that transform a collection of parts into a functional system, also is needed. Thus the quest to understand a complex system requires a combination of reductionist and holistic strategies. Computing also is an essential component of this complex systems science approach. For the diverse systems investigated by BER scientists, a variety of models and software tools is needed to integrate the large streams of data and information emanating from these studies. The rapid evolution of computing power and analytical technologies over the last few decades has enabled scientists to take on systems and problems of increasing complexity. This trend is expected to continue in the coming decades.

Defining a Long-Term Scientific Vision for BER

Given the extent to which biological, environmental, and climate research has advanced in the last two decades, it is exciting to imagine where BER science could be in another 20 years. Although the research planning process often tends to focus on near- to mid-term time frames of 3 to 5 years, defining longer-term directions for science also is important. In March 2010, the Biological and Environmental Research Advisory Committee held a workshop to identify scientific opportunities and grand challenges in the coming decades and to develop an overall strategy for drafting a long-term vision for BER. Key workshop goals included:

- Identifying the greatest scientific challenges in biology, climate, and the environment that DOE will face over a 20-year time horizon.
- Describing how BER should be positioned to address those challenges.
- Determining the new and innovative tools needed to advance BER science.
- Suggesting how the workforce of the future should be trained in integrative system science.

The following chapters describe the grand research challenge recommendations that workshop participants identified for biological systems, climate research, energy sustainability, computing for biological and environmental research, and education and workforce training. The vision and challenges in this report establish an initial framework of ideas expected to be further refined at follow-up workshops engaging the scientific community.

2 Grand Challenges in Biological Systems

> *"It appears to me that beyond this stratum of molecular biology, or above it, as some . . . would say, is a second stratum, a stratum which contains problems of strategy, of programming, of how to use the various and ingenious molecular devices invented by creatures to make a creature or society. To this class of problems I give the name systems biology. . . This appears to be a problem in the programming—programming of the use of the information contained in the genetic material."*
>
> **— James Bonner. 1960. California Institute of Technology Archives.**

Complexity is a major challenge for most scientific disciplines in the 21st century, especially biology (see sidebar, Why Are Biological Systems So Complex?, p. 7). The functions of any complex biological system—including the microbial and plant systems important to the Department of Energy's (DOE) Office of Biological and Environmental Research (BER)—arise from the extraordinary arrangements and dynamic interactions among numerous components that form higher-order structures at each level of biological organization. Transformational advances in the capacity and speed of genome sequencing, high-throughput experimental technologies, and computational resources over the last 20 years now enable global analyses of molecular species at the whole-cell level. Computational and mathematical tools now permit scientists to acquire, store, transmit, integrate, mine, and finally model data to begin to convert the growing wealth of data into accessible knowledge or information. These powerful new tools and techniques are helping biologists examine how numerous molecular-scale phenomena collectively lead to more complex, larger-scale behaviors.

As modern experimental and computational tools rapidly transform biology from an observational to an informational, data-intensive science, one manifestation of this evolution is the expansion of the complementary fields of systems and synthetic biology. While systems biology provides the approaches needed to address biological complexity effectively and achieve a systems-level understanding, synthetic biology tests this understanding by using it to design and construct novel biological systems for useful purposes. Building on a long tradition of successful biological research, BER is playing a leadership role in applying the tools of systems and synthetic biology to critical 21st century problems in energy security, carbon and greenhouse gas (GHG) management, adaptation to climate change, and development of bioinspired materials and processes central to environmental sustainability (see box, Summary of Research Recommendations, p. 6, and sidebar, Biological Research to Advance DOE Missions, p. 8).

Systems Biology

Systems biology is a comprehensive, quantitative analysis of the manner in which components of a biological system interact functionally over time and space (Aderem 2005). Although not a new concept, systems biology has piqued a recent explosion of interest driven by the increasing availability of full genome sequences, which has enabled the expanding development and application of new tools for a systems-level analysis of cellular function. The ultimate goal is a new, predictive view of biological function, augmenting the older, descriptive understanding.

Summary of Research Recommendations

Enabling Predictive Biology

2.1 Develop a simulation model of a single cell for accurately predicting phenotype from genotype-environment interactions.

2.2 Establish new model organisms for relevant ecological process understanding.

2.3 Use robust biochemical, functional, and experimental evidence to enhance genome and metagenome annotation.

2.4 Develop the biological understanding for generalizing and applying models from simple systems to more complex systems, working toward an ultimate goal of modeling the evolution and dynamics of a complex biological system under environmental stress conditions.

2.5 Identify general design principles used in natural systems, articulate these, and then use them to distill our understanding of biology.

Measuring and Analyzing Biological Systems

2.6 Apply advanced computational and analytical capabilities to characterize the information molecules and network interactions used by biological systems.

2.7 Improve capabilities for imaging a single cell at a resolution of one molecule per cell.

2.8 Measure microbial processes and interactions in the real world and in experimental simulations.

2.9 Provide standardizable, reproducible sampling protocols and observations enabling functional characterizations of systems and synthetic design.

2.10 Define the range of analytes, ligands, and fluxes of materials to be measured.

2.11 Define scalability requirements for measurement technologies and develop capabilities for *in situ* and *in vitro* sensors.

Exploring Ecosystem Function and Elemental Cycling

2.12 Understand, predict, and manipulate the types and rates of ecosystem responses and feedbacks that result from and influence climate change.

2.13 Deploy synthetic (or nonsynthetic) biology to understand and manipulate ecosystem function.

2.14 Determine the molecular basis of robustness, fitness, and selection.

2.15 Develop a complete understanding of the biogeochemical cycles important to regulating carbon flux through biological systems.

2.16 Apply functional metagenomics to enable mass balance closure for biogeochemical cycles and transfer this information to biosystem design.

2.17 Manage plant and microbe stress response to control carbon biosequestration and remediation of metals and radionuclides.

2.18 Define the fundamental microbial basis for permafrost carbon-methane transformations.

2.19 Develop designs for optimizing carbon flow for biomass production, carbon allocation, and biosequestration to reduce rates of atmospheric CO_2 accumulation and to increase terrestrial carbon storage by 50% in 20 years.

2.20 Determine carbon and nutrient dynamics in natural systems.

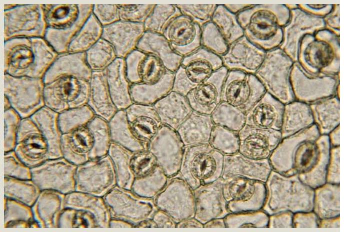
Similar to the scientific revolution of molecular biology, systems biology represents a fundamental change in the way researchers ask questions. In addition to providing a global view of all biological system components, systems biology is about programming and the strategies needed to bring a collection of biological parts to life as a dynamic, functional system. This holistic approach to biology is as much about information flow and feedbacks as it is about the flow of biological substances. In order to understand the vast quantities of molecular components and information shared in just a single cell, large bioinformatic databases and the latest computational tools and advances in spatial and temporal modeling are essential to the success of systems biology.

Why Are Biological Systems So Complex?

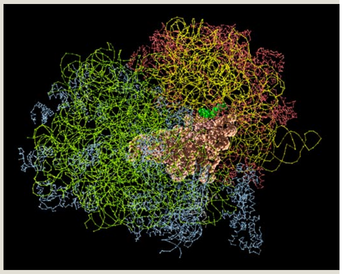
- **Hierarchical and multiscalar levels of biological organization and activity.** Biological systems—from a single bacterial cell to a forest ecosystem—are the dynamic products of many hierarchical levels of components interacting with each other and their environment across many scales of space and time. However, the connections between these different levels of interactions are poorly understood.
- **Vast diversity of components within biological systems.** At each level of organization, a biological system contains numerous parts—thousands of proteins, nucleic acids, and metabolites at the molecular level within a cell, or millions to billions of different cells at the microbial community or tissue level. All of these parts have diverse functions, and they interact in complex ways. These functions and interactions underpin the phenotypes observed for a particular set of environmental conditions.
- **Environmental dependency of biological phenotype and response.** An organism's phenotype or the behavior of a biological system is highly dependent on environmental conditions. Biological function and response depend not only on a system's inherent properties (e.g., its genome), but also on how it interacts with other biotic and abiotic components of its environment. Environmental changes (e.g., fluctuations in nutrient or water availability, temperature change, or the introduction of invasive species) can have significant impacts on these interactions, so understanding and predicting environmental changes are essential parts of predicting biological response.
- **Emergent properties.** Even if all the parts of a biological system are thoroughly characterized, unanticipated functions and features can emerge at all levels of biological organization. The collective response of myriad microscale events, or even rare or isolated occurrences, can lead to seemingly unpredictable larger-scale outcomes. This nonlinear and adaptive nature of self-organizing biological systems is a hallmark of biological complexity and makes extrapolations difficult. Workshop participants suggest that a new mathematics is needed to model biological systems to accurately capture issues of dynamic growth, interactions, and change.

Organisms and Ecosystems

Tissues and Organs



Cells



Molecules

Activities and interactions at the molecular level underpin the functions and interactions that occur at higher levels of biological organization, such as cells, tissues, organs, organisms, and ecosystems. [Image credits: iStockphoto.]

Synthetic Biology

Synthetic biology is (1) "the design and construction of new biological parts, devices, and systems; (2) the redesign of existing, natural biological systems for useful purposes" (syntheticbiology.org); and (3) the inspiration for new materials and processes derived from the design, functions, and products of living cells. Often considered a component of systems biology, synthetic biology represents a tool chest of methods and approaches for understanding natural systems using simplified or controllable analogs such as "minimal cell architectures," genetic circuits, metabolic networks, and constructed metabolic pathways. By integrating with computational science, synthetic biology can enable

Biological Research to Advance DOE Missions

For over six decades, the Department of Energy's (DOE) Office of Biological and Environmental Research (BER) and its predecessor programs have supported landmark research applying and developing the latest technological advances to study biological problems important to DOE's energy, environment, climate, and basic science missions. Since initiating its Human Genome Program in 1986, BER has spearheaded the development of modern genomics-based systems biology and played a major role in seeding and fostering the contemporary biotechnology revolution.

Although several federal agencies support and benefit from the continued advances in systems biology, BER—with its rich history in interdisciplinary, problem-focused research in biology, climate and carbon cycle science, subsurface biogeochemistry, and computational science—is particularly suited to play a leadership role in understanding biological systems at the energy-climate interface. The following examples of BER science have driven the emergence of an integrated multidisciplinary research agenda that fully embraces the tools of molecular biology, genomics, and computational analysis.

- **Plant and microbial genomics.** BER's Genomic Science program provides the science and technology development needed for a systems-level understanding of DOE-relevant organisms. This program and BER support for genome sequencing at the DOE Joint Genome Institute have been major platforms for significant progress in plant and microbial genome sciences.
- **DOE Bioenergy Research Centers.** The three DOE Bioenergy Research Centers are applying cutting-edge technologies to understand the processes essential for the production of cellulosic biofuels. These centers provide collaborative research environments well suited for applying systems and synthetic biology approaches to advance DOE BER missions.
- **Subsurface biogeochemistry to inform remediation of legacy waste.** Long-term BER support for subsurface biogeochemical research has contributed to a new understanding of the microbial systems and the interrelated hydrological, geophysical, chemical, and biological processes controlling contaminant fate and transport in the subsurface, an area of unique and urgent importance to DOE.
- **Structural biology leveraging DOE scientific user facilities.** BER supports experimental stations at several DOE Office of Science facilities. These include the synchrotron light sources located at Argonne National Laboratory, Brookhaven National Laboratory, Lawrence Berkeley National Laboratory, and SLAC National Accelerator Laboratory, as well as neutron beam sources located at Los Alamos National Laboratory and Oak Ridge National Laboratory. These facilities provide imaging and analytical capabilities that drive research in fundamental materials science and structural biology critically important in computational biology and dynamic modeling.
- **Molecular Systems Science.** BER's Environmental Molecular Sciences Laboratory is advancing understanding of physical, chemical, and biological processes at the molecular and systems levels. This research is leading to new strategies for bioenergy production and carbon biosequestration, improved catalysts and materials for industrial applications, tools for managing and predicting the movement of subsurface legacy wastes such as radionuclides and heavy metals, and approaches for mitigating climate change.

the design and construction of essentially new materials, organisms, pathways, or information-processing systems.

The analytical tools of synthetic biology already are being used—with traditional workhorse organisms of microbiology and biotechnology—to metabolically engineer new biofuel pathways (Brynildsen and Liao 2009; Atsumi et al. 2009). Key issues for synthetic biology involve characterizing what nature already has assembled and determining how to access, harness, and improve biological components or processes to meet future challenges or understand system responses to change. In a broad systems context, a grand challenge is advancing the use of synthetic biology's tool chest within natural ecosystems to investigate complex dynamics and adaptations, molecular (even quantum) interspecies interactions (e.g., microbe-microbe and plant-microbe), and emergent properties.

Importance of *In Situ* Analyses for DOE Biological Systems

Many of the biological challenges and opportunities important to DOE missions involve the additive complexity of understanding microbial and plant systems within the context of a particular real-world environment at scales ranging from molecular to ecosystem levels. For example, we can obtain the entire genome sequence of a bioenergy crop such as switchgrass, but predicting the photosynthetic rate or some other measurable phenotype requires understanding how this plant functions within a specific environment. The metabolism for a microbe studied in the laboratory may look nothing like the metabolism of the same microbe living in the soil—intimate associations taking place in the microscale environment can have a profound influence on microbial metabolism. The overall vision is to understand, predict, and perhaps manipulate the relationship among microbial and plant systems relevant to DOE missions, the environment in which such organisms exist, and the effects that changes to one have on the other.

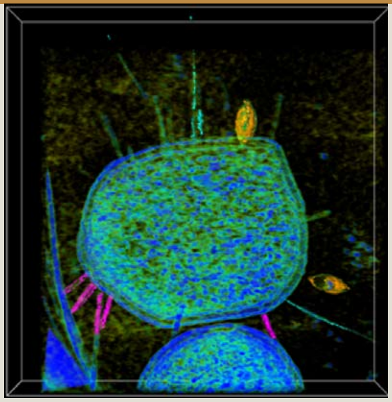
With metagenomics and metaproteomics, the systems biology paradigm is moving out of the laboratory and into the environment where biological systems can be studied *in situ* (see sidebar, Metagenomics of Natural Microbial Communities, p. 10). Previously laborious and slow methodologies have yielded to significant advances in analytical tools, microarrays, RT-PCR (reverse transcription polymerase chain reaction), high-throughput sequencing, and modern mass spectrometry. These tools now provide the capacity for near high-throughput environmental analysis of DNA, mRNA, proteins, and metabolites.

For a relatively simple microbial community consisting of a small number of species and thriving in highly acidic drainage from a mine, BER supported groundbreaking metagenomic research that reconstructed near-complete genomes of novel microbial species, characterized the community's structure, and revealed important insights into the community's shared metabolic network (Tyson et al. 2004). Subsequent research characterized the metaproteome of this community (Ram et al. 2005). The greater complexity of most natural microbial communities, however, currently limits the applicability of these metaomic methods to other soil communities of interest to DOE.

An important part of understanding biological processes and systems in natural environments is informational biology—the flow and pools of information molecules in ecosystems that are elaborated by abiotic or biotic transmitters and responded to by receivers in the environment. This physical flow of information can include nucleic acids and other information-rich molecules, smaller metabolites, and signaling compounds. Characterization of these molecules shared among different organisms in ecosystems—coupled with the explosive growth anticipated for metaomic analysis of diverse environments—will produce a tremendous bioinformatic resource base. If successfully managed, this resource will be ripe for synthetic exploitation across many areas of enzyme, organism, and system development and improvement.

Ultimately, computational resources will have the daunting task of dealing with the large datasets needed to address the temporal and spatial heterogeneity of environmental analysis. Such capabilities will be necessary to provide critical dynamic information on organismal (gene) distributions and responses (messages) in ecosystems. Though very promising, current research in this area has yet to address major issues like quantitative standard analysis, sampling representativeness and replication,

Metagenomics of Natural Microbial Communities



Ultrasmall ARMAN cells have a compact genome with only 1 million base pairs and a diameter of about 200 to 400 nanometers, roughly one-fifth the size of the average bacterial cell. The pink rod-like extensions and the gold lemon-shaped objects at the cell surface are viruses. [Image credit: Luis R. Comolli, Lawrence Berkeley National Laboratory.]

Transcending the limitations of investigating isolated organisms cultivated in the laboratory, metagenomics—the sequencing and analysis of genetic material extracted from natural microbial habitats—is expanding our understanding of the structure and function of microbial communities. Although only a few decades old, the ability to recover and analyze nucleic acids from environmental samples (representing the indigenous biological community) has revolutionized the approach to measuring the biomass, diversity, and functional activity of microbes in the environment. Recommendations for advancing metagenomics research, which has great potential for contributing to the missions of many governmental agencies and commercial stakeholders, were described in the National Research Council report, *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* (2007).

Metagenomic sequencing of microbial communities is the foundation underlying measurements of gene transcripts, proteins, metabolites, and other molecular species actively expressed by these communities. Genomics, transcriptomics, proteomics, metabolomics, and other global measurements of the molecular-level activities of microbial consortia can collectively be described as "metaomics." As sequencing and high-throughput experimental technologies continue to advance, these different metaomic approaches will become more integrated. For example, concomitant advances in PhyloChip and functional gene arrays have produced significant capabilities for characterizing small subunit rDNA biodiversity and functional gene abundance in complex environmental samples. However, chip-based arrays are expected to be overtaken by high-throughput sequencing methods to produce an integrated genomic and transcriptomic analysis of environmental samples. Growing evidence suggests that pyrosequencing approaches are progressing rapidly to provide snapshots of 16S rDNA diversity in environmental analysis.

Metaomics will inform and reinforce the continued support of genomic and systems biology research on isolated microbial populations such as the critical subsurface anaerobes *Geobacter*, *Shewanella*, and *Desulfobacter,* as well as other microorganisms important to the Department of Energy's Office of Biological and Environmental Research. Studies of cultured organisms can provide key reference points for interpreting metaomic data, and metaomics can help enable the growth and analysis of microbes considered "unculturable" by shedding light on the metabolic and environmental requirements of these microbes.

Complementary laboratory-based and *in situ* approaches to exploring the microbial world are revealing new insights and hypotheses addressing the functions and relationships among community members. For example, metagenomic sequencing of a low-complexity community thriving in acid mine drainage (AMD) led to the discovery of unique microbes so small that they are practically invisible to a light microscope (Baker et al. 2006). Known as ARMAN (archaeal Richmond Mine acidophilic nanoorganisms), these microbes represent a completely new phylum and contain many genes that have never been seen before. Using 3D cryogenic electron tomography to image AMD community samples (Comolli et al. 2009), researchers were able to determine the size and structure of ARMAN cells, which appear to be free-living, and several are associated with viral particles (see figure). Although the role of these microbes in the AMD community is not yet known, this research shows that combining metagenomics with 3D imaging can provide some valuable insights into the physiology and functions of uncultivated organisms.

efficiency in extracting organisms or nucleic acids and proteins, and the ascendant challenges of data analysis. Fully attaining this capability also requires the assembly and bioinformatic analysis of databases likely to be orders of magnitude larger than current genomic databases for the hundreds of species thus far examined by DNA sequencing. Investigating the vast array of cellular constituents (e.g., RNA, proteins, and metabolites) present under many different biological and environmental states for every genome of interest will significantly multiply the data generated from future genomic studies.

