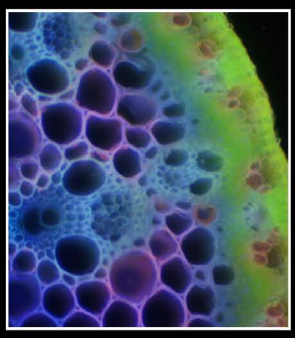
DOE/SC-0121

U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE

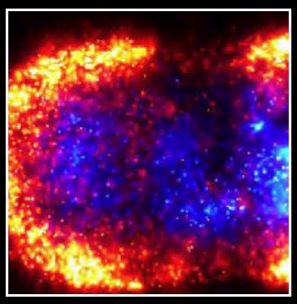
New Frontiers in Characterizing Biological Systems

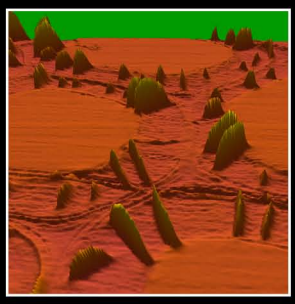
Report from the May 2009 Workshop

Organisms in Ecosystems

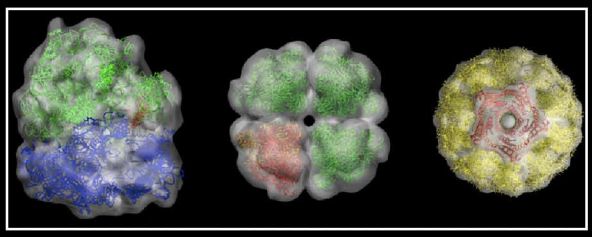


Microbial Communities and Plant Tissues





Cellular Systems



Molecular and Subcellular Systems

U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Biological and Environmental Research

New Frontiers in Characterizing Biological Systems

Report from the May 2009 Workshop

Convened by

U.S. Department of Energy
Office of Science
Office of Biological and Environmental Research

Organizers

Dean Cole
dean.cole@science.doe.gov
301.903.3268

Arthur Katz
arthur.katz@science.doe.gov
301.903.4932

---

www.science.doe.gov/Program_Offices/BER.htm

As part of the U.S. Department of Energy's (DOE) Office of Science, the Office of Biological and Environmental Research (BER) supports fundamental research and technology development aimed at achieving predictive, systems-level understanding of complex biological and environmental systems to advance DOE missions in energy, climate, and environment.

To promote development of a new generation of characterization technologies, BER hosted the workshop, New Frontiers in Characterizing Biological Systems, in May 2009. Experts from scientific disciplines relevant to DOE missions and from the enabling technologies (e.g., optical spectroscopy, genomic sequencing technology, electrochemistry, electron microscopy, and mass spectrometry) met to determine the opportunities and requirements for identifying and developing new tools and analytical approaches for characterizing cellular- and multicellular-level functions and processes that are essential to develop solutions for DOE missions in biofuels, carbon cycling and biosequestration, low dose radiation, and environmental stewardship. The intent of the workshop was to broadly explore future technology capabilities that are needed, not current technologies and their development. Discussion of specific technologies was intended to be illustrative of some promising approaches, but no attempt was made to be exhaustive.

To define such capabilities, participants were organized into working groups based on areas of biological scale: cellular, multicellular, and interface processes. This report outlines the workshop's findings and highlights key opportunities for establishing a framework for research activities.

Suggested citation for this report: U.S. DOE. 2009. *New Frontiers in Characterizing Biological Systems: Report from the May 2009 Workshop*, DOE/SC-0121, U.S. Department of Energy Office of Science (genomicscience.energy.gov/characterization/).

DOE/SC-0121

New Frontiers in Characterizing Biological Systems

Report from the May 2009 Workshop

Workshop Dates: May 13–14, 2009

Publication Date: October 2009

U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Biological and Environmental Research

This document is available at the DOE Genomic Science website (genomicscience.energy.gov/characterization/).

Contents

Executive Summary v

1. Introduction 1

2. Biological Challenges 5
   - 2.1 Understanding the Cell and Its Response to Chemical and Physical Perturbations 6
   - 2.2 Understanding Interactions Between Cells: From One to Many 8
   - 2.3 Understanding Dynamic Biological Systems Across Multiple Scales of Time and Distance 13
   - Summary of Key Biological Challenges 17

3. Technological and Capability Needs 19
   - 3.1 Expanding Global Characterization Capabilities 19
   - 3.2 Identifying and Measuring Rare Events, Molecules, and Cells Within Complex, Heterogeneous Biological Systems 23
   - 3.3 Seeing It All: Simultaneous Measurements Across Multiple Dimensions 26
   - 3.4 Integrating Information 32
   - Summary of Key Technological and Capability Challenges 38