DOE Relevance and Potential Impact of Systems and Synthetic Biology

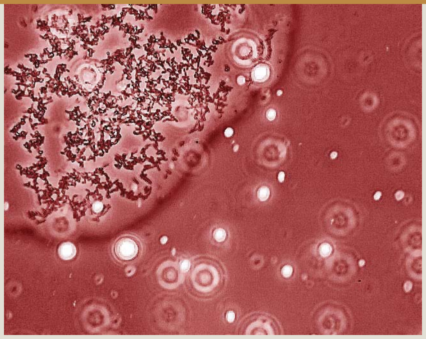
To address the issues of climate change, energy, and remediation of legacy waste, DOE will lead in integrating systems and synthetic biology approaches with conceptual and numerical models that elucidate and predict the role of biological systems in the changing *in situ* environment. Research activities and developments in these complementary fields of study naturally converge to produce an analytical framework to:

- Contribute fundamental new knowledge about the mechanisms of biological response of ecosystems to stressors from environmental change.
- Provide mechanistic insight into collective, adaptive, and emergent properties of complex systems.
- Promote development of biobased and bioinspired materials, processes, and technology to meet energy, fuel, feedstocks, remediation, and carbon management challenges.

Applying this analytical framework to natural ecosystems will extend systems biology concepts and synthetic biology tools to the investigation of complex dynamics and adaptations of microbial communities, molecular interspecies interactions (e.g., microbe-microbe and plant-microbe), and emergent properties. In addition to advancing the understanding of (1) community gene networks, (2) regulatory circuits, and (3) information processing underlying fundamental microbial and plant processes and their interactions within the environment, this integration of systems and synthetic biology with ecosystem research will directly contribute to the following objectives:

1. Developing new organisms and genetic resources to create novel, optimized processes for bioenergy alternatives ranging from gasoline replacements and renewable biodiesel to methane (CH_4), hydrogen (H_2), and electricity (see sidebar, Applying Systems and Synthetic Biology: A Bioenergy Example, p. 12).
2. Mechanistically understanding control points and fitness determinants that may be modified within ecosystems to permit rapid adaptation to changing climate regimes and global change stresses.
3. Improving the capacity, productivity, and sustainability of biomass resources to serve as reliable and plentiful feedstocks for fuel and fiber.
4. Optimizing carbon cycle flow for greater terrestrial carbon uptake and soil sequestration to create an intermediate sink for excess fossil carbon release.
5. Interceding in carbon mineralization pathways of geologically stored labile carbon to inhibit release as CH_4 and other GHGs in response to global warming.
6. Manipulating soil-plant-microbe interactions to ensure sustainable contaminant and radionuclide remediation in the soil and deeper subsurface.
7. Creating modular components of the biofuel production system to provide essential ecosystem services based on life-cycle assessments and to support sustainable bioenergy choices and paradigms.
8. Discovering, understanding, and using environmental processes to optimize designs of biomaterials and bioinspired products and reduce net energy intensity.

Applying Systems and Synthetic Biology: A Bioenergy Example



A mass of *Escherichia coli* cells (upper left corner) sequester themselves from the oily biodiesel droplets they produce. [Image credit: Eric Steen, DOE Joint BioEnergy Institute.]

A sustainable biofuel economy depends on achieving high biomass production at low cost, generating minimal (or ideally beneficial) environmental effects, and using lands not suited for food production. Biological research is needed to contribute to five major objectives in the biofuel economy:

1. Improve plant biology important to productivity and develop plant products that are more easily converted to liquid fuels.
2. Address potential plant-yield reduction resulting from the impact of various biotic and abiotic stresses.
3. Optimize microbial processes for efficiently converting plant carbon to advanced fuel products.
4. Understand and improve soil-water-plant-microbe relationships so that the microbes aid sustainable, low-cost plant productivity on marginal lands envisioned for use by the biofuel economy.
5. Identify the suite of environmental and socioeconomic conditions under which bioenergy crop production is most suitable.

For microbial production of improved biofuels, metabolic engineering and systems and synthetic biology approaches already have demonstrated major success in defining and developing gene networks and control strategies. These approaches have led to the development of controllable cells and metabolic activity for process analysis, sensing, and new synthesis routes for alternative liquid fuels such as biodiesel (see figure) and isobutanol (Atsumi et al. 2009; Brynildsen and Liao 2009; Steen et al. 2010). Results from this research also are helping to characterize microbial response networks associated with different biofuel products and providing key insights into mitigating stress from increased biofuel production. The availability of knowledgebases relating genes to metabolism and the use of genetic tool kits such as "BioBricks" to create novel architectures in well-characterized microorganisms are paving the way for continued success in the production of biofuels that can directly replace petroleum products. The long-term challenge is to integrate and "scale up" this new level of reductionist understanding to remove or minimize current inefficiencies in the overall biofuel production process.

Grand Challenge Research Recommendations

Integrated applications of new systems and synthetic biology approaches are helping to build connections from genome- and molecular-level research to investigations at physiological and ecosystem levels. Though discussed as individual challenges, the future long-term research opportunities described in this section have a natural connection and progression among biological systems important to DOE. For example, achieving the goal of accurately modeling a cell, a community of cells, and an ecosystem of organisms provides the ability to predict or, indeed, manipulate how all of these respond to (or affect) local or global environmental change.

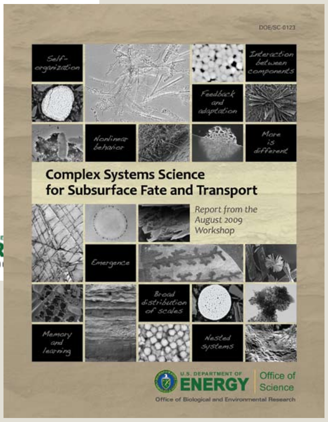
The following grand challenge recommendations start with research needs focused on systems-level understanding at the cellular level and then move

BER Workshop Reports Focus on Characterizing Complex Biological Systems

Advancing the resolution, throughput, speed, and temporal and spatial requirements of research technologies is essential to revealing the connections between the genome and functional processes occurring at cellular and higher scales. Thus the importance of improving tools, techniques, and probes to characterize biological and ecological systems is well recognized by the research community. In addition to the needed observational, experimental, and computational capabilities identified in the section, Grand Challenge Research Recommendations, p. 12, the reader is encouraged to examine two recent Department of Energy (DOE) Office of Biological and Environmental Research (BER) reports that summarize relevant research and technology development needs defined by the scientific community.

In May 2009, BER supported a workshop on developing a new generation of technologies for characterizing cellular- and multicellular-level functions related to DOE missions. The diverse scientists and engineers participating in this workshop presented their findings in the report *New Frontiers in Characterizing Biological Systems* (DOE/SC-0121).

DOE/SC-0121
U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE
New Frontiers
in Characterizing Biological Systems
Report from the May 2009 Workshop



DOE/SC-0122
Complex Systems Science
for Subsurface Fate and Transport
Report from the August 2009 Workshop
U.S. DEPARTMENT OF ENERGY Office of Science
Office of Biological and Environmental Research

In another BER workshop held in August 2009, researchers with expertise in complex systems and environmental, microbial, and Earth sciences described tools and approaches needed to quantify and predict complex behavior in the subsurface. Output from this workshop was published in the 2010 report *Complex Systems Science for Subsurface Fate and Transport* (DOE/SC-0123).

on to challenges associated with investigating systems of increasing biological complexity. The recommendations also address improvements in the technologies and techniques that support the science described in these grand challenges. For other recent BER reports summarizing needed tools and approaches identified by the research community, see sidebar, BER Workshop Reports Focus on Characterizing Complex Biological Systems, this page.

Enabling Predictive Biology

2.1 Develop a simulation model of a single cell for accurately predicting phenotype from genotype-environment interactions.

Genomics, transcriptomics, proteomics, metabolomics, high-resolution imaging, structural genomics, and structural dynamics have begun to provide a "parts list" of individual cells. In addition, systems biology approaches have revealed partial schematics of cellular systems, but we still do not fully understand the workings of a complete cell from any one organism. A fully functional, predictive model of a single cell—the smallest unit of a living system—represents a grand challenge that underpins many of the recommendations in this chapter.

A dynamic model of an entire cell will require selecting one or more cell types (e.g., a eubacterium or some other single-celled microorganism or a plant cell) and integrating new research results with all existing data and scientific studies relevant to that cell type (see sidebar, Holistic, Dynamic Model of a Single Cell, p. 16). Workshop participants suggest the following criteria for choosing a single-cell system for intensive study:

- DOE relevance (e.g., biogeochemical cycling and cellulose metabolism).
- Multiple genome sequences available for comparison.
- Genetically tractable (e.g., organisms amenable to mutagenesis and gene transfer).
- Extant genome-scale models.
- High-throughput phenotyping available.
- Ability to make comparisons within an evolutionary context.

The entire genetic, transcriptomic, and metabolic networks of the cell should be represented *in silico*, and where protein, metabolite, and kinetic parameters do not yet exist, they should be range estimated from the closest, best-available data subject to parameter refinement as new data are developed. The model should be able to simulate microbial growth and reproduction from a minimal growth environment.

2.2 Establish new model organisms for relevant ecological process understanding.

Identifying and developing new organismal models highly relevant to bioenergy production and carbon-cycle science are critical research needs for BER. *Escherichia coli* remains the primary model for most microbial systems biology research. As we move to a complete understanding of all of its network and metabolic functions under diverse environmental regimes, *E. coli* will continue to serve as a foundational model for all other organisms. Capturing the greater network and metabolic versatility prevalent in the environment is needed for organisms and processes representing key biogeochemical cycling components (e.g., cellulose degradation, methanogenesis, and methane oxidation), aspects of plant host susceptibility, and other fundamental ecological properties. These organisms should be elevated to the equivalent state of network understanding anticipated for *E. coli* over the next decade. BER has contributed to an advanced understanding of some potential models for subsurface biogeochemistry—*Geobacter, Shewanella*, and *Desulfovibrio*. Clearly needed are genetic systems for efficiently characterizing and modifying new model organisms relevant to the carbon cycle, especially photosynthetic organisms and rhizosphere heterotrophs and azotrophs.

2.3 Use robust biochemical, functional, and experimental evidence to enhance genome and metagenome annotation.

The predictability and functionality of a comprehensive single-cell model or any newly designed gene networks, circuits, organisms, and communities highly depend on the robustness and accuracy of genome annotation. For existing models, there is an uncertainty concerning the veracity of annotated genomes—an issue that will be confounded further as more synthetic biology strategies are undertaken on a greater diversity of network elements and organisms. For metagenomics, this issue likely will become a critical element to future progress in design and application.

2.4 Develop the biological understanding for generalizing and applying models from simple systems to more complex systems, working toward an ultimate goal of modeling the evolution and dynamics of a complex biological system under environmental stress conditions.

From the single-cell level, we should be able to advance to predictive models of communities of cells, a microbial colony, a leaf, an organ, organisms, and then to communities of organisms. This will involve improving our understanding of gene networks, epigenomics, cell-cell communication, information flow, tissue interactions, horizontal gene transfer, and hybridization. Workshop participants envisioned a long-term goal of predicting how a particular crop variety would grow in different environments simply by knowing its genotype. Achieving this higher level of predictability will require understanding how organisms and communities of organisms drive or respond to environmental change.

Complex biological systems and ecosystems evolve in response to a variety of environmental stresses (Comte et al. 2010; Zhang et al. 2009).

Understanding, diagnosing, and predicting these changes and their dynamics are critical to sustaining natural systems and designing new synthetic, stable biological systems. By identifying key markers and impact factors, researchers will be able to develop capabilities for detecting, modeling, and predicting the loss of community diversity, changes in metabolic potential, and the emergence of increased sensitivity to additional stresses.

With continued improvement in the availability and utility of models for specific cellular and developmental systems within a selected species, a natural extension will be to investigate and dynamically model the *evolution* of such systems as they interact with their environment and respond to various stresses. Types of fundamental evolutionary changes include both changes of molecular identities within roughly constant networks and changes of network superimposed on roughly constant molecular components (e.g., major changes in expression and morphology resulting from regulatory alterations in a key developmental control gene, as reported in Chan et al. 2010). Reconstruction of ancestral biomolecular networks may become possible, providing clues to some otherwise unobservable ancient network functionalities. Understanding the past evolutionary dynamics of cellular and multicellular networks, especially at the level of dynamic models, also could contribute significantly to predicting future evolutionary response to altered environments. An example problem for evolutionary projection is estimating the number of generations required for a given adaptation, which can be very large (Blount et al. 2008) or very small (Grant and Grant 2006).

When moving outside the laboratory or beyond studies of artificially created communities, we are immediately faced with the complexity and unique environmental context that must be addressed when studying natural systems. Efforts should be made early and continuously to integrate knowledge from models of less complex organisms (see sidebar, Holistic, Dynamic Model of a Single Cell, p. 16) or environments and then test these models against natural communities. In the field of metagenomics, analyses of low-complexity microbial communities (Tyson et al. 2004; Ram et al. 2005) are providing the foundational insights, research capabilities, and computational tools needed to help enable investigations of more complex natural communities important to DOE.

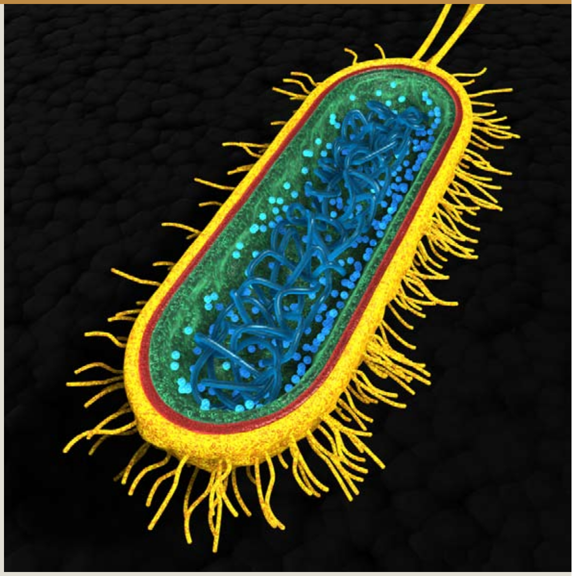
2.5 Identify general design principles used in natural systems, articulate these, and then use them to distill our understanding of biology.

One possibility offered by a systems approach to investigating biological function is the identification of general design principles for biology at different levels. We already have some information about the design of gene circuits, which are comprised of small modular circuit elements such as feedback and feedforward modules (Alon 2007). These modules appear to be the building blocks of gene circuit evolution (Gao and Davidson 2008). Synthetic gene circuits have been assembled from modules of regulatory genes to connect genomic information to network understanding of the control and regulation of proteins involved in particular cellular processes (Hasty et al. 2002). Similar circuit diagrams are being drawn now for biology at the cellular level, with circuit elements for cell-cell interactions also showing modular construction. These circuits, however, have additional emergent properties that arise from cellular signaling, leading to changes in gene regulatory networks, and from new types of interactions (such as plant hormone transport), adding a spatial element to cell-level interactions that are absent in gene regulatory networks (e.g., Jönsson et al. 2006). These tools and approaches also can provide a conceptual framework for analyzing a metacommunity and facilitate the development of additional tools to probe network connectivity within the environment. For example, research on metagenome and metaproteome networks and modules could reveal insights into the fundamental regulation of community or ecosystem processes. In a conceptual sense, these systems are considered complex and adaptive (Levin 2002; Hartvigsen et al.1998) and should

Holistic, Dynamic Model of a Single Cell

Can all the molecules required for life and their interactions be comprehensively described for one cell?

An ultimate goal is to have computational, predictive models that describe organisms in their environment, simulate the dynamic interactions within these organisms and their communities, and incorporate the relationships and feedbacks that occur among organisms, their environment, and the climate system. However, the complexities of these interactions currently are beyond our understanding and likely will continue to be so 20 years hence. Therefore, an initial step toward this ultimate goal is identifying a simpler, more tractable system for which a truly holistic model could be achieved. Workshop participants suggested developing a comprehensive, dynamic model of a single cell (e.g., a bacterium) as an attainable goal over the next two decades. Creating this "simpler" model would provide insights needed to build the larger community or ecosystem models that we hope to have someday.



Developing a comprehensive, dynamic model for a bacterial cell with some manageable number of genes, cellular components, and outputs could help us determine how much and what kinds of biological information are needed to attain reliable predictions. [Image credit: TurboPhoton.]

The DNA sequence of a cell's genome allows us to predict, to a first approximation, what the entire proteome will look like, at least in the case of prokaryotes. However, the detailed map of the composition, structures, and interactions of all other macromolecules (e.g., RNAs and polysaccharides) and small molecules (e.g., metabolites and the lipids that maintain membrane structure) remains a mystery. Even the least complex bacterial cells have hundreds of thousands of such molecular species. Vast improvements in modern analytical instruments, particularly mass spectrometers, present the possibility of determining the molecular weights and structures of all the molecules required by a cell—from the smallest metabolite to the largest macromolecular structure. For the first time, we would be able to access the full "parts list" of a single cell. Comparable to the publicly funded Human Genome Project, which required 400 DNA sequencers, determining the full molecular contents of a single cell also may require only a few hundred dedicated instruments, together with supplies, personnel, and informatic support, focused on this one goal. This research also would need to characterize the interactions between the cell's molecules and analyze the dynamic changes that occur in response to environmental perturbations.

In addition to empowering systems and synthetic biology, the impacts of this project could be as diverse and far-reaching as those of the Human Genome Project. Once the structures of all molecules required for life are known and standards become routinely available, this information undoubtedly would find direct application to the study of other organisms, both prokaryotic and eukaryotic. A single-cell model represents a cost-effective first step toward predicting *in silico* what the effect of any genetic or environmental perturbation would be on that cell's biochemical machinery. In essence, having this predictive capability would place all of life's biochemical processes under our control, and, extended in the future to more complex organisms, it could transform currently unpredictable, hit-or-miss paradigms for agricultural production of energy crops or other biology-driven practices to predictable engineering endeavors.

yield to the same analytical approach. Many new modules remain to be discovered, and the computational environment for using them as building blocks to understand basic principles of gene regulation and organismal development—and hence to invent new types of networks—is only just emerging. These same design principles and circuit motifs may operate at other biological levels—not only in more complex types of cell-cell communication (as in the interactions between plant roots and soil microbes), but at higher levels of biological organization. Such levels include interactions between tissues and organs, interactions between members of the same species in populations, interactions with other species in ecosystems, and interactions with the abiotic environment. Discovering and applying these organizational principles are the great challenges in biology for the next 20 years.

Measuring and Analyzing Biological Systems

2.6 Apply advanced computational and analytical capabilities to characterize the information molecules and network interactions used by biological systems.

Computational resources and the ability to resolve small information molecules, particularly by mass spectrometry, are giving rise to rapid expansion of the field of infochemicals (Dicke and Sabelis 1988). Many of these informational chemical signals and transfers are the result of recognition by receptor proteins, most of which remain uncharacterized. This area presents promising opportunities for structural biology research to access powerful computational tools and protein docking models to understand the molecular basis of an interaction. This approach also could aid in predicting the nature and ecosystem function of unknown information molecules and receptors in the environment. New biologically inspired research in this area can take advantage of tremendous computational resources (see Chapter 5: Grand Challenges in Computing for Biological and Environmental Research, p. 43) and complex protein characterization capabilities rapidly expanding within DOE laboratories. This is a massive research area potentially yielding hundreds of thousands of single nucleotide polymorphisms (SNPs) for kinases, transporters, transcription activators, and other proteins with recognizable domains but unknown biochemical confirmation within the metatranscriptome and metaproteome. Thus this effort requires coordinated, high-throughput, collaborative biochemical analysis among investigators nationally and perhaps internationally.

2.7 Improve capabilities for imaging a single cell at a resolution of one molecule per cell.

New demands are increasing for single organism–single molecule resolution to further reduce variation and experimental noise, increase throughput, and reduce cost. Great progress has been made in areas such as single-cell isolation by fluorescent-activated cell sorting, single-cell resolution of fluorescently probed or reporter cells, single-cell PCR amplification, and single-cell gene introductions and silencing. However, ease of use, convenience, and throughput for these technological advances need to be improved to meet future demands. Additional improvements are needed in developing multichannel formats to meet dynamic throughput requirements, sequencing from one molecule in a single cell, and achieving dynamic detection of subattomole concentrations (perhaps down to a single protein molecule or metabolite) to address subcellular rates and fluxes.

2.8 Measure microbial processes and interactions in the real world and in experimental simulations.

In situ or *in vivo* observations are difficult to make. Often, the acts of sampling and sample processing introduce uncertainty into the analysis by perturbing genetic expression and physiological function or by debilitating and killing resident organisms and interactions. The biofilm and environmental research communities deal with these issues by applying robust experimental simulations. For systems-level modeling, real-world physiology must be defined and kinetic parameters quantified.

The nonculturability issue also needs further testing and experimental verification to predict the degree of manipulation and control that can be exerted over the system either from a synthetic biology or control engineering standpoint. If indeed large portions of the community are found to be truly nonculturable, then new genetic tools, gene transfer systems, or nanophysical techniques are needed to permit manipulation and experimental observation of organisms in their complex environmental settings.

2.9 Provide standardizable, reproducible sampling protocols and observations enabling functional characterizations of systems and synthetic design.

Key issues for implementing synthetic biology at the environmental scale are accurate, precise observations of fundamental network elements to understand the system(s) of interest to DOE. Given the complexity of the environment, a critical need is establishing protocols that will provide the necessary observational information and will enable comparability across laboratories and environments. The types of observability needed to achieve computational design include information flow (genes and metabolites) within and among populations, knowledge of how stochastic processes are used and propagated in the system, influences of circadian rhythms and diurnal cycles, and metagenomic to proteomic component data.

2.10 Define the range of analytes, ligands, and fluxes of materials to be measured.

We anticipate that virtually all ranges of biomolecules and their life spans will need to be identified as systems and synthetic biology strive to develop a more complete understanding of host metabolic network interactions under dynamic environmental conditions. These biomolecules might include numerous small molecules (e.g., signaling molecules and metabolites), as well as proteins, and require visualization on the scale of nanometers to centimeters. Initially, these requirements are likely to manifest themselves at the level of segmentation and compartmentalization within the cell or at the organism level, but they will rapidly expand to more dynamic issues relative to tracking and trafficking across tissues and to the environment, including biofilms, and, ultimately, into complex matrices such as soils. Measurement of carbon and nitrogen fluxes within these systems, along with determination of phenotypic growth rates, will be a challenge to overcome.

2.11 Define scalability requirements for measurement technologies and develop capabilities for in situ *and* in vitro *sensors.*

Current measurement capabilities range from remote satellite imaging to monitoring individual cell-cell communications and applying point-of-use sensors. Hyperspectral satellite imaging is required to achieve individual plant species quantification and physiologic function within a landscape. At the microbial level, attaining a similar capability within a water body would be desirable. Although measuring physiologic function at the microbial species level is unlikely, measuring certain community-level metabolic activities—such as differential chlorophyll dynamics for discriminating photosynthetic production rates—is achievable. Pigment production, marl formation, and artificial reporter tags (e.g., fluorescent proteins and bioluminescence) are further examples of targets for remote sensing with systems-level applications. In subsurface- and rhizosphere-oriented research, measurements of environmental parameters, such as pH, redox, and specific chemistries, are needed within communities and also at the level of cell-cell interactions. These needs could be met with a range of chemical sensors on scalable platforms, biosensors, and surrogate biosensors and bioreporters. Process- and product-specific gene(s) or protein targets need to be defined in order to guide sensor development in these areas. In addition to sensors that measure specific properties at a specific scale, approaches or tools are greatly needed for providing integrated signatures of macroscopic processes that are indicative of complex system transitions and behavior at scales relevant to ecosystem or climate predictions.

Exploring Ecosystem Function and Elemental Cycling

2.12 Understand, predict, and manipulate the types and rates of ecosystem responses and feedbacks that result from and influence climate change.

As environmental and biological systems respond to climate change, new populations could begin to dominate communities, or important species could be lost, potentially leading to unanticipated consequences on the delicate balance of ecosystems. Relatively modest changes in climate could cause significant changes in organisms and ecosystems that provide essential functions and services. For example, the spread of new tree pests could threaten forest health, thereby amplifying carbon emissions through the loss of forest biomass. Another possible outcome could be the emergence of new plant pathogens or reservoirs of human pathogens that cause ripple effects on society, industry, the economy, and the environment.

Research should continue to fully explore the diversity, complexity, and feedbacks at work in natural environments subjected to multiple stress conditions including large-scale stress such as climate change. The challenge then is to understand the feedbacks between climate change and evolutionary changes in ecosystems—from the molecular to global level—and predict how these feedbacks will impact natural and human-altered environments. Meeting these challenges will require DOE investment in the development of new methods to explore and model ecosystem dynamics, evolution, and interactions. This will involve investigating evolutionary adaptation by population changes, genotype changes, plasticity, and the evolution of ecosystems when new organisms are introduced. A true systems biology approach to this problem needs a combination of experimental and computational methods to understand how things happen and predict what is likely to happen *in situ* and under various scenarios. To achieve a predictive understanding of biological response to climate and environmental change, research support for mathematical modeling should be integrated with data gathering and experiments designed to probe mechanisms for climate-driven changes in organisms and their ecosystems.

The connection between ecosystems and sustainable energy requirements is a key reason why BER should be a leader of this research involving biological systems at the climate-energy interface. For example, biologically derived fuels likely will play an increasing role in meeting U.S. energy needs and reducing carbon emissions in the near future, and the ability to produce such fuels on a large scale depends heavily on the interaction between climate and local ecosystems. In response, ecosystem changes can lead to shifts in patterns of human settlement, activity, and energy use. In another example, geoengineering has been discussed as a possible strategy to mitigate global warming (see sidebar, Understanding the Potential Consequences of Geoengineering, p. 32). For geoengineering to be considered a serious option, the potential large-scale effects—which cause planet-wide changes in climate—and local impacts on ecosystems and biological processes must be understood and put into context.

2.13 Deploy synthetic (or nonsynthetic) biology to understand and manipulate ecosystem function.

Even rudimentary developments of current synthetic biology can be tested and exploited for environmental and ecological use, albeit largely under laboratory or mesocosm simulation. To obtain useful, realistic measurements and kinetic data from natural and synthetic microbial communities without going into the field, a key challenge identified by the workshop participants is the ability to construct laboratory mesocosm environments that are quantitatively accurate representations of actual natural environments. These "realistic" mesocosms would provide a suitable environment for testing synthetic microbial systems in a controlled setting. BER has supported pioneering the use of designed organisms to enhance process understanding at the field scale (Ripp et al. 2000). Recent advances in genetic containment and control strategy greatly improve the prospects for more expanded fundamental studies in this area to meet future needs. An example of a natural system where systems and synthetic biology approaches could be used to advance our mechanistic understanding of microbial activity is the interplay between microbial-geochemical transformations and water advection in the subsurface. This microbe-environment interaction exhibits intermittent and threshold-governed feedbacks characteristic of complexity. Systems-level laboratory and field experimentation could be tightly coupled with new top-down modeling strategies to address questions that could advance our ability to predict and manipulate microbial communities in dynamic natural environments. Key research questions include:

- Which environmental variables control microbial community structure and function?
- What are the relationships between structural and functional heterogeneity in microbial communities, and how do these relate to the changing properties of the natural physical and geochemical environment?
- Which forces drive or limit adaptation of individual populations and the microbial community to changing environmental conditions?

Targeted research on these and other topics can be projected to provide information on failure or success rates. Such research would aid in gathering fundamental knowledge to understand critical environmental parameters, structural and mechanistic processes, and impediments to future success.

2.14 Determine the molecular basis of robustness, fitness, and selection.

A major challenge in importing new functionality or populations into ecosystems is a lack of fundamental understanding on how new genotypes and phenotypes are selected and maintained within the systems. Whether they can be established in an unoccupied niche or must displace others from occupied niches is not well understood. Also unclear are the fundamental properties contributing to the fitness of organisms or genes to be introduced into and maintained in the system. Intricate networks, genetic exchange and information flow, and modularity of system components are expected to contribute to these properties, which need to be identified and characterized at the ecosystem level. However, these are largely untested hypotheses that must be evaluated to achieve future success. Indications suggest that modern agribusiness already may be exploring these issues from a systems perspective to delineate selective characteristics for which certain breeds may be more successful than others.

2.15 Develop a complete understanding of the biogeochemical cycles important to regulating carbon flux through biological systems.

Earth's natural carbon cycle is driven by primary productivity and organic carbon decomposition. The dynamics of this cycle are essential to (1) food, fiber, and biofuel production; (2) carbon storage in biomass, soils, and sediments; and (3) atmospheric CO_2 concentration. The extant biology and its environment determine the fluxes in the carbon cycle. Understanding, predicting, and wisely managing this cycle require a systems-level understanding of microbes, plants, biological communities, and ecosystem biogeochemical processes. This systems approach is necessary because of complex interactions—both within biological communities and with controlling environmental parameters such as moisture, nutrients, soil structure, temperature, and light. In particular, the carbon cycle is highly dependent on nitrogen and phosphorus and, in certain geographic regions, other nutrients like sulfur, iron, and micronutrients. Thus, the carbon cycle cannot be productively understood in isolation. A challenge for the next 20 years is ascertaining how to link knowledge from the systems biology of the cell with that from ecosystem science in ways that provide mechanistic understanding at different levels of resolution. If successful, systems biology—now practiced at the level of the molecule, cell, and organism—becomes "systems biology of the biosphere."

2.16 Apply functional metagenomics to enable mass balance closure for biogeochemical cycles and transfer this information to biosystem design.

Given the need to couple dynamic biogeochemical cycling submodels with Earth system models for climate-consequence projections, applying metagenomics and systems biology approaches can help identify and understand uncharacterized biological components of the carbon cycle and other biogeochemical cycles. These insights can help reduce the uncertainty of long-term projections for atmospheric carbon levels and climate change. Our knowledge of biogeochemical cycles is not static—only a decade ago, essential new processes were discovered, elucidating anaerobic ammonia oxidation in marine sediments. This discovery helped to complete the mass balance closure for nitrogen cycling in that system. Coupling carbon decomposition with dynamic aspects of GHG release in marine and terrestrial ecosystems appears essential for coupling of climate and global temperature predictions and related feedback responses.

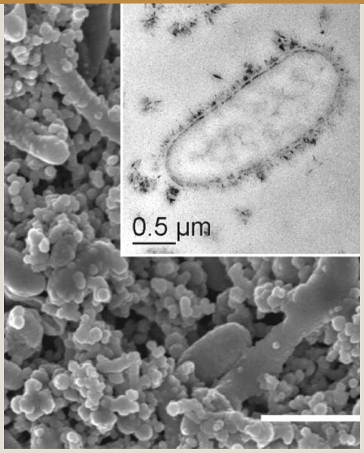
Sequencing from environmental samples has enormous potential for advancing gene discovery, but the power of metagenomic analysis is severely limited in the ability to reconstruct more than a few genomes, let alone hundreds or thousands that exist in a given sample. Metatranscriptomic and metaproteomic approaches, as well as more advanced analytical techniques for exploring system response, must contend with the fact that recovery of nucleic acids and proteins from environmental samples is low (often <20%) and highly variable. These advanced analyses are critically important if gene network architecture can be developed to understand regulatory controls of processes within the system. Thus resolving low-abundance genes and transcripts in a complex, high-biomass background represents another important challenge. Perhaps a user facility, similar to the DOE Joint Genome Institute, to support metacommunity analyses should be established.

2.17 Manage plant and microbe stress response to control carbon biosequestration and remediation of metals and radionuclides.

Using plants and microorganisms in remediation of contamination is well established but far from optimal in its application at broad scales (see sidebar, Application of Systems Biology: Environmental Remediation, p. 22). To achieve new solutions for environmental remediation, greater progress in engineering interactions between these biotic partners is needed. Another barrier to research progress in this area has been the limited interest in interchanging component parts to create new products or processes that can support large-scale engineering use. Developing rhizosphere communities to help plants deal with nutrient and essential element acquisition during hydrologic or temperature stress may be one approach to maintain sustained rates of primary production during less-than-optimum growing conditions. Likewise, channeling excess photosynthate to the rhizosphere could promote biopolymer synthesis for metal immobilization or control the redox environment to accentuate metal or radionuclide reductive conditions. We may need to conceptualize a "virtual rhizosphere" (perhaps analogous to the "virtual liver" for toxicogenomic studies by the U.S. Environmental Protection Agency and the National Institutes of Health) to begin assembly of the tools, models, and databases that can provide the necessary capabilities for complex system research in ecosystem analysis.

The response of natural ecosystems to environmental stress is largely a result of interactions among species comprising the ecosystem. Positive and negative feedbacks of these interactions are analogous to those observed for gene regulatory networks and metabolic flux within individual cells. These interactions are hierarchical and may be manifest at the system level but unobservable. Regardless, they may act as buffers or mediators of change, acclimation, adaptation, or system evolution. In natural soil and subsurface environments,

Application of Systems Biology: Environmental Remediation



Background image is a scanning electron micrograph of microbes and particles sampled from a subsurface community at a contaminated site in Rifle, Colorado. Inset shows transmission electron micrograph of a single microbe from this community with dark nanocrystals of iron and sulfur forming at the cell membrane. [Image credit: Williams et al. 2009.]

Contamination of soils and sediments by radionuclides, heavy metals, and organic chemicals at sites that were once part of the nuclear weapons complex remains a longstanding legacy for the U.S. Department of Energy and its predecessor agencies. Some of these contaminants have escaped their original containment structures, and those that are soluble have moved with the groundwater and present human and environmental risks. Certain microorganisms within natural subsurface communities can reduce problematic radionuclides and metals, rendering them insoluble and blocking their further movement (see figure). Several can oxidize organic pollutants and some metals, and others can bind or sequester pollutants. Metal reduction and other types of remediation processes will need to be sustainable over many years. Scientific progress that can bring lasting solutions to these problems will require a new understanding of very long term interactions among microorganisms, minerals, migrating fluids, and dissolved constituents over a wide range of scales. The long-term challenge is to advance from the mechanistic knowledge of potentially useful microbial processes to a systems-level, predictive understanding that enables optimization of microbial functions *in situ*.

interactions between the microbial community, minerals, dissolved constituents, and water flow occur within a heterogeneous physical framework. Future research on interactions within these environments needs to transcend reductionistic mechanisms and integrate with metacommunity analysis. A particular focus should be placed on understanding biogeochemical dynamics at the mineral-microbe interface, microbial community responses to variable environmental conditions, biogeochemical reaction rates in heterogeneous media, and feedbacks between biogeochemistry and water flow. This continuous, iterative process ultimately will lead to useful models and improved understanding of *in situ* biological systems.

Horizontal and interspecies transfer of genetic information is well understood. Mechanistically, these processes, whether by conjugation, transformation, or transduction, have been demonstrated at the microbial level in the environment. Transfer mechanisms from microbe to plant also are well known, co-evolutionarily linked, and biotechnically exploitable. Unfortunately, there are few genetically developed systems for organisms from nature, but comparative genomic information already has provided further evidence that microbe-plant transfer mechanisms are perhaps more common than anticipated. This is indicated by the frequency of prokaryotic sequences residing in the genomes of sequenced host species. A recent example is the occurrence of prokaryotic intracellular endophytes in *Populus* and the co-occurrence of gene homologs in the plant genome (personal communication Jerry Tuskan, Oak Ridge National Laboratory).

2.18 Define the fundamental microbial basis for permafrost carbon-methane transformations.

The consequences of large-scale carbon storage in permafrost or in labile carbon pools in frozen Russian Yedoma sediments remain unknown and problematic. The mass of material, if released as GHGs in short time intervals, would aggravate atmospheric CO_2 greatly. Yet, turnover of plant residues is a complex process with many routes, which may or may not lead to GHG releases. Many microbial guild assemblages participate in the process, often competing for carbon sources, terminal electron acceptors and donors, and dependency for balanced micro- and macronutrients—all of which influence terminal carbon releases. The opportunity exists to intercede at any one of these levels to alter the course of carbon releases.

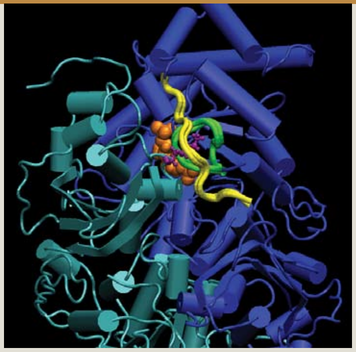
2.19 Develop designs for optimizing carbon flow for biomass production, carbon allocation, and biosequestration to reduce rates of atmospheric CO_2 accumulation and to increase terrestrial carbon storage by 50% in 20 years.

An estimated 55% of the CO_2 released annually from human activities remains in the atmosphere; the rest is taken up by the oceans and terrestrial environment (IPCC 2007). Since 1960, the percentage of human emissions incorporated into natural carbon sinks has decreased 5% and may be expected to diminish further with time (see sidebar, Optimizing Terrestrial Biosequestration of Carbon, p. 24). Microbial biomass is the dominant form of fixed carbon in short-term pools, exceeding that of higher plants. It is entirely feasible to design systems to increase net primary production and CO_2 fixation, alter heterotrophic ability to fix CO_2 in soils, inhibit methane production and escape, and store carbon in novel exo-biopolymers. A challenge should be undertaken to demonstrate this capability at the microcosm and mesocosm scale, with the intention to have a field-deployable system available within 20 years.

2.20 Determine carbon and nutrient dynamics in natural systems.

Critical discrimination among anthropogenic, natural, and "background" carbon sources is needed to predict biotic fluxes and the efficacy of remediation efforts, as well as to measure, report, and verify reductions in carbon emissions. Improvements in the ease-of-use or availability of isotopic labeling techniques also are necessary to provide analytical fingerprints and differentiate fossil-derived nutrient and carbon sources from those generated via contemporary cycling from biomass sources. The ability to discriminate between "natural" (ecosystem) driven versus purely anthropogenic carbon emissions from agricultural, urban, and industrial sources is needed for both experimental and modeling purposes. These needs extend not only to the atmospheric environment, but also to soil, water, and plant and microbial biomass, as well as to the small metabolite pool in the environment. In addition to advancing isotopic techniques, these improvements also are needed for developing reference fingerprints relative to carbon or nutrient pool composition with particular relevance to discriminating slow (recalcitrant) and fast turnover pools.

Optimizing Terrestrial Biosequestration of Carbon

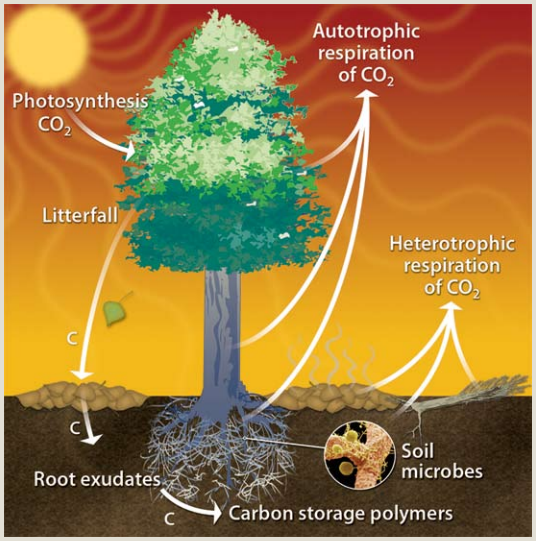


The ribulose bisphosphate carboxylase enzyme (RuBisCO) fixes CO_2. [Image credit: Sandia National Laboratories.]

In the past 50 years the capacity of terrestrial and ocean sinks to take up CO_2 has decreased by 5%, and this ominous trend is expected to intensify. Given the current anthropogenic release of 10 petagrams of carbon per year (Pg C/y), the cumulative effect of losing this natural capacity for carbon storage will continue to exacerbate the atmospheric CO_2 level. As a result, elevated CO_2 may contribute to changes in forest-coupled climate-carbon dynamics that potentially could release 100 to 510 Pg total carbon by 2100. An additional 100 to 750 Pg could come from permafrost thawing and decomposition by 2100. Since land sinks are estimated to take up 30% (or 3 Pg C/y) of annual anthropogenic carbon emissions (Canadell et al. 2007), a 50% improvement in this number would represent a very significant carbon mitigation effect.

Mechanistically, synthetic biology could accomplish this mitigation by increasing carbon sequestration in above- and belowground biomass through a variety of approaches that could channel plant carbon allocation to the rhizosphere, produce belowground carbon storage biopolymers, and shunt carbon metabolites away from pathways that release the greenhouse gases CO_2 and methane (CH_4). Channeling excess photosynthate or metabolites for eventual rhizosphere excretion into productive pathways for biosynthesis would be the goal. Biological approaches to address this goal include

- Enabling the conversion of C3 plants to C4 using pathway reconstruction, perhaps by adding a phosphoenolpyruvate carboxylase.
- Directly modifying ribulose bisphosphate carboxylase by reducing O_2 fixation and increasing enzyme efficiency for CO_2 fixation.
- Increasing heterotrophic CO_2 fixation competency by engineering carboxysomes (bacterial compartments containing carbon-fixation enzymes) into soil microbes.
- Altering root cellulose or lignin content with the intent of long-term renewable carbon storage harvestable as belowground biomass for biofuel feedstocks.
- Developing rhizosphere inocula with pathways that could shunt carbon into storage polymers. In addition to enhancing intermediate-term carbon storage, these polymers would improve soil condition and water-holding capacity. For example, novel pathways for high-molecular-weight heteropolysaccharides are being discovered for wastewater microorganisms, which produce the polymers during growth on low-molecular-weight organic acids such as acetate (a prime substrate for methanogenesis) (Allen et al. 2004). Pathways from these wastewater microbes could be engineered into novel rhizosphere bacteria that generate carbon-storage polymers from substrates for methanogenesis.



By enhancing microbe-root interactions and altering root content, carbon storage in plant and soil biomass could be improved.

3 Grand Challenges in Climate Research

> ***"The Earth is the only world known so far to harbor life. There is nowhere else, at least in the near future, to which our species could migrate. Visit, yes. Settle, not yet. Like it or not, for the moment the Earth is where we make our stand."***
>
> **— Carl Sagan. 1994. *Pale Blue Dot: A Vision of the Human Future in Space.***

Today, we face the challenge of providing energy services for a growing and increasingly affluent human population (especially in the developing world) while maintaining a sustainable climate system. Although difficult, this problem is one of the most important for society to confront and overcome. The Department of Energy's (DOE) Office of Biological and Environmental Research (BER) is poised to play a central role in meeting this challenge through its support of the basic science underpinning potential solutions (see sidebar, BER Contributions to Climate Research, p. 27).

Twenty years from now, policymakers will be planning how to adapt to predicted climate change and to effectively mitigate further changes. Climate science will be able to meet those needs—with more accurate climate predictions and more detailed assessments of the overall impacts of new technologies—provided we improve scientific understanding of the climate system and human interactions with it. We envision six grand challenges (see box, Summary of Research Recommendations, p. 26). Surmounting them will require a deeper insight into the following Earth system processes:

- Ice sheet and glacier dynamics.
- Clouds, aerosols, and precipitation.
- Global biogeochemical cycling of carbon, water, and other natural compounds.
- Soils and terrestrial vegetation, including human decision making about crop choices and management.
- Groundwater resources.
- Biological interactions and feedbacks.
- Anthropogenic greenhouse gas (GHG) emissions.
- Interactions among all of these processes in the human-influenced Earth system.

Elucidating these processes will allow us to more adequately model their effects on climate and the degree to which change in the climate system affects these processes.

Comprehensive observation, analysis, diagnosis, and prediction of all Earth system components are required for successfully addressing the grand challenges. Today's operational analysis is adequate for weather prediction but disconnected from comprehensive Earth system modeling. Observing technologies developed 20 years from now should be able to assimilate the broadest possible range of observations into comprehensive Earth system models (ESMs). The resulting ESMs will document the changing state of the Earth system and predict many aspects of global change. Such integrated analyses and predictions will provide fundamental information to policymakers balancing decisions about mitigation of and adaptation to climate change with other societal needs.

Summary of Research Recommendations

Earth System Models

3.1 Develop higher-resolution models to integrate many more relevant processes than current models and to describe climate change over much longer time scales.

Cloud and Aerosol Processes

3.2 Improve parameterizations and basic knowledge about aerosols that affect clouds and factors that control their number and concentration for microphysics, radiative transfer, and turbulence processes to quantify indirect aerosol forcing and resulting precipitation changes.

Ocean and Terrestrial Processes

3.3 Develop ecosystem-observing systems to monitor biogeochemical cycles, estimate critical process parameters, and provide model tests in ocean and terrestrial biospheres, including subsurface soils.

Biological Processes

3.4 Advance understanding of important biological interactions and feedbacks to identify potential tipping points and possible mitigation strategies such as carbon biosequestration.

Human Interactions

3.5 Improve integration of anthropogenic climate forcings into Earth system models (ESMs) and develop new techniques to evaluate these coupled models at both global and regional scales.

Observing Systems

3.6 Establish new observational technologies and use them to comprehensively compare ESM predictions with observations and to quantify model errors.

Grand Challenge Research Recommendations

Earth System Models

3.1 Develop higher-resolution models to integrate many more relevant processes than current models and to describe climate change over much longer time scales.

A major scientific challenge is to create models that accurately represent the evolution of the entire Earth system. Models have evolved from ones representing only the atmosphere to those that couple the atmosphere, ocean, and land surface and recently to ESMs with active carbon cycles. Now abundantly clear is that a much more complete, process-based model representation of the Earth system will be needed to understand in detail how the system operates and to provide information about how it might evolve. This representation will take into account not only physical processes but also biogeochemical cycles, hydrological and biological interactions and their feedbacks, and human decision making.

ESMs of 2030 will address coupled interactions among the atmosphere, ocean, land, sea ice, land ice, biogeochemistry, dynamic vegetation, lakes and soils, aerosols, chemistry, and changes in energy technologies and resource management. They will represent the reservoirs and evolving flows of energy, water, carbon, nitrogen, and other species.

These comprehensive ESMs also must incorporate human activities—such as global energy systems and land use—that affect climate and in turn respond to changes in climate, freshwater availability, and sea level. Policymakers will be able to use comprehensive ESM simulations to explore alternative combinations of adaptation and mitigation strategies that will shape changes in both society and climate for the next century and beyond. Therefore, ESMs must enable investigations of future evolutions of energy and land-use systems, the consequences for the physical climate system and biogeochemical cycles, and potential human adaptations. To pursue targeted model refinements and use predictions effectively, the uncertainties associated with ESM predictions must be quantifiable and also attributable to their underlying causes.

Increases in model resolution will be dramatic, with approximately 1-km global grid spacing for the atmosphere and 1- to 5-km resolution for oceans. Achieving even higher resolutions would be useful but may be unfeasible for climate simulations in 2030. A 1-km grid spacing is sufficient to allow many processes treated very simply in today's models to be represented using a "first-principles" approach. Such resolution also is comparable to the scale at which many observational measurements are made. Surface boundary condition components (including transient current, historical, and potential future conditions) related to soils and vegetation must be improved greatly to simulate surface-atmosphere interactions fully and more effectively at local and regional scales.

The 1-km atmospheric grid—combined with more realistic parameterizations of turbulence, cloud-particle microphysics, and radiation—will significantly improve simulations of clouds and convection as well as major weather events like tropical cyclones. This resolution also will allow more accurate integration of atmospheric chemistry, land- and ocean-surface carbon and nitrogen interactions, and soil and groundwater processes into the models. Such integration subsequently will enable analysis and better prediction of the radiative forcing by short-lived species such as aerosols. Even at this resolution, breakthrough developments are needed to determine how to translate information about disparate, coupled processes and phenomena across scales—from the molecular and pore to the local, regional, or

BER Contributions to Climate Research

The Office of Biological and Environmental Research (BER), within the Department of Energy's (DOE) Office of Science, has played a major role in advancing climate change research, both through its own research programs and the provision of major facilities and resources to further the scientific community's progress. Specifically, BER has collaborated with the National Science Foundation and University Corporation for Atmospheric Research to support the development of the Community Earth System Model (formerly the Community Climate System Model). Through collaboration with DOE's Office of Advanced Scientific Computing Research, BER also has made available the country's most advanced computational facilities for climate research, including simulations for Intergovernmental Panel on Climate Change (IPCC) assessments.

Climate model intercomparisons, which are now an important part of the international climate research enterprise, began in the late 1980s under BER support, notably continuing through the Atmospheric Model Intercomparison Project. BER's Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory has provided essential data management support to IPCC and other major national and international climate modeling projects. Additionally, BER's support for integrated modeling of energy and land-use systems has provided important insights into the various ways that anthropogenic perturbations to the carbon cycle might evolve.

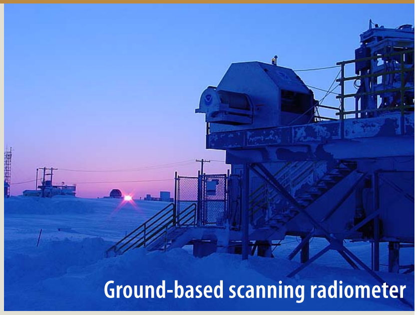


Image credit: Department of Energy's Atmospheric Radiation Measurement Climate Research Facility.

BER also has been very active in the collection of climate data. The Atmospheric Radiation Measurement Climate Research Facility has gathered invaluable field data used for testing the cloud and radiation components of climate models. BER's long-term support for fundamental research on the carbon cycle, especially its terrestrial components, and the response of plants to enhanced atmospheric CO_2 has provided important process-level knowledge about the current carbon cycle, as well as some of the responses to a changed climate future.

global scale of the prediction. The models and approaches might be mechanistic, mathematical, statistical, or phenomenological in nature. New variants of volume averaging, hybrid mixture theory, statistical and continuum mechanisms, and decomposition methods are necessary for addressing important scale transitions.

These very high resolution models will be made possible in part by exponential increases in computer power. Multiple machines must be dedicated to running ESMs, and each must be capable of delivering a sustained exaflop (10^{18} floating point operations per second) or more. Although this requirement is well recognized, its importance cannot be overstated.

ESMs are complex in terms of their conceptual basis, embodiment as computer programs, and sheer number of variables and parameters. New approaches are needed to deal with this complexity, especially for analyzing and understanding model results. As ESMs evolve to include more complex and interactive processes, a system approach, as well as improved process representations and understanding, will be needed to explore nonlinearity in the human-Earth system. Such investigations will help improve capabilities for predicting abrupt climate changes.

Despite the emphasis on increasing resolution, ESMs with lower resolutions still will play a role in climate modeling, especially for long-term simulations. Changes in ice sheets, for example, are key to understanding and predicting critically important shifts in sea level. Simulations of ice-age cycles with low-resolution ESMs will provide crucial tests of ice-sheet models and other ESM components. They also will offer powerful evidence for the ability of ESMs to predict future climate change and will serve as useful checks of many ESM components, including representations of the carbon cycle. These simulations will reveal insights into questions such as, what determines the lower and upper limits of CO_2 and methane concentrations in the cold and warm phases of an ice-age cycle? Can the time scales of past high sea-level stands and the subsequent initiation of ice ages be properly simulated? Simulations of ice-age cycles also will provide new knowledge about the nature and predictability of abrupt climate fluctuations that modulate longer cycles.

Research challenges related to other key processes requiring better integration into ESMs are outlined in the following sections.

Cloud and Aerosol Processes

3.2 Improve parameterizations and basic knowledge about aerosols that affect clouds and factors that control their number and concentration for microphysics, radiative transfer, and turbulence processes to quantify indirect aerosol forcing and resulting precipitation changes.

Higher model resolution cannot solve all problems. With 1-km grid spacing, parameterizations still will be needed for at least three types of physical processes:

- Microphysics of the creation and evolution of cloud particles and aerosols.
- Radiative transfers of the flows of solar and terrestrial (Earth-emitted infrared) radiation.
- Turbulence.

The greatest remaining uncertainty in atmospheric models in 2030 probably will be associated with microphysics. However, uncertainties in the interactions of turbulence dynamics, radiation, and microphysics are coupled in ways not included in today's global models. Moreover, the most poorly understood of these small-scale processes (e.g., formation of liquid and ice particles) are precisely the targets of many geoengineering proposals (see sidebar, Understanding the Potential Consequences of Geoengineering, p. 32).

A combination of new and traditional approaches will be needed to make progress in this area. Traditional approaches include field experiments in which the most detailed data are gathered from ground, aircraft, and balloon platforms.

Extended support is crucial to advance current instruments used for *in situ* measurements and to develop additional technologies, especially for fully characterizing ice crystal properties and emulating cloud-particle formulation or evolution (e.g., ice nucleation). New approaches can include direct numerical simulation of the evolution of cloud particles in a "control volume" and experimental manipulation of cloud systems (e.g., to study the effects of aerosols on liquid and ice water content, effective radius, particle size distribution, and precipitation). New observing systems and a means of accurately interpreting resultant observations are needed to detect such effects and transfer the information into a modeling framework. Links with the natural biological system also should be quantified (e.g., bacteria can seed cloud particles, and organic particles are produced from vegetative emissions).

Today's climate models are unable to resolve interactions among radiation, microphysics, and turbulence dynamics. New approaches also include extending the ability of these large-scale models to appropriately address their inadequacies, either through very high resolutions in limited domains (e.g., 50-m resolution) or through improved parameterizations. Such approaches will allow us to better quantify indirect forcing by aerosols and changes to precipitation driven by aerosols.

Ocean and Terrestrial Processes

3.3 Develop ecosystem-observing systems to monitor biogeochemical cycles, estimate critical process parameters, and provide model tests in ocean and terrestrial biospheres, including subsurface soils.

A 1- to 5-km grid for modeling oceans will resolve eddies and frontal and coastal processes, all of which are crucial for the exchange of properties with the atmosphere. For example, the vertical transport of properties in convectively unstable regions (such as the Nordic Seas) clearly is affected by secondary circulation of baroclinic eddies, which need to be resolved. We now know that only when sea surface temperatures in the Gulf Stream frontal zone are resolved down to tens of kilometers is their signature expressed in the tropopause, affecting the position and strength of storm tracks. Coastal regions 1 to 10 km wide are where much of oceanic upwelling and downwelling occur. Here, vertical transport of heat and nutrients has important feedbacks on atmospheric circulation (e.g., the position of the intertropical convergence zone) and economic impacts on fisheries.

Through sustained monitoring of both the oceanic uptake of heat, carbon, and other atmospheric constituents and the associated ocean acidification and sea-level rise, we can evaluate high-resolution models and use them to simulate remediation scenarios. Parameterized ocean turbulence from the microscale, including internal waves, will remain a source of uncertainty. At this scale, process models and experiments will remain the methods of choice for understanding turbulence and designing realistic parameterizations.

The terrestrial surface also responds to change in ways that promote significant positive feedbacks to climatic warming. There are model-based indications of such feedbacks, as well as biogeographically based computations and observations of CO_2 change lagging warming in the Little Ice Age and Pleistocene deglaciation events. Arguments suggest that these feedbacks may be strongly positive, neutral, and even negative. Clearly, this global feedback question needs to be better understood in the context of other major Earth systems (atmosphere and ocean).

Developing the ability to predict changes in soil moisture and groundwater on seasonal and longer time scales is imperative for modeling terrestrial ecosystems and their interactions with the rest of the climate system. Soil moisture strongly controls plant productivity in many terrestrial ecosystems and also influences runoff, stream flow, and numerous other key ecosystem services. Below the upper half meter of soil, water flow through a fractured medium and hydraulic redistribution by deep roots are two of many poorly understood hydrological processes that alter soil moisture near the surface and vertical distribution of moisture to the water table.

The subsurface—actively linked to the atmosphere and biosphere through the hydrological and carbon cycles—serves as a storage location for much of Earth's fresh water. Coupled hydrological, microbiological, and geochemical processes occurring within the subsurface environment cause the local and regional natural chemical fluxes that govern water quality. To advance our predictive capability of feedbacks to the climate system and the impacts of climate on groundwater availability and quality, long-term watershed-level observing systems are needed. These systems would include measurements of the subsurface and biosphere, perhaps focusing on the Arctic and tropics, which are terrestrial systems with large carbon stores.

As predicted years ago, climate change has been particularly rapid in the Arctic, as the insulation capacity of thick sea ice changes and the positive sea ice–albedo feedback operates. Furthermore, releases of permafrost carbon stores in a warming Arctic represent another possible strong positive feedback on climate warming (see sidebar, Potential Permafrost Loss Poses Growing Concern, p. 31). Future ecosystem-observing systems might be modeled after those developed by DOE's Atmospheric Radiation Measurement Climate Research Facility to monitor clouds and radiation.

Soil microbiology is an essential component of the Earth system. Although dynamic vegetation models have been developed, dynamic soil models still are needed. Such models should represent potential future changes in microbial populations regulating the release and consumption of trace gases to the atmosphere, while acknowledging important scaling limitations. Better representation of the processes controlling trace-gas emissions also could provide important synergies with BER efforts in microbial genomics and sequencing. Next-generation models also need to include additional subsurface resolution and processes that control a large fraction of carbon and nitrogen dynamics, water availability for plant growth, and potential hydrological feedbacks to the physical climate system. To support development and evaluation of soil and vegetation models, we need—as a first priority—to create databases on global soil characteristics, soil depth, and above- and belowground vegetation properties. Particularly significant in these calculations is the depth of soil carbon active in the global carbon budget under different conditions. Changes in the active soil depth strongly affect the amount of soil carbon, which already is the dominant storage pool of terrestrial carbon from a global perspective. Also required are detailed, long-term soil measurements in multiple landscapes across the globe.

Oceans and land contribute to atmospheric composition too. These contributions include reactive gases and particles that are significant because of their role in total radiative forcing, influences on human health, and impacts on biological productivity. The carbon feedback has not been well quantified, although some studies suggest that it may be strongly positive. Nitrogen dynamics also are important for determining primary production, particularly because of the large anthropogenic magnification of the nitrogen cycle. We need to build a predictive, process-based capacity for understanding trends in the abundances of various atmospheric chemical species and enhance the ability to simulate and monitor potential futures.

Biological Processes

3.4 Advance understanding of important biological interactions and feedbacks to identify potential tipping points and possible mitigation strategies such as carbon biosequestration.

There is broad agreement that the types of changes being projected in the physical climate system will affect not only a multitude of biophysical and biogeochemical processes, but also the future viability of both terrestrial and marine ecosystems. Disturbance processes such as fires and direct human disturbances like deforestation, reforestation, and agricultural intensification need to be incorporated into ESMs. Responses of these major ecosystem components and their sensitivities to climate variability and change need to be understood to simulate longer-term consequences.

The paired biological processes of photosynthesis and respiration dominate global carbon cycling. Most organic matter formed by photosynthesis is rapidly returned to the atmosphere as CO_2 through respiration by plants and microbial decomposers, but a small fraction is temporarily removed from active cycling by biosequestration. Current estimates are that terrestrial processes

Potential Permafrost Loss Poses Growing Concern

Image credit: iStockphoto.

Permafrost (i.e., permanently frozen subsoil) currently occupies 18.8 x 10^6 km^2, or 16%, of global soil area. More important, permafrost now sequesters an estimated 1672 Pg (1.67 × 10^{12} metric tons) of organic carbon. Concern over this soil has emerged because current models for global heating indicate an 80% to 99% reduction in permafrost area by 2100. This warming and the concomitant action of soil microbial populations could convert stored permafrost carbon into CO_2 or, under anaerobic conditions, CH_4 (methane)—both potent greenhouse gases (GHGs). Although model simulations differ, permafrost thawing would be expected to add 100 to 750 Pg (1 to 7.5 x 10^{11} metric tons) of carbon to the atmosphere. For comparison, total anthropogenic carbon releases are estimated to be 10 Pg annually. Consequently, permafrost thawing has tremendous potential for significantly increasing the levels of GHGs released into the atmosphere as a result of heightened microbial activity.

Currently, no model can accurately predict the effects of heating on permafrost. Turnover of plant residues as permafrost thaws is a complex process with numerous routes that may or may not lead to predictable releases of GHGs. Many microbial guild assemblages participating in the process often compete for carbon sources, terminal electron acceptors and donors, and micro- and macronutrients—all of which influence terminal carbon releases. Understanding the systems biology of microbial cells and microbial communities at both subzero and above-zero temperatures and under a variety of soil moisture and geochemical conditions could reveal novel physiology and regulatory controls. Such insights would enhance our understanding and prediction of changes already under way in permafrost regions and their feedbacks on global change. Indeed, opportunities may exist to intercede at any one of these levels to alter the course of carbon releases at both the molecular and organismal scale. From a design perspective, organisms or pathways that alter carbon flow from methane into biopolymers requiring gluconeogenesis exist and could be employed in large-scale mitigation strategies.

The biology and climate of permafrost regions—including the effects of both soil microbial and plant communities—represent an important and exciting opportunity to integrate the climate and environmental research missions of the Department of Energy's Office of Biological and Environmental Research.

have prevented about 30% of anthropogenic emissions from remaining in the atmosphere. First-generation climate–carbon cycle models predict that the biosphere's capacity to sequester anthropogenic carbon will peak by midcentury, and current evidence suggests that the airborne fraction of these emissions began to increase in the first decade of this century. The lack of adequate understanding about so-called "tipping points" introduces enormous uncertainty into longer-term climate projections. Examples of these tipping points include the potential for huge organic carbon releases resulting from warming of high-latitude permafrost or the increased frequency of forest disturbances, such as large-scale beetle outbreaks that become more prevalent as temperatures rise. Although critically important, many of these biological carbon cycling processes are absent or only minimally represented in first-generation climate models.

Twenty years from now, the new generation of ESMs will benefit from improved understanding of the biological processes driving the global carbon cycle and greater integration of them into biogeochemical models. The biosphere also plays a vital role in regulating the global water cycle, although the processes involved are more dynamic and take place at smaller geographic and temporal scales. A pressing need will be to develop tests of climate models so that their predictions can be used confidently to foresee future climate and ecosystem services and to prioritize mitigation strategies based on their effectiveness. These tests and current estimates rely on empirical land data that are highly uncertain. In contrast, emissions from fossil fuel use and cement production are relatively certain, large,

Understanding the Potential Consequences of Geoengineering

Clearly, humans are strongly perturbing the Earth system, and their impacts are increasing with time. These impacts are inadvertent today, but there is a possibility that humanity will choose to deliberately manipulate—or "geoengineer"—small components of the Earth's climate system. This process introduces a host of potential risks and rewards, and research is needed to explore potential consequences and outcomes to the planet. So far, two geoengineering approaches have been suggested:

- Carbon Dioxide Removal (CDR)
- Solar Radiation Management (SRM)

We must be able to predict, *with error bars*, the potentially broad consequences of geoengineering on society and the Earth system.

Some possible early steps in geoengineering provide an opportunity to improve our knowledge of fundamental aspects of the climate system critical for predicting climate change. For example, an experiment designed to perturb the brightness of a climatically negligible cloud system for brief periods over a small area could verify our understanding of cloud-aerosol interactions. These interactions are crucial components of the climate system and are responsible for major uncertainties in characterizing climate change. However, care must be taken to quantify both large- and smaller-scale changes that may take place. Comparing small but measurable perturbations to the predicted behavior in a model provides an unprecedented opportunity to confirm our understanding and conduct real hypothesis testing. This kind of experiment provides vital information about the climate system and also is a critical first step for an SRM "cloud whitening" strategy. Similar opportunities can be constructed for other components of the climate system and for different geoengineering strategies.

Image credit: Department of Energy's Atmospheric Radiation Measurement Climate Research Facility.

and rapidly increasing (Le Quéré et al. 2009, van de Werf et al. 2009, GCP 2009). Therefore, efforts need to be directed at improving not only models, but their underlying data as well.

Human Interactions

3.5 Improve integration of anthropogenic climate forcings into ESMs and develop new techniques to evaluate these coupled models at both global and regional scales.

Energy production, its transformation and use, and direct land-use change make up the major anthropogenic forcings of the current climate system that release CO_2 , other GHGs, and aerosols (including short-lived species) to the atmosphere. In addition, land-use practices (e.g., rice cultivation and fertilizer applications on crops) also are responsible for significant fractions of the non-CO_2 GHG and reactive gas budgets affecting atmospheric composition both regionally and globally. Furthermore, land-cover changes and water management both control the disposition of energy from the surface into radiative, sensible, and latent heat fluxes—potentially altering climates on local scales.

Integrated assessment models (IAMs) simulate these types of emissions as functions of economic decisions on energy, technology, and land use but do not incorporate the social, political, or cultural influences of land-use change (CBES 2009; see also Chapter 4: Grand Challenges in Energy Sustainability, p. 37). Increasingly, IAMs interact with ecosystem and atmosphere models and also incorporate responses to mitigation and adaptation measures.

Because humans play such an important role in shaping anthropogenic climate forcing through industrial, energy, and land systems, coupling IAMs to ESMs will enable a more complete representation of the Earth system. Understanding the history of land-use change and its contributions to total atmospheric forcing over the past several centuries also is a critical component of such work. Nevertheless, it is important to recognize that, despite the utility of developing comprehensive ESMs, the ability to project effects of human decisions is limited. New techniques thus will be needed to evaluate coupled ESM and IAM models at both global and regional scales.

Observing Systems

3.6 Establish new observational technologies and use them to comprehensively compare ESM predictions with observations and to quantify model errors.

To create a comprehensive analysis of the Earth system's evolving state, a very broad range of observations will need to be combined and the data assimilated with ESMs. Key goals are to:

- Monitor evolving reservoirs of carbon, heat, and freshwater and quantify and attribute errors in model results.
- Understand the diversity of forces influencing changes in land management and type, potentially through *in situ* measures of land changes. Observing spectrally resolved trends in radiation flows—at both the top of the atmosphere and the Earth's surface—also may be possible, along with relating those trends to changes in clouds, water vapor, aerosols, snow and ice, land use, and groundwater resources.
- Determine, through observations and modeling, the rates of oceanic uptake of carbon and heat and the subsequent impact on marine productivity, ocean acidification, and sea-level rise, among other things. Measurements of the distributions and flows of carbon, hydrogen, and oxygen isotopes also can be particularly useful for documenting both physical and biological processes.
- Obtain much better information on surface boundary conditions, including vegetation characteristics and the global distribution of soil properties (e.g., depth, texture, infiltration, permeability, and nutrient levels). Soils, and belowground processes in general, are poorly understood and frequently misrepresented in current models. This inadequacy and poor representations of hydrological processes in current

models result in significant model disagreement on how land-cover changes affect climate. These challenges must be resolved before models can be used effectively for regional-scale attribution of climate change.

Major investment in new observational technologies and tools is needed to enable comprehensive comparisons of ESM predictions with observations, as well as quantification of model errors. This will entail new strategies for assimilating high-resolution spatiotemporal observations of the hydrological cycle into improved atmospheric models that can resolve the coupling between hydrological processes and atmospheric motions at the measured scales. Similarly, new observational technologies are needed that can provide much more complete spatial sampling of the system under study. Models actually can guide observational strategies by revealing where measurements will be most useful. For example, measurements can be concentrated where models indicate sensitivity to poorly understood processes. Moreover, observing systems must be based on both remotely sensed and *in situ* technologies to benefit from the strengths of both and the synergy that occurs when they are used together.

Important Research Payoffs

Detailed knowledge of how the hydrological cycle, terrestrial and oceanic carbon cycles, land-use changes, and fossil fuel emissions are evolving in response to shifts in forcing will allow real hypothesis testing with ESMs. Monthly and yearly monitoring of operational Earth system forecasts will enable us to identify and attribute errors—an approach very similar to the one used today by operational weather forecasting centers. Over decades, this will lead to vastly improved climate projections.

An important application of these emerging models will be decadal projections. High-resolution ESMs will use observations of the initial climate state and estimates of evolving forcing to generate predictions of the system's state one to three decades in the future. These decadal forecasts will capture El Niños and decadal modes of climate variability and their influences on extreme weather. They also will quantify regional time-evolving climate change to inform adaptation policies. Both private and public enterprises can benefit from such predictions. Information on even longer time scales is needed for setting infrastructure goals and policy directions. Monitoring system response to policy actions will enable improved projection capabilities over time.

Twenty years from now, scientifically based observation and modeling systems will produce much more detailed analyses and predictions of both weather and climate, including additional regional-scale diagnostics such as rainfall and extreme event statistics. These analyses and projections will continue to increase in value, driven largely by better ways to translate model output into information that society needs. Equally important as improvements in the quality and quantity of this information is better access to it.

Models will be used for predicting not only climate change, but also the consequences of particular climate change policies on ecosystems. Needed developments include tools that predict energy consumption and the consequences of energy use and support energy infrastructure planning.

Climate change attribution (i.e., connecting observed changes to particular causes) is just as important as prediction. Attribution is particularly challenging on regional scales. Regional climate changes—already seen in the Arctic—will become more visible over the next few decades.

Ensuring a Competent Workforce

A skilled workforce will be needed to provide climate services, including projections of energy requirements and freshwater needs. For example, a new discipline could evolve whose practitioners have a background in climate modeling and statistical analysis, as well as training in applications and stakeholder interactions. This new element of the labor force would work at the interface of science and society, translating results from models into usable information for stakeholders. The demands on stakeholders also will change because some understanding of the relevant science will be essential. For example, a water manager who is told that there is a 40% chance of drought in the Southwest must be able to optimally use that information.

For decades, finding scientifically knowledgeable people in the general population has become increasingly difficult. The population at large has a deplorably low level of scientific literacy, and there are reasons to think this situation will worsen. Schools at all levels, from primary to postgraduate, are under tremendous financial stress, with little prospect for relief. Our society simply must use all possible means to improve this situation. Useful steps would include incentives for students at all levels to consider and ultimately pursue careers in science and engineering. Students who pursue nontechnical careers should acquire at least a reasonable understanding of what science actually is: namely, a way to learn about the world.

A further issue associated with meeting the above challenges is increasing the doctorate-level workforce in the area of climate change and ESMs. Huge increases in the complexity of these models are projected, but only minor improvements are possible if just a small number of people are actively working in this field. Today, international assessments—such as those by the Intergovernmental Panel on Climate Change (IPCC)—aid progress through the efforts of many scientists. But progress also is hampered because, for example, as the IPCC focuses attention on the biggest current challenges, it also draws time away from efforts needed to overcome them. An increased doctorate-level workforce will improve this situation significantly.

Creating "DOE Cooperative Institutes" at universities would be one approach for achieving a more highly educated workforce. Providing a mechanism for continuing investment in graduate education, these institutes would focus on issues relating to climate, energy, and the environment. For example, an institute devoted to permafrost and climate change could be created at a northern university with natural interests in this area and ready access to permafrost experimental sites. (For more on workforce development, see Chapter 6: Grand Challenges in Education and Workforce Training, p. 51.)

4 Grand Challenges in Energy Sustainability

"… . Reliable and affordable energy is essential for meeting basic human needs and fueling economic growth… [yet] the harvesting, transport, processing, and conversion of energy using the resources and technologies relied upon today cause a large share of the most difficult and damaging environmental problems society faces."

— J. P. Holdren. 2008. "Science and Technology for Sustainable Well-Being," *Science* 319(5862), 424–34.

The present organization of human society and its well-being rest on sustaining the conditions and processes related to economic, sociopolitical, and environmental systems (Holdren 2008). These processes and conditions are created, maintained, and evolved by the systems associated with them. The systems, however, do not exist in isolation; they are interlinked, and we and our technologies have major effects on them. Society would not be worried about changes in climate and, more generally, the Earth system, if human activities and the externalities of the technological choices being employed were not altering the conditions under which we live.

We choose how to structure and live our lives, adopting and incorporating responses and solutions according to our preferences and access to resources. Witness the differences between the United States and Europe regarding acceptance of nuclear energy, biotechnology, and genetic engineering, as well as the degree to which behaviors are embedded and resistant to change. An essential component of energy sustainability is fundamental knowledge of relevant natural and physical processes—from the foundational constituents of microbes to the combinations that make up ecosystems and produce the functions and services they provide to the characteristics of global atmospheric circulation. A major thrust of this understanding is preserving these processes and the conditions under which they exist for the benefit of humanity (see box, Summary of Research Recommendations, p. 38, and sidebar, What Is Sustainability?, p. 39). An equally essential component of energy sustainability is an integrated understanding of how humans perceive and influence these processes.

Research challenges discussed in other chapters—biological systems, climate, and computing—form the foundation of energy sustainability research for the Department of Energy's Office of Biological and Environmental Research (BER).

Grand Challenge Research Recommendations

Outlining the Challenges

Endeavors to achieve energy sustainability suggest the need to understand the consequences and processes necessary for mitigating the problems and externalities resulting from how the energy and industrial sectors are organized. Research must address problems caused by present activities (e.g., CO_2 emissions), develop future alternatives, and anticipate potential problems and consequences of these alternatives so that they can be minimized. Such understanding is partly technological, but mostly scientific.

Additionally, our primary energy sources will have to evolve from those relying on sunlight stored in the chemical bonds of fossil fuels to those that rely more on present-day sunlight and other renewable resources. Such energy sources

Summary of Research Recommendations

4.1 Analyze and compare potential approaches to organize land use, water use, and energy systems in ways that achieve sustainable energy, food, biodiversity, and ecosystem functioning.

- **Develop the understanding needed to produce enough energy to support more people at a higher standard of living in ways that sustain growth and minimize negative environmental impacts.**
- **Determine the means to double, over 20 years, the share of energy needs met by bioenergy in environmentally and economically sustainable ways.**
- **Develop affordable and competitive options for energy supply and conversion that minimize negative direct, indirect, and life-cycle impacts on climatic, environmental, and ecological systems.**

4.2 Identify and characterize potential Earth system drivers, feedbacks, and vulnerabilities to state changes so that their consequences and triggers might be avoided.

- **Characterize the spatial and temporal variabilities of biological systems throughout Earth and be able to understand, predict, and manipulate their rates of change.**
- **Develop technologies to remotely monitor ecosystem services in real time and at landscape scales.**

4.3 Develop unifying models and frameworks capable of testing and evaluating the significance of potential global change issues, including energy, land use, and water. The impacts these issues have on both society and the environment also must be tested and assessed.

- **Develop the scientific basis for life-cycle analysis and full-cost accounting of ecosystem services for extant and next-generation energy technologies.**
- **Advance the science to detect, understand, and mitigate the environmental and security issues associated with the potential resurgence of nuclear power.**
- **Advance the discipline of ecological informatics and incorporate social and policy elements into science-based sustainability research.**
- **Develop the science perspective necessary to set target resolutions for the spatial and temporal dimensions underlying systems of interest and be able to convert global change estimations into model visualizations.**
- **Determine the degree to which future sustainable energy facilities should be centralized.**

could include (1) direct conversion of sunlight by engineered technology; (2) wind; (3) extraction of energy stored in biota via photosynthetic and other processes; and (4) various other ocean and tidal, hydroelectric, geothermal, solar, or biological means. These energy pathways are mostly related to technology deployment, but the development of technologies and their associated processes must be based on a strong scientific foundation.

Research also needs to address not only technical and scientific issues and constraints, but also the degree to which solutions are implemented and the subsequent consequences. Technological solutions undoubtedly will face social, physical, and economic constraints, such as the present difficulties in siting wind farms (e.g., off the coast of Cape Cod) or the environmental concerns limiting the deployment of solar thermal technologies in California. Varied ethical, legal, economic, and other societal implications are inherent in the science and technology of energy systems, and the degree to which solutions are adopted affects both social and natural systems.

What Is Sustainability?

Sustainability is defined differently depending on various individuals' perspectives, but a broad definition is the capacity and ability to maintain and endure. Given the interactions within and between relevant systems, sustainability is associated with the notion that providing resources and satisfying current needs in one arena should not overly inhibit the current or future functioning of another. Energy sustainability involves supplying and maintaining adequate, reliable, and affordable energy sources in ways that do not compromise the future ability to provide useful energy and energy services or the vitality of the associated systems.

Image credit: U.S. Department of Agriculture Natural Resources Conservation Service.

Also important is quantifying the extent to which different entities link systems, perhaps through the resources that these entities provide or the degree to which the resources are impacted. Land use and water resources and quality are prime examples of such linkages. Understanding not only the role of these linkages, but also how science can relax the constraints imposed by them is necessary.

Sustainability, by its very nature, calls for multidisciplinary research that spans spatial and temporal scales. Compartmentalized research, development, and associated agendas can inform understanding of system components, but sustainability is an area of study that must address the overlap and interrelationships of systems. In particular, biologists and ecologists must engage with social scientists to understand the forces and implications of changes in energy sustainability. Although we cannot predict with certainty the makeup of tomorrow's society, the future Earth and climate systems will be different from today's, and there is great concern that our desired future may elude us if we continue on our current trajectory. The choices we make are determined partly by values, and different kinds of futures call for different kinds of information, tools, and responses. Many issues and challenges fall under the realm of sustainability, as do inextricable linkages with economic and social processes. Although this chapter primarily focuses on the pervasive challenges and opportunities facing the biophysical sciences, there are equally pervasive challenges and opportunities facing the social sciences. Moreover, many of the most interesting and important energy sustainability questions are at the interface of natural and human systems, requiring integrated research across these two domains. Previously, BER has not included human roles in much of its work, but these are important determinants of and vital contributors to energy sustainability. Future cross-agency partnerships may be a valuable consideration for addressing important social questions in coupled natural human energy systems.

Energy Sustainability Research Recommendations

Humanity clearly faces enormous challenges with respect to energy production, transmission, conversion, use, impacts, and sustainability. Many of these challenges and their associated research agendas can be categorized into the following three overarching recommendations.

4.1 Analyze and compare potential approaches to organize land use, water use, and energy systems in ways that achieve sustainable energy, food, biodiversity, and ecosystem functioning.

Which milestones have to be reached, and what must be accomplished to meet them in a reasonable time frame? Several related grand challenges are to:

- Develop the understanding needed to produce enough energy to support more people at a higher standard of living in ways that sustain growth and minimize negative environmental impacts. By 2100, the world's population—which will seek a higher standard of living—is projected to increase by 40% over today's level. If society is bordering on or already has crossed into an unsustainable regime, how do we provide a standard of living that sustainably supports a much larger population? One path forward would be to triple the amount of degraded land on which perennial energy crops are planted in a manner that increases soil carbon storage and water quality (Tilman et al. 2009).
- Determine the means to double, over 20 years, the share of energy needs met by bioenergy in environmentally and economically sustainable ways. Such a pathway requires developing transformational bioenergy alternatives that deliver more energy, goods, and services than are available currently. Today, C4 plants most efficiently fix CO_2, but can we develop sustainable and productive plants that more efficiently use sunlight, water, and other nutrients; remediate environmental toxins; provide other ecosystem services; and are ideally suited for bioenergy applications?
- Develop affordable and competitive options for energy supply and conversion that minimize negative direct, indirect, and life-cycle impacts on climatic, environmental, and ecological systems. Developing means, methods, and technologies that provide useful energy with minimal greenhouse gas (GHG) emissions and other impacts requires a substantial shift away from systems that emit CO_2. It also involves understanding how new solutions can impact and alter different Earth system components.

Nascent and probably future technologies to extract useful hydrocarbons from resources that are increasingly lower quality (e.g., oil from tar sands) or still virtually unexploited (e.g., methane from hydrates) require methods that use fewer resources for harnessing present and future stocks. Future energy and resource (e.g., water) intensities of harvesting primary energy sources should be equal to or less than the present intensity of conventional activities producing these resources (e.g., exploration and production of oil and natural gas by drilling into conventionally located reservoirs). Science should strive to understand the consequences of exploring and exploiting these resources (e.g., mineral leaching, nutrient flow, and seafloor destabilization).

Current approaches to mitigate CO_2 and other GHG emissions have non–emission related processes and impacts that also must be understood and either enhanced or mitigated. For example, understanding and enhancing the chemical, geochemical, and biogeochemical methods and processes for CO_2 separation and geologic storage can provide cheap and secure means to mitigate CO_2 emissions from existing and future facilities. Siting renewable energy facilities such as solar- and wind-generated electricity farms involves building new roads and establishing new facilities, the footprints of which impact ecosystems. At larger scales, concentrated solar power and biomass used for bioenergy can require significant water withdrawals, thereby affecting water quality. At even larger scales, wind turbines can impact climate and atmospheric patterns by increasing the roughness of Earth's surface and extracting energy from blowing wind (Keith et al. 2004).

4.2 Identify and characterize potential Earth system drivers, feedbacks, and vulnerabilities to state changes so that their consequences and triggers might be avoided.

We lack sufficient understanding of Earth, environmental, and ecological systems and their vulnerability and resilience in a changing climate. Some associated grand challenges are to:

- Characterize the spatial and temporal variabilities of biological systems throughout Earth and be able to understand, predict, and manipulate their rates of change. For example, what are the roles of subsurface hydrological and biogeochemical processes in climate change and GHG exchange? Similarly, we also must understand current microbial activities in permafrost (see sidebar, Potential Permafrost Loss Poses Growing Concern, p. 31). How and under what conditions might these permafrost activities change? How can these activities be mitigated? What is the potential for capturing and using the products of this activation (e.g., CH_4)? What are the consequences for the climate, Earth system, and physical infrastructure of permafrost-loss areas? More broadly, we must understand how to identify, quantify, and model the mechanisms of present carbon sinks in plants and soils to increase their effectiveness at removing CO_2 from the atmosphere.

- Develop technologies to remotely monitor ecosystem services in real time and at landscape scales. We need to understand how rapidly species can migrate (e.g., by using dynamic vegetation models) and how individual and species-level responses translate into ecosystem-level change. Also needed is the ability to characterize and assess differential ecosystem sensitivity and the extent to which ecosystem vulnerability is determined by *in situ* responses or by potential species migration.

4.3 Develop unifying models and frameworks capable of testing and evaluating the significance of potential global change issues, including energy, land use, and water. The impacts these issues have on both society and the environment also must be tested and assessed.

Scientific developments related to energy have a long history of being hailed as the next solution to societal needs and problems. In the 1960s, nuclear energy from fission was touted as the next great technology to provide cheap and accessible electricity, and nuclear fusion always has been touted as just a decade or two away. More recently, the hydrogen economy was popular, and now biofuels, plug-in electric vehicles, and geoengineering are solutions being emphasized. We need to develop the means to put such potential solutions into context, where uncertainty is naturally incorporated to ultimately achieve a fully integrated and overarching assessment model. Some related grand challenges are to:

- Develop the scientific basis for life-cycle analysis and full-cost accounting of ecosystem services for extant and next-generation energy technologies.
- Advance the science to detect, understand, and mitigate the environmental and security issues associated with the potential resurgence of nuclear power.
- Advance the discipline of ecological informatics and incorporate social and policy elements into science-based sustainability research. Determining the role of biodiversity and human activity for ecosystem functioning is one direction this research could pursue.
- Develop the science perspective necessary to set target resolutions for the spatial and temporal dimensions underlying systems of interest (e.g., 1 km and 1 year, respectively, for climate models) and be able to convert global change estimations into model visualizations. At a smaller scale of interest, a significant challenge will be to develop (1) integrated approaches to model and understand biogeochemical systems *in situ*; (2) unified and transferable models of biogeochemical processes; and (3) the means and methods for multimodal microscopy, spectroscopy, and visualization and imaging of microbiological and environmental chemicals. Accurately modeling and analyzing global climate change will require developing and building more and better computing facilities. Such facilities are essential, given the computational complexity inherent in all systems of interest, their different spatial and temporal characteristics, the necessity of modeling a system at a level larger than the unit of interest, and the interactions of systems within and across their characteristic spatial and temporal scales.

- Determine the degree to which future sustainable energy facilities should be centralized. Such an understanding integrates the characteristics of a technology and its inherent returns to a scale engendered by the cost functions of its installation, operation, and disposal; by environmental impacts throughout its life cycle; and by the land, water, and ecosystem resources necessary for and impacted by it. On one end of the spectrum, distributed electricity generation from renewable sources can bring those sources closer to the point of use and minimize their impacts on local ecosystems. However, such decentralization can cause coordination issues and impede certain activities. CO_2 capture and storage, for example, cannot be applied cost effectively to diffuse and distributed CO_2 sources.

5 Grand Challenges in Computing for Biological and Environmental Research

Data Representation, Integration, and Knowledgebase Development to Support BER's Long-Term Vision

> ***"Since at least Newton's laws of motion in the 17th century, scientists have recognized experimental and theoretical science as the basic research paradigms for understanding nature. In recent decades, computer simulations have become an essential third paradigm. . . . As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform*** **data-intensive science.*"***
>
> **— G. Bell, T. Hey, and A. Szalay. 2009. "Beyond the Data Deluge,"** ***Science*** **323 (5919), 1297–98.**

As experimental systems become increasingly digital, more accurate, and efficient, the resultant datasets continue to grow exponentially and encompass an expanding diversity of data types. Examples relevant to the Department of Energy's (DOE) Office of Biological and Environmental Research (BER) include measurements of greenhouse gas fluxes and subsurface contaminant concentrations, as well as climate simulation output. In biology, this enormous growth in information is especially acute as DNA sequencers transition to next-generation technologies. These new sequencers are expected to produce data volumes equivalent to a human genome (3 gigabases) for $15 by 2015. By comparison, the first human genome was completed in 2003 in a 13-year, $3 billion effort by the Human Genome Project. The exponential growth in genomic sequences over the last two decades will become super-exponential growth over the next 20 years. Furthermore, as the cost of obtaining these data dramatically declines, sequencing technologies will become more accurate and cost-effective replacements for microarrays in transcriptomic analyses. The increase in scientific data has been faster than anticipated over recent decades because of improvements in electronics, computing, and experimental systems. Growth over the next 20 years is expected to be significantly greater, given the goals of improved models and anticipated further advancements in experimental techniques.

In climate science, the largest volume of data has come from the many environmental satellites in operation since 1979. Perhaps the more important climate science data relevant to DOE's mission, however, are the data available from BER's Atmospheric Radiation Measurement (ARM) Climate Research Facility archive. This program started collecting data more than 20 years ago (see sidebar, BER Contributions to Climate Research, p. 27), and the volume has now grown to 200 terabytes, with approximately 2 to 3 more terabytes added monthly. Data volume will increase by 4 to 7 times this amount in the near future. The key to using these data in climate model verification is an understanding of the location and time resolution of the data sample collection and how and whether these data can be adapted for comparison with climate model data.

A theme shared across all BER science is that as the research community develops common "models" of scientific understanding, the approaches and new knowledge then are captured in computational systems. These systems reflect with increasing accuracy detailed scientific understanding and, ultimately, the reality of natural systems

Summary of Research Recommendations

5.1 Establish a new data management paradigm for data-intensive science with ontologies as a basis for semantic data representations; standards for experimental protocols and data exchange; and an open-access, open-development data management infrastructure.

5.2 Create a new publishing paradigm that credits and rewards researchers for publishing peer-reviewed datasets or analytical methods in addition to conventional peer-reviewed journal articles.

5.3 Develop new computing paradigms capable of meeting the enormous parallel processing and data-intensive analysis needs now emerging for biological, climate, and environmental data.

5.4 Standardize experimental and computational protocols and methods to increase data integration, data usability, and system interoperability to improve research productivity.

5.5 Improve data usability and model accuracy by ensuring that appropriate data quality standards are created and stored with the accompanying data.

5.6 Design and build software solutions that provide researchers with better access to increasingly large, complex, and interrelated datasets.

5.7 Develop virtual laboratories and tools to more fully engage human cognitive faculties and provide richer opportunities for scientific collaborations.

being studied and the experimental systems probing them. Such computational models also would be designed to simulate a natural system or its components and the associated experimental systems to provide predictions that could then be further validated. Currently, these models are elements such as analysis programs and climate models, database schema, and annotations and/or metadata. Substantial increases in data—along with possible limits in the growth of computer processing speed—will demand a more focused approach to direct modeling of scientific concepts into computational systems (see box, Summary of Research Recommendations, this page). Similarly, experimental protocols and the reported data should accurately describe current scientific understanding and experimental results and provide the supporting ontological and semantic information to enable machine-readable support for subsequent searching, analysis, and reuse. Ultimately, this general concept of data modeling will be captured entirely in computational systems that then will provide an automated basis for data storage, integration, query, and retrieval. These systems will enable computational simulations and projections that allow further assessment and understanding, resulting in new directions for future experiments and model improvements.

This approach is becoming critical because the increasing scale and complexity of data require computing for consistently and coherently managing and integrating large sets of data and information. Such a perspective drives our view of what computational or informational science is and the objectives that must be achieved over the next 20 years.

Grand Challenge Research Recommendations

5.1 Establish a new data management paradigm for data-intensive science with ontologies as a basis for semantic data representations; standards for experimental protocols and data exchange; and an open-access, open-development data management infrastructure.

As data generation becomes exponentially more cost effective, the number of scientists will not increase proportionately. Current methods researchers use to analyze entire datasets (such as with Excel) will break down when applied to very large and complex datasets, and uniform manual curation of these data will be impractical. New paradigms thus are required to address these challenges. Improved computational methods and resources are becoming increasingly important to continued scientific progress. Since the volume and diversity of data will grow substantially beyond that which was typical in the past, new computational requirements will include capabilities for managing and processing very large datasets.

Although growth in computer disk storage space will be cost effective for the foreseeable future, computer processor speed and network bandwidth may not continue to grow at past rates (see recommendation 5.3, p. 47). Therefore, a likely computing paradigm in the future will be data-intensive science (Bell et al. 2009) that entails (1) developing ontologies as a basis for semantic representation of biological, climatological, and environmental systems science concepts; (2) establishing standards for experimental protocols and data exchange; and (3) building a data management infrastructure that is open access and open development in all aspects.

The ARM Climate Research Facility represents one such "open-access" model for data. Indeed, the development of this program's "value-added products," which combine the best observational data to yield new structures or products with high value for atmospheric science, is a data-provision model that also yields high value to scientists comparing the data with computer model output.

Increasingly, data will be preprocessed and compressed to make them more manageable. For example, an experimental system that produces image files, which typically are large, may preprocess them to extract important data and then discard the original image files. This effort saves disk space, but an improved image-analysis algorithm could render old data obsolete if the images are not still available to rerun. Tradeoffs depend on the costs of storage and re-running the experiment and on the probability that a new algorithm will be found.

Beyond basic file storage, data also can be organized in a relational database system, which allows structured queries to retrieve information quickly. New "query" paradigms are based on novel ways to organize data from XML and its derivatives into a software framework (e.g., MapReduce/Hadoop) that supports applications for processing large amounts of data in parallel. Potentially, such a system could enable the testing of new, value-added products that might become part of a data archive in the future. Although not suitable for all problems, these approaches are beginning to be developed and seem to hold promise for dealing with large-scale data processing in parallel implementations such as "cloud" computing.

"Google Maps" is a familiar example of an application that integrates data from different sources and scales and also processes queries quickly. We envision maps of cellular and subsurface systems or atmospheric data displayed similarly (see sidebar, "Google" Life, p. 46). These maps also could have browsable annotation. Because manual annotation is becoming increasingly prohibitive due to data growth, one idea is that future annotations would be generated automatically. However, this approach is recognized as not ideal. Informed users are encouraged to contribute manual annotation as they encounter opportunities to make improvements (such as browsing a Google Map of pathways). Such contributions would be vetted by a peer-review system and could be considered a type of mini-publication.

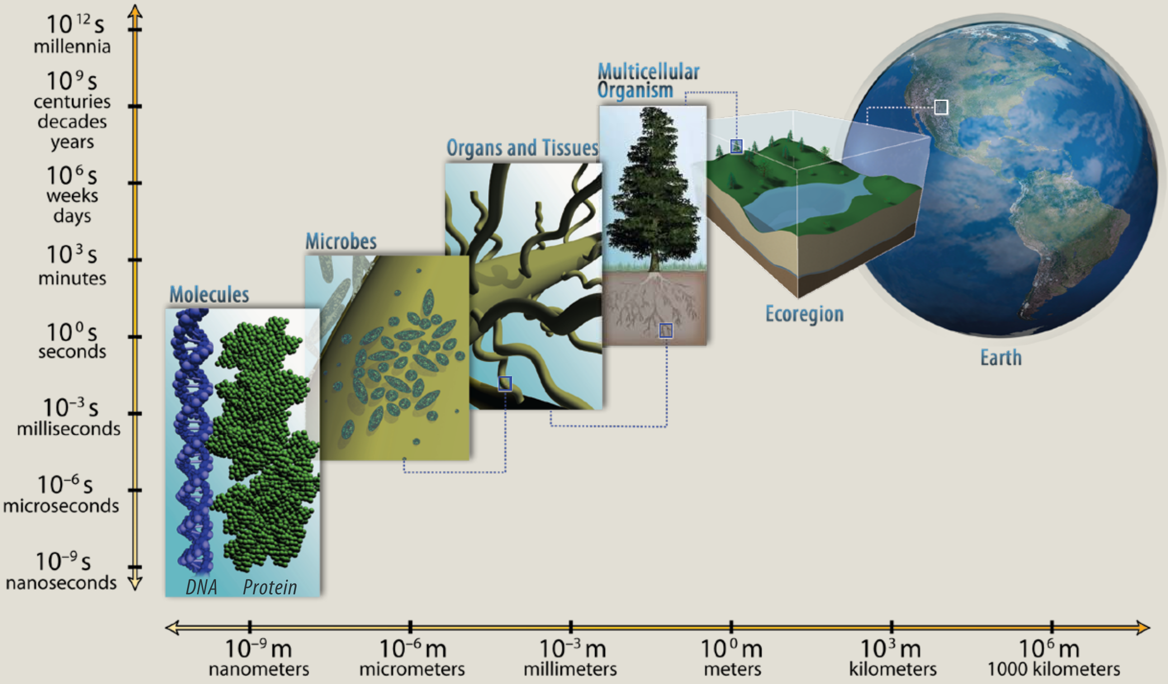
In 20 years, most researchers analyzing data often may not know how the data were generated or results derived. This problem is analogous today to commonly used systems like BLAST or the use of complex but accepted laboratory techniques. To conduct science, such analysis systems must be well documented in a generally accessible way so that interested researchers can learn how the system works and verify its integrity to their satisfaction. The computational community has the responsibility of making tools accessible, both in terms of providing detailed documentation of the system use, methods, uncertainty, and code, as well as guiding correct usage. Also critical is that such systems be assessed and accredited by authoritative external reviewers conducting an unbiased peer review. Without such characteristics, these systems do not meet the rigors of the scientific method.

"Google" Life: Computing Biological and Climatological Interactions Across Scales

Imagine a computational resource allowing us to move seamlessly across biological or atmospheric scales. From an initial view of molecules inside a cell nucleus, we would continue outward to see the cellular components, then the whole cell, the cell in the context of a tissue interacting with adjacent cells, the tissue as part of an organism (e.g., a tree in a forest, the forest as part of a geographical region, and then as part of a continent), and finally to a view of the whole Earth. (A related example might be the rhizosphere associated with a tundra biome that has carbon and climate impacts.) For atmospheric science, we might envision seeing the data taken within a column above an Atmospheric Radiation Measurement Climate Research Facility site placed in context within the broader image of the region surrounding the site and its evolution over time.

Computational modeling at multiple scales is required for atmospheric systems and understanding biological systems within a bacterium or in the context of environmental and ecological interactions that ultimately impact the carbon cycle and climate. Whatever the scale of study, each discipline has existing models and will continue to refine and improve them as new experimental data become available and new analysis methods are developed.

Models could interact at each scale, but today, there is no continuity from models of microbial systems to those that predict global climate. Researchers in each discipline are focused on improving current scientific understanding in their domain and generally view today's models as crude. As models become refined and increasingly precise at each scale, the question of how a model can interact meaningfully with those at adjacent scales will present opportunities to gain further insight.



An important scientific tenet is that publication of research implies that enough information is provided to allow the work to be replicated. As science has moved forward into a more data intensive, analytical world, we have lost sight of this tenet. Data provenance no longer is being captured sufficiently for results to be validated. This issue must be addressed and become a standard operating principle (SOP). Provenance includes, for example, information about who conducted the experiment and when, how it was done, and the software version used. In many environmental datasets, a "quality flag" is often added to the data to indicate that it has been quality-controlled in a way that meets the standard for that dataset.

In 20 years, all researchers will have easy access to a community knowledgebase that will include a repository of standard experimental data, an archive of analysis tools, and an ontology library. Each researcher will have a login that provides a private computational workspace that can be shared with collaborators. They also will be able to contribute to any of the knowledgebase elements in a manner similar to that of a peer-reviewed journal and ultimately publish components of their private work in this public space.

5.2 Create a new publishing paradigm that credits and rewards researchers for publishing peer-reviewed datasets or analytical methods in addition to conventional peer-reviewed journal articles.

Today, most experimental results are published in peer-reviewed journals. Looking forward, science could advance much more quickly if results were saved online in a machine-readable format connected with associated metadata, as well as ontologies and data values with their statistical properties. The experimental procedure also would be documented with an SOP. This approach would enable analyses that integrate and compare multiple independent experiments and that could be conducted in a community knowledgebase if the experimental data deposited therein were compatible with a particular standard.

Although "publishing" such data, analytical methods, or code in, for example, the DOE Systems Biology Knowledgebase (see genomicscience.energy.gov/compbio/) or elsewhere beyond traditional journals would provide value to the research community, such contributions currently are neither commonly required by sponsors or publishers nor recognized as accomplishments for career advancement. Indeed, data within the ARM Climate Research Facility archive generally do not come with specific authorship after a specific experimental procedure has been accepted. Other examples are manual curation, or correction of automated annotation, and ontology development. Given the continuing need for developing electronic, machine-readable forms of experimental results and conclusions, new paradigms for peer-review accreditation must be an ongoing topic for consideration.

5.3 Develop new computing paradigms capable of meeting the enormous parallel processing and data-intensive analysis needs now emerging for biological, climate, and environmental data.

Moore's Law generally has kept pace with the growth in both biological sequence data and environmental data. However, the increase in processor clock speed appears to be subsiding even as data volume is growing ever faster. Consequently, investigating alternative analysis methods is becoming critical. One new approach is parallelism based on cloud computing methods. Although traditional high-performance computing (HPC) works well on problems such as climate modeling and molecular dynamics, most data-intensive problems in biology are expected to parallelize most effectively in a cloud computing architecture. Such efforts are only just beginning and will require analysis and recoding methods.

Analysis systems necessarily become more complex as scientific models become more accurate or encompass more components. Additional layers of technical complexity usually come at the cost of adding more computing parameters. The inverse also is true. In general, a reduction in data dimensionality will make computing more efficient and can reduce noise. However, along with this theoretical simplification, the model may become less accurate. As problems reach computability limits, this approach will allow us to use tools like mathematics to make simplifications that require fewer computational resources but usually result in less accurate models. An example would be using least squares to derive an equation that approximates measured data. The equation is a simplification that eliminates data noise, but the data might be represented more accurately by a more complex error model. The model, however, would require much greater computational resources.

"Google Translate" is an example of another kind of simplification. As an automatic translator, this program uses statistical learning techniques to build a translation model based on data analysis. Typical translation systems have been rule based and require predefined vocabularies and grammars. Google Translate instead has analyzed billions of words of text and aligned examples of human translations. When results of Google Translate are incorrect, a knowledgeable user can submit a better translation that then is added to the database. This is a type of manual curation.

Methods analyzing statistical co-occurrences of words are readily applicable to climate, environmental systems science, and biological and physical systems. Although such "machine learning" methods are not 100% accurate, they can be remarkably sufficient, if properly understood, and provide automated methods that may be necessary in the near future.

5.4 Standardize experimental and computational protocols and methods to increase data integration, data usability, and system interoperability to improve research productivity.

As the research community gains increasing access to new experimental technologies and large datasets, standardized experimental and computational protocols and methods will be required to enable comparative analyses that provide a true basis for scientific assessment and further investigation. Standards development can be a delicate issue. Although standards can be onerous and require substantial additional work to support a data submission, integrating and comparing experimental results become almost impossible without them. Initial focus for standards development should be on the largest and most significant experimental results having the greatest impact on the research community. This triage will continue for the foreseeable future, but standardization needs to become as serious a professional activity in biology, climate, and environmental systems science as it is in the telecommunications industry because research productivity will suffer without it.

5.5 Improve data usability and model accuracy by ensuring that appropriate data quality standards are created and stored with the accompanying data.

Data usability and model accuracy depend on data quality. The quality of data can be assessed by analyzing their statistical properties and comparing the data with independent validation. In general, most data are collected without careful consideration of these properties. The research community needs to establish and maintain a culture that recognizes the need for and value of quality assessments to properly conduct experiments and analyses. Without such an approach, results are of questionable value and undermine scientific credibility. Much of the data in databases today are of varying quality, yet user displays present the data as if their quality is uniform and presumably good. Because all downstream analyses depend on the quality of input data, including into the analyses quality measures that are appropriately propagated is important. Quality is related to the reproducibility of results, an important science objective. A related factor is provenance, the historical record of the sources of data and analytical methods. As computational models are built, their accuracy also depends on input quality. Differences between experimental measures and model predictions determine the next experiment. Poor input quality leads to erroneous conclusions and potential experimental waste. Additionally, experimental consistency among laboratories must be assessed. Without progress on these issues, our ability to integrate multiscalar data of potentially varying quality will become increasingly compromised.

5.6 Design and build software solutions that provide researchers with better access to increasingly large, complex, and interrelated datasets.

The ultimate limit on scientific advances is not computers, but scientists. Software needs to be designed so that it does not interfere with what researchers are trying to do, but rather enhances their research experience and productivity. Consequently, the research community needs to be involved in designing and developing computational systems such as the DOE Systems Biology Knowledgebase. With the huge scaling up of experimental data, the future bottleneck will be computational data management and analysis. No researcher should be limited by lack of access to computational and analytical capabilities.

Many software systems used for analysis today were developed by individual scientists who created these tools to advance their own research but are not rewarded for making them easier to use. A shift therefore is needed toward an open-community environment that encourages code to be easier to use and provides further incentives based on usage and value to the research community. This development environment would provide rapid feedback and support for high-quality user interfaces for a range of tools—from free-form data analytics to researcher-driven, semantic-based discovery. In this context, ease of use includes enabling researchers to discover and learn more about how a software system or tool works.

Future research activities will be conducted in an environment in which researchers who analyze the data possibly will not understand how the data were generated or what assumptions were made, all of which could lead to incorrect conclusions. Computational tools will need to be developed and hardened to allow noncomputational scientists to gain insights from large datasets. Ultimately, we might have to expand computational infrastructure—computers, networks, and methods—to gain more scientific productivity.

5.7 Develop virtual laboratories and tools to more fully engage human cognitive faculties and provide richer opportunities for scientific collaborations.

What do scientists do when they are not physically conducting either a computational or observational experiment? Mostly, they click, type, read, talk, and think—all acts of nearly pure symbol manipulation. If mapped into virtual environments, human symbol manipulation can be given a much wider diversity of physical expression and more fully engage human cognitive faculties. For example, scientific objects and processes could be represented by virtual macroscopic objects and visible computational processes. Such virtual environments will be essential for future scientists, both to retain their interest and to greatly amplify cognition with computation and collaboration.

What could be represented in a virtual scientific environment? Almost anything if it is important enough to justify the initial software development effort. Genes, proteins, reactions, biomolecular networks, kinetic models, and diverse datasets from both atmospheric and biological systems are well on their way already. Observations, hypotheses, informal models, scientific notebooks, instruments, and entire replicable experiments represented as workflows could follow. For example, particularly for biological systems, many hypotheses could be represented using a diagrammatic, process-modeling language that can be manipulated like a circuit diagram drawn in a three-dimensional or hyperbolic space, permitting connections to other such process diagrams. Large-scale laboratory and computer experiments in progress could be represented using a virtual factory floor. Collaborative environments supporting remote interaction also may be among the drivers for virtualization. These developments may require applying current and near-future artificial intelligence technologies, including machine learning and limited natural language understanding. Such technologies increasingly are applied in the nonscientific video games and virtual environments now being used by our future scientists. Labor cost savings from the successful transplantation and replication of experimental protocols alone could pay for the required software development for virtual scientific environments.

6 Grand Challenges in Education and Workforce Training

"The key . . . to improving our health and well-being, to harnessing clean energy, to protecting our security, and succeeding in the global economy will be reaffirming and strengthening America's role as the world's engine of scientific discovery and technological innovation. And that leadership tomorrow depends on how we educate our students today, especially in those fields that hold the promise of producing future innovations and innovators. And that's why education in math and science is so important."

— President Barack Obama, November 23, 2009, on the "Educate to Innovate" campaign.

The grand challenges articulated in this document reveal sweeping changes in the language, methods, and conceptual frameworks that dominate research supported by the Department of Energy's (DOE) Office of Biological and Environmental Research (BER), shifting from a focus on the parts to a quest for the whole. As noted in Chapter 2, tackling complexity really is THE grand challenge in biology. However, biological education has not switched to this new track. Biologists continue to teach discrete components—parts of the cell, individual molecules, separate organ systems, and striated taxa—rather than presenting a unified image of systems with emergent properties. For example, only recently have BER and the subsurface research community come to a consensus regarding the urgent need for new approaches to adequately detect, quantify, and interpret complex behavior in the subsurface. In this area, hydrological and biogeochemical processes interact within a heterogeneous framework over a hierarchy of scales ranging from subnanometer to kilometer, and one component process alone can exhibit complex behavior (U.S. DOE 2010). The imperative need is to educate our future workforce to think about properties of whole systems, which rarely can be explained simply as an accumulation of parts.

This new research framework necessitates reorienting education from teaching the known to exploring the unknown, an approach advocated vehemently by scholars and science educators to make systems-based biology an agent for the long-hailed overhaul of science education. Education research dictates that the teaching of biology should capture the nature of inquiry, presenting the science as a method of investigation rather than an extant body of knowledge. Inquiry-driven teaching impels instructors to use problem-based learning, open-ended laboratory exercises, and collaborative discussion—all techniques shown to foster higher-order thinking skills as well as retention of information (see box, Summary of Recommendations, p. 52).

BER has a long-standing commitment to science education and workforce development. This is demonstrated principally through the support of graduate students pursuing advanced degrees; postdoctoral students who work on research projects; and, to a lesser extent, hands-on research opportunities for undergraduate students and K–12 educators, as well as informal experiential learning opportunities for K–12 students. Many of these individuals use DOE research facilities and work alongside the scientific and technical staff at the national laboratories.

Summary of Recommendations

6.1 Engage science educators.

6.2 Develop a centralized education mission.

6.3 Initiate interdisciplinary fellowships.

6.4 Enhance career development programs.

6.5 Support engineering education to address crucial shortages in engineering disciplines relevant to BER's grand research challenges.

6.6 Install education experts at the national laboratories.

6.7 Develop collaborative teaching programs.

One example of BER education and workforce development efforts is the Global Change Education Program (GCEP), which promotes undergraduate and graduate education and training supportive of BER's global change research activities. GCEP has two components: the Summer Undergraduate Research Experience (SURE) and the Graduate Research Environmental Fellowships (GREF). SURE's primary goal is to involve undergraduate students at the end of their sophomore or junior year in BER-supported global change research, continuing this undergraduate experience during subsequent summer studies. To further improve the quality of emerging scientists in disciplines related to global change research, SURE students are encouraged to apply for GREF and graduate education opportunities. GREF's primary goal is to support research designed and conducted collaboratively among graduate students, their faculty advisors, and DOE researchers at national laboratories or universities. GREF graduates are encouraged to continue climate change careers in postdoctoral and permanent positions in academia, government laboratories, and industry.

In addition, several of the research programs within BER offer their own science education portfolio. Examples are provided below.

- The DOE Environmental Molecular Sciences Laboratory (EMSL) offers two fellowship programs. The Wiley Visiting Scientist program is designed to enable distinguished scientists to spend extended periods of time at EMSL focusing on their research and helping to plan or develop EMSL capabilities. The William Wiley Post-Doctoral Fellowship gives highly qualified Ph.D. scientists the opportunity to conduct creative original research using EMSL capabilities.
- BER's Genomic Science program provides resources for educators to enhance existing curricula (genomicscience.energy.gov). These include primers on biofuels and microbial genomics, an image gallery, and Genomic Science program reports.
- Each of the three DOE Bioenergy Research Centers has active outreach and educational programs. For example, the DOE BioEnergy Science Center has created a website devoted to K–12 education (bioenergycenter.org/students-and-kids/). It provides teachers and students with information, games, and links to other websites that teach them how to make a difference in their own lives with energy conservation. Likewise, the DOE Great Lakes Bioenergy Research Center (GLBRC) has a significant education effort (www.glbrc.org/education/). The mission of GLBRC Education and Outreach is to inform a variety of audiences about bioenergy research, energy concerns, and sustainability issues affecting Earth. Its goal is to broaden the understanding of current issues in bioenergy for the general public and for students and educators at the K–12, undergraduate, and graduate levels. Finally, the DOE Joint BioEnergy Institute (JBEI) supports the training of high school teachers and students who can then transfer this training and knowledge back to their own classrooms.

The above examples are effective educational efforts that seek to engage and excite K–12 students and teachers and support and encourage graduate and postgraduate research scientists. Yet we must consider whether these disparate programs reflect the very best education pedagogy for training a workforce in the complex systems biology approaches required to meet the grand challenges proposed in this report. Even cursory examination of these grand challenges reveals that mathematics is a central theme of the science and must be, as biology has become, an information science.

Recognizing the widespread concern about science, technology, engineering, and mathematics (STEM) workforce development, the Secretary of Energy in 2006 commissioned a review by the Secretary of Energy Advisory Board. That review demonstrated a clear role for DOE in STEM education and concluded that partnerships should be the primary vehicle for achieving its goals.

Grand Challenge Education and Training Recommendations

BER clearly is poised to play a crucial role in enhancing STEM education. To begin, BER must develop a clear, unifying vision for education and workforce development, emphasizing the integration of mathematics and biology and tackling complexity through the use of multidisciplinary teams. Workshop participants made recommendations on ways BER might achieve this goal, and several are summarized below.

6.1 Engage science educators.

BER must work with professional science educators to inform and direct the development of its educational vision and resulting programs. Embracing the most compelling pedagogy available will require rigorous, routine evaluation to ensure programs meet BER's changing workforce needs.

6.2 Develop a centralized education mission.

With numerous distinct educational efforts scattered among its various centers and programs, BER should leverage this diversity of resources and talent to create a coordinated educational approach. Emphasizing the informational aspect of biology and relying on team-based approaches, each program can then develop its own component in this overarching educational mission. All efforts should highlight the relevance of mathematics.

6.3 Initiate interdisciplinary fellowships.

BER should create a program of undergraduate and graduate interdisciplinary training that requires students to work with teams of scientists. One model could be an approach similar to that developed through the National Science Foundation's Integrative Graduate Education and Research Traineeship program (IGERT), in which the host institutions create faculty teams to address complex research questions. An alternative model might focus on encouraging scientists from national laboratories, academia, and industry to forge new collaborations supported by graduate traineeships that involve students moving among the team's institutions.

6.4 Enhance career development programs.

Continuing the existing early-career fellowship program, BER also should implement a program targeting postdoctoral scholars who wish to move between fields such as biology and mathematics.

6.5 Support engineering education to address crucial shortages in engineering disciplines relevant to BER's grand research challenges.

To address crucial shortages in engineering disciplines that directly impact DOE- and BER-relevant missions, BER should enhance and support engineering education using approaches similar to those described above.

6.6 Install education experts at the national laboratories.

Education experts at the national laboratories could engage the research scientists more directly and increase their awareness of educational opportunities. These individuals could coordinate educational activities between BER programs and the laboratories and carry out the necessary evaluations of education programs.

Energy and the Next Generation: Grand Challenge in Education

Imagine every school in the nation involved in energy creation and conservation. In a manner similar to the U.S. Environmental Protection Agency's Energy Challenge, students would take a pledge to measure their school's baseline energy consumption and then identify and implement practices to reduce consumption by 10%. Different school grades would focus on different aspects of the energy challenge, ranging from identifying renewable and nonrenewable energy sources in their schools (elementary students) to measuring baseline energy consumption (middle school students) to creating and implementing alternative energy sources (high school students). Schools would create multidisciplinary teams to confront and help solve their school's different energy challenges, with input from classes including mathematics, physics, biology, and geology.

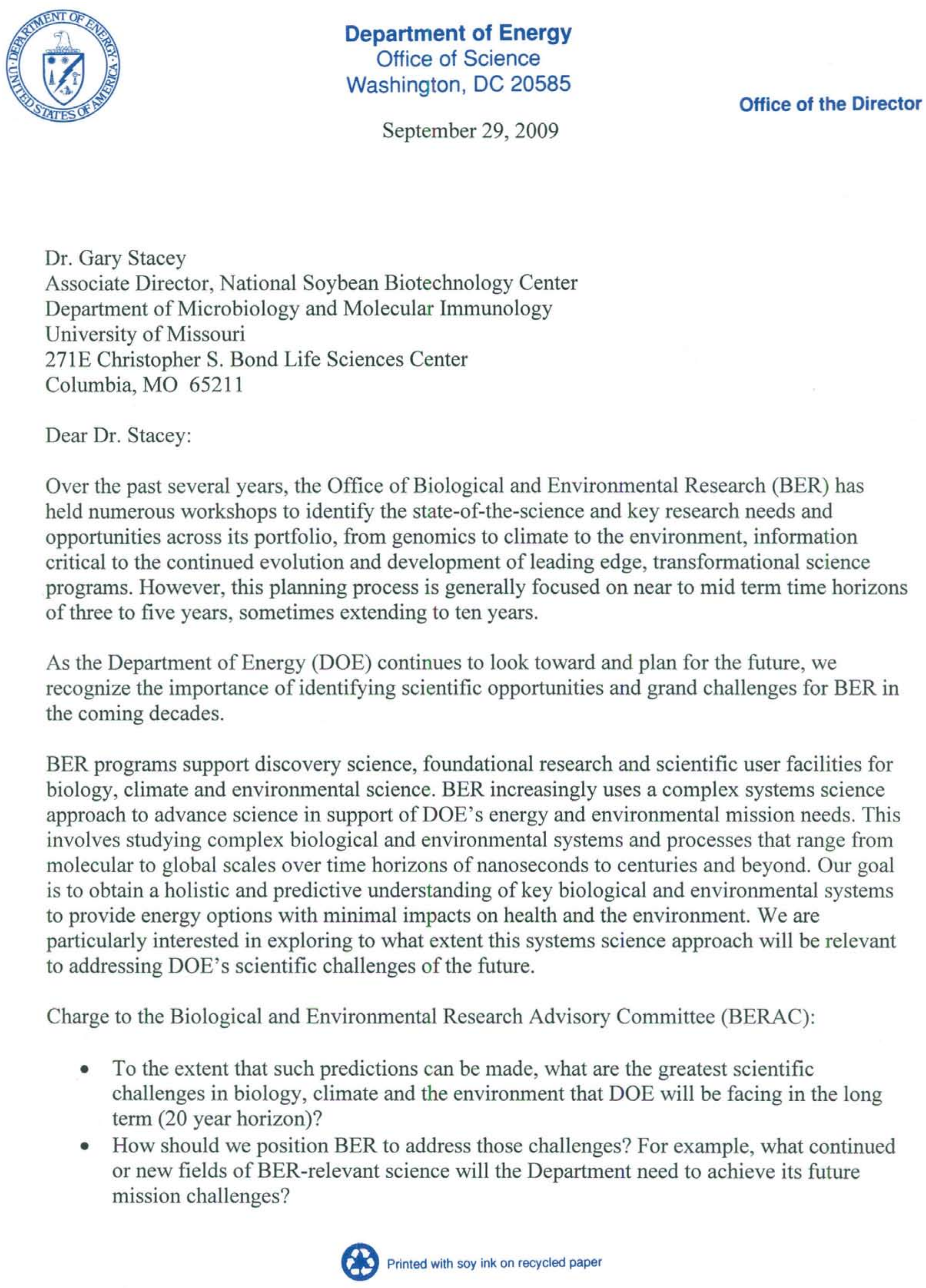
Image credit: Greater New Bedford (Massachusetts) Regional Vocational Technical High School.

BER would engage appropriate governmental agencies to develop K–12 curricula focused on energy conservation and alternative energy production. This educational program would have a direct and significant impact on STEM (science, technology, engineering, and mathematics) education and workforce development across the nation by (1) engaging students in activities that take science out of the classroom and into their daily lives, (2) encouraging multidisciplinary team-based investigation of problems and solutions, and (3) emphasizing the statistical and mathematical methods required to discuss energy consumption.

6.7 Develop collaborative teaching programs.

Interactions are needed among university scientists, K–12 teachers, national laboratory scientists, and students. An example of a Grand Challenge that would result in such interactions is provided in the sidebar, Energy and the Next Generation, this page.

Appendix 1: Charge Letter



Department of Energy
Office of Science
Washington, DC 20585

Office of the Director

September 29, 2009

Dr. Gary Stacey
Associate Director, National Soybean Biotechnology Center
Department of Microbiology and Molecular Immunology
University of Missouri
271E Christopher S. Bond Life Sciences Center
Columbia, MO 65211

Dear Dr. Stacey:

Over the past several years, the Office of Biological and Environmental Research (BER) has held numerous workshops to identify the state-of-the-science and key research needs and opportunities across its portfolio, from genomics to climate to the environment, information critical to the continued evolution and development of leading edge, transformational science programs. However, this planning process is generally focused on near to mid term time horizons of three to five years, sometimes extending to ten years.

As the Department of Energy (DOE) continues to look toward and plan for the future, we recognize the importance of identifying scientific opportunities and grand challenges for BER in the coming decades.

BER programs support discovery science, foundational research and scientific user facilities for biology, climate and environmental science. BER increasingly uses a complex systems science approach to advance science in support of DOE's energy and environmental mission needs. This involves studying complex biological and environmental systems and processes that range from molecular to global scales over time horizons of nanoseconds to centuries and beyond. Our goal is to obtain a holistic and predictive understanding of key biological and environmental systems to provide energy options with minimal impacts on health and the environment. We are particularly interested in exploring to what extent this systems science approach will be relevant to addressing DOE's scientific challenges of the future.

Charge to the Biological and Environmental Research Advisory Committee (BERAC):

- To the extent that such predictions can be made, what are the greatest scientific challenges in biology, climate and the environment that DOE will be facing in the long term (20 year horizon)?
- How should we position BER to address those challenges? For example, what continued or new fields of BER-relevant science will the Department need to achieve its future mission challenges?

- What new and innovative tools should be developed to advance BER science? For example, what new tools might allow the integration of data from different fields to advance systems science?
- What scientific and technical advances are needed to train the workforce of the future in integrative science, including complex system science?

With these questions in mind and others that may occur to you, we request that BERAC and the subcommittee that you will establish develop an overall strategy for drafting a long-term vision for BER. We expect that this effort may begin with an initial, overarching workshop that can develop a framework for this process. We also expect that this long-term vision will serve as a catalyst for follow up workshops organized by BER Program Managers engaging the scientific community in further developing and refining ideas for the future of BER. Many thanks for your contributions to this important effort.

Sincerely,

W. F. Brinkman
Director

cc: Anna Palmisano
David Thomassen
Sharlene Weatherwax

Appendix 2: Grand Challenges Workshop Agenda

Gaithersburg, Maryland
March 2–5, 2010

March 2, 2010

Evening arrivals

March 3, 2010

8:30 a.m.	Opening of meeting: Description of meeting goals and expected outcomes: Anna Palmisano and Gary Stacey
9:00 a.m.	**Chris Field**, Climate and Life
9:30 a.m.	**Richard Murray**, Synthetic Biology
10:00 a.m.	Break
10:30 a.m.	**Virginia Dale,** Sustainability
11:00 a.m.	**David Hill**, Systems Biology
11:30 a.m. – 12:30 p.m.	Lunch on site
12:30 p.m. – 3:00 p.m.	Breakout groups I
	1. Climate Change, chaired by David Randall
	2. Systems Biology, chaired by Gary Stacey
	3. Information and Synthetic Biology Systems Integration Framework, chaired by Gary Sayler
	4. Research Framework for Energy Sustainability, chaired by Peg Riley
3:00 p.m. – 3:30 p.m.	Refreshment break
3:30 p.m. – 4:30 p.m.	Breakout groups I
	1. Climate Change
	2. Systems Biology
	3. Information and Synthetic Biology Systems Integration Framework
	4. Research Framework for Energy Sustainability
4:30 p.m. – 5:30 p.m.	Reports of breakouts
5:30 p.m.	Adjournment

March 4, 2010

8:00 a.m.	Remarks by Undersecretary Steve Koonin
8:30 a.m.	**Leroy Hood**, The Future of Systems Biology
9:00 a.m.	**Isaac Held**, Climate Dynamics
9:30 a.m.	**Doug Landis,** Sustainability

10:00 a.m. – 12:30 p.m.	Breakout groups II 1. Understanding Systems across Temporal (Milliseconds to Millennia) and Spatial Scales (Microns to Ecosystems), chaired by Greg Petsko 2. Meeting the Workforce and Education Needs to Address the Grand Challenges, chaired by Jo Handelsman 3. Data Integration and Knowledgebase Development, chaired by Bob Cottingham 4. Novel Tools, Techniques and Probes, chaired by Joe Ecker
12:30 p.m. – 1:00 p.m.	Lunch
1:30 p.m. – 3:00 p.m.	Reports and discussion
3:30 p.m.	Adjournment

March 4 evening and March 5 (until 5:00 p.m.)

Writing committee (selected members) stays over to complete draft of workshop report.

Appendix 3: Grand Challenges Workshop Participants

Definitions of affiliation acronyms follow list of workshop participants.

Arkin, Adam (LBNL)
Bader, David (ORNL)
Baliga, Nitin (Institute for Systems Biology)
Bielicki, Jeffrey (ORNL)
Birdsey, Richard (USDA Forest Service)
Braam, Janet (Rice University)
Buford, Marilyn (USDA Forest Service)
Cessi, Paola (Scripps Institute of Oceanography)
Church, George (Harvard University)
Collins, William (LBNL)
Cottingham, Robert (ORNL)
Dale, Virginia (ORNL)
Denning, Scott (Colorado State University)
Dickinson, Robert (University of Texas, Austin)
Doney, Scott (Woods Hole Oceanographic Institution)
Easterling, David (NOAA)
Ecker, Joseph (The Salk Institute for Biological Studies)
Edmonds, James (PNNL)
Ehleringer, James (University of Utah)
Feddema, Johannes (University of Kansas)
Field, Christopher (Carnegie Institution of Washington)
Fowler, Joanna (BNL)
Freilich, Michael (NASA)
Fridlind, Ann (NASA GISS)
Fung, Inez (University of California, Berkeley)
Gilna, Paul (University of California, San Diego)
Goulden, Michael (University of California, Irvine)
Greenberg, E. Peter (University of Washington)
Greenberg, Jean (University of Chicago)
Handelsman, Jo (Yale University)
Hanson, Paul (ORNL)
Hecht, Alan (EPA)
Held, Isaac (NOAA GFDL)
Herzog, Howard (Massachusetts Institute of Technology)
Hill, David (Harvard Medical School)
Hood, Leroy (Institute for Systems Biology)
Hubbard, Susan (LBNL)
Hurtt, George (University of New Hampshire)
Jackson, Robert (Duke University)
Janetos, Anthony (Joint Global Change Research Institute)
Joachimiak, Andrzej (ANL)
Landis, Doug (Michigan State University)
Large, William (NCAR)
Leung, Ruby (PNNL)
Liao, James (University of California, Los Angeles)
Loeffler, Frank (Georgia Institute of Technology)
Long, Stephen (University of Illinois, Urbana-Champaign)
Looger, Loren (Howard Hughes Medical Institute)
Lucas, Robert (University of Southern California)
Mace, Jay (University of Utah)
Marvin, Jonathan (Janelia Farm Research Center)
Masiello, Carrie (Rice University)
Maslov, Sergei (BNL)
Meehl, Jerry (NCAR)
Meyerowitz, Elliot (California Institute of Technology)
Michalak, Anna (University of Michigan)
Murray, Richard (California Institute of Technology)
Ort, Donald (University of Illinois, Urbana-Champaign)
Otto-Bliesner, Bette (NCAR)
Padgette, Stephen (Monsanto Company)
Pakrasi, Himadri (Washington University in St. Louis)
Palsson, Bernard (University of California, San Diego)
Patrinos, Ari (Synthetic Genomics Inc.)
Penner, Joyce (University of Michigan)
Petsko, Gregory (Brandeis University)
Prather, Michael (University of California, Irvine)
Prince, Roger (ExxonMobil Biomedical Sciences, Inc.)
Randall, David (Colorado State University)
Remington, Karin (NIH)

Riley, Margaret (University of Massachusetts, Amherst)

Rogers, Alistair (BNL)

Roth, Fritz (Harvard University)

Sayler, Gary (University of Tennessee, Knoxville)

Schneider, Edwin (Center for Ocean-Land-Atmosphere Studies)

Simpson, Michael (ORNL)

Smith, Melinda (Yale University)

Stacey, Gary (University of Missouri)

Sussman, Michael (University of Wisconsin, Madison)

Thornton, Peter (ORNL)

Wall, Judy (University of Missouri)

Washington, Warren (NCAR)

Weinberger, Leor (University of California, San Diego)

Wilbanks, Thomas (ORNL)

Wildung, Raymond (PNNL)

Wiscombe, Warren (NASA GSFC)

Xu, Dong (University of Missouri)

Zachara, John (PNNL)

Zhang, Minghua (State University of New York at Stony Brook)

Affiliation Acronyms

ANL	Argonne National Laboratory
BNL	Brookhaven National Laboratory
EPA	Environmental Protection Agency
LBNL	Lawrence Berkeley National Laboratory
NASA	National Aeronautics and Space Administration
NASA GISS	NASA Goddard Institute for Space Studies
NASA GSFC	NASA Goddard Space Flight Center
NCAR	National Center for Atmospheric Research
NIH	National Institutes of Health
NOAA	National Oceanic and Atmospheric Administration
NOAA GFDL	NOAA Geophysical Fluid Dynamics Laboratory
NREL	National Renewable Energy Laboratory
ORNL	Oak Ridge National Laboratory
PNNL	Pacific Northwest National Laboratory
USDA	U.S. Department of Agriculture

Appendix 4: Bibliography

Aderem, A. 2005. "Systems Biology: Its Practice and Challenges," *Cell* **121,** 511–13.

Allen, M. S., et al. 2004. "Analysis and Glycosyl Composition of the Exopolysaccharide Isolated from the Floc-Forming Wastewater Bacterium *Thauera sp.* MZ1T," *Environmental Microbiology* **6**, 780–90.

Alon, U. 2007. "Network Motifs: Theory and Experimental Approaches," *Nature Reviews Genetics* **8**, 450–61.

Atsumi, S., et al. 2009. "Engineering the Isobutanol Biosynthetic Pathway in *Escherichia coli* by Comparison of Three Aldehyde Reductase/Alcohol Dehydrogenase Genes," *Applied Microbiology and Biotechnology* **85**(3), 651–57.

Baker, B. J., et al. 2006. "Lineages of Acidophilic Archaea Revealed by Community Genomic Analysis," *Science* **314**, 1933–35.

Bell, G., T. Hey, and A. Szalay. 2009. "Beyond the Data Deluge," *Science* **323**, 1297–98.

Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. "Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of *Escherichia coli*," *Proceedings of the National Academy of Sciences USA* **105**, 7899–906.

Brynildsen, M. P., and J. C. Liao. 2009. "An Integrated Network Approach Identifies the Isobutanol Response Network of *Escherichia coli*," *Molecular Systems Biology* **5**, 1–13.

Canadell, J. G., et al. 2007. "Contributions to Accelerating Atmospheric CO_2 Growth from Economic Activity, Carbon Intensity, and Efficiency of Natural Sinks," *Proceedings of the National Academy of Sciences USA* **104**(47), 18866–870.

Center for BioEnergy Sustainability (CBES), Oak Ridge National Laboratory. 2009. *Land-Use Change and Bioenergy: Report from the 2009 Workshop*, ORNL/CBES-001, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, and Oak Ridge National Laboratory, Center for Bioenergy Sustainability (www.ornl.gov/sci/besd/cbes.shtml).

Chan, Y. F., et al. 2010. "Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer," *Science* **327**, 302–305.

Comolli, L. R., et al. 2009. "Three-Dimensional Analysis of the Structure and Ecology of a Novel, Ultra-Small Archaeon," *The ISME Journal* **3**, 159–67.

Comte, L., et al. 2010. "Assessment of Stream Biological Responses under Multiple-Stress Conditions," *Environmental Science and Pollution Research* **17**(8), 1469–78**.**

Dicke, M., and M. W. Sabelis. 1988. "Infochemical Terminology: Based on Cost-Benefit Analysis Rather than Origin of Compounds?" *Functional Ecology* **2**(2), 131–39.

Gao, F., and E. H. Davidson. 2008. "Transfer of a Large Gene Regulatory Apparatus to a New Developmental Address in Echinoid Evolution," *Proceedings of the National Academy of Sciences USA* **105**, 6091–96.

Global Carbon Project (GCP). 2009. *Carbon Budget 2008* released November 17, 2009 (www.globalcarbonproject.org).

Grant, P. R., and B. R. Grant. 2006. "Evolution of Character Displacement in Darwin's Finches," *Science* **313**(5784), 224–26.

Hartvigsen, G., A. Kinzig, and G. Peterson. 1998. "Use and Analysis of Complex Adaptive Systems in Ecosystem Science: Overview of the Special Section," *Ecosystems* **1**, 427–30.

Hasty, J., D. McMillen, and J. J. Collins. 2002. "Engineered Gene Circuits," *Nature* **420**, 224–30.

Holdren, J. P. 2008. "Science and Technology for Sustainable Well-Being," *Science* **319**(5862), 424–34.

IPCC 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Eds. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Cambridge University Press, Cambridge and New York.

Jönsson, H., et al. 2006. "An Auxin-Driven Polarized Transport Model for Phyllotaxis," *Proceedings of the National Academy of Sciences USA* **103**, 1633–38.

Keith, D. W., et al. 2004. "The Influence of Large-Scale Wind Power on Global Climate," *Proceedings of the National Academy of Sciences USA* **101**, 16115–120.

Le Quéré, C., et al. 2009. "Trends in the Sources and Sinks of Carbon Dioxide," *Nature Geoscience* **2**, 831–36.

Levin, S. A. 2002. "Complex Adaptive Systems: Exploring the Known, the Unknown, and the Unknowable," *Bulletin of the American Mathematical Society* **40**, 3–19.

Lüthi, D., et al. 2008. "High-Resolution Carbon Dioxide Concentration Record 650,000–800,000 Years before Present," *Nature* **453**, 379–82.

Ram, R. J., et al. 2005. "Community Proteomics of a Natural Microbial Biofilm," *Science* **308**, 1915–20.

Ripp, S., et al. 2000. "Controlled Field Release of a Bioluminescent Genetically Engineered Microorganism for Bioremediation Process Monitoring and Control," *Environmental Science and Technology* **34**(5), 846–53.

Steen, E. J., et al. 2010. "Microbial Production of Fatty-Acid-Derived Fuels and Chemicals from Plant Biomass," *Nature* **463**, 559–63.

Tans, P. 2008. *Trends in Atmospheric Carbon Dioxide: Mauna Loa*. NOAA Earth System Research Laboratory (www.esrl.noaa.gov/gmd/ccgg/trends/).

Tilman, D., et al. 2009. "Beneficial Biofuels—The Food, Energy, and Environment Trilemma," *Science* **325**, 270–71.

Tyson, G. W., et al. 2004. "Community Structure and Metabolism Through Reconstruction of Microbial Genomes from the Environment," *Nature* **428**, 37–43.

U.S. DOE. 2010. *Complex Systems Science for Subsurface Fate and Transport: Report from the August 2009 Workshop*, DOE/SC-0123, U.S. Department of Energy Office of Science (www.science.doe.gov/ober/BER_workshops.html).

van der Werf, G. R., et al. 2009. "CO_2 Emissions from Forest Loss," *Nature Geoscience* **2**, 829.

Williams, K. H., et al. 2009. "Geophysical Monitoring of Coupled Microbial and Geochemical Processes During Stimulated Subsurface Bioremediation," *Environmental Science and Technology* **43**(17), 6717–23.

Zhang, Y., et al. 2009. "Responses of Soil Bacteria to Long-Term and Short-Term Cadmium Stress as Revealed by Microbial Community Analysis," *Bulletin of Environmental Contamination and Toxicology* **82**(3), 367–72.

DISCLAIMER

Printed with soy ink on recycled paper