4. Path Forward 39
   - 4.1 Technology: Meeting the Measurement Challenge 40
   - 4.2 Research Strategy: Linking Observations from Cells to Natural Communities 43
   - 4.3 Knowledge Visualization: Overcoming Challenges of Data Integration and Systems Analysis 44
   - 4.4 Integration: Teaming at Disciplinary Interfaces To Advance Frontiers in Biological Research 45

Appendix 1: Agenda for New Frontiers in Characterizing Biological Systems Workshop 47

Appendix 2: Workshop Writing Team and Participants 49

Appendix 3: Bibliography 51

Appendix 4: Figure Credits 55

Acronym List, Reference Table of Numeric Prefixes ..... Inside back cover

List of Sidebars

Chapter 2. Biological Challenges

Real-Time Chemical and Structural Analyses Are Needed To Improve Enzymatic Degradation of Lignocellulosic Biomass 9

Microbial Environmental Remediation of Chlorinated Solvents 10

Uncovering Key Interdependent Relationships Among Organisms To Leverage Beneficial Capabilities 11

Working Toward Real-Time Measurements of Biological Responses: The Low Dose Radiation Challenge 13

Targeting the Interfacial Region Between Microbial Cells and the Geosphere 14–15

Chapter 3. Technological and Capability Needs

Whole-Genome Amplification Holds Promise for Single-Cell Genome Drafting 21

The Value of Understanding Rare Events and Cells 24

Investigating Rare Microbes' Function and Phylogeny at the Single-Cell Level Using EL-FISH NanoSIMS 25

Single-Molecule Measurements Reveal Phenotype Switching in Genetically Identical Cells 27

Super Resolution Light Microscopies Enable the Imaging, Counting, and Localization of Single Proteins in Microbes 30

Challenges for Imaging Live Plant Cells 31

Integration of Proteomics and Genomics Reveals a Core Proteome 34

Executive Summary

Understanding the relationship between the genome and functional processes is the most significant challenge and potentially enabling advancement that faces modern biology. Elucidating this connection presents opportunities for realizing sustainable energy solutions and responsible management of natural resources. Understanding the function of the genome is at the core of the Department of Energy's (DOE) Genomic Science program and is central to realizing DOE's mission goals in bioenergy research, carbon management, and environmental stewardship. Just as genomic science is central to these mission goals, technology advancements are central to genomic science and to unlocking the connections between the genome and functional processes occurring at cellular to global environmental scales. New developments in characterization technologies will be essential for driving advances in genomic science and in our understanding of the genomic bases of natural processes.

In May 2009, DOE's Office of Biological and Environmental Research (BER) held the New Frontiers in Characterizing Biological Systems workshop to address the next generation of challenges in genomic science and its connection to functional systems. The workshop included a diverse array of scientists and engineers with expertise in the mission-relevant biological and environmental sciences and in the analytical and physical sciences. Working groups were focused on defining the challenges associated with studies at the cellular, multicellular, and interfacial levels. Common themes and priorities emerged from the different groups. There was universal agreement that appropriate advances in characterization technologies will first depend on articulation of the major challenges that face the biological and environmental science communities. To that end, this report—rather than comprehensively discussing currently available technologies—highlights the major challenges and outlines the future technological capabilities required to meet them. Workshop participants identified numerous knowledge gaps that inhibit the understanding of biological systems, and these can be distilled into three major challenges:

- **Understand the Cell and Its Response to Chemical and Physical Perturbations.** The characterization of genome sequences and their products has highlighted the need for identifying and characterizing the other parts that comprise the cell. Many of these components are difficult to identify or quantify. Completing the "parts list" of the cell and determining how cellular networks composed of these parts respond to local physical and chemical changes are priorities.
- **Understand the Interactions Between Cells.** Many cellular interactions are poorly understood, such as how cells communicate, regulate their genetic information in response to other cells, and combine their capabilities for higher-order functions. A needed advance involves routinely interpreting how multiple cells, with similar or different genetic content, combine to process information, energy, and materials.
- **Understand the Functioning of Biological Systems Across Multiple Scales of Time and Distance.** A connection between the genome and biological function at physical scales as diverse as an individual cell and an ecosystem or at temporal scales as diverse as seconds and years can be apparent but difficult to define. The dynamic nature, spatial and temporal ordering, and nonlinearity of system responses confound interpretations at any level of inquiry. The ability to design experiments, identify and model appropriate system components, and predict function are challenges that must be addressed.

These primary knowledge gaps are relevant to understanding the processing of biomass into different chemical forms, the cycling of carbon, and the transformation of contaminants in the environment. They are fundamental challenges intrinsic to diverse biological and environmental concerns. Timely resolution of these problems will revolutionize our understanding of biological systems and significantly advance DOE mission science. Achieving these goals will depend on a transformation of current measurement capabilities. Numerous technological approaches can be considered.

Regardless of approach, the specific technical capabilities needed to fill these knowledge gaps include:

- **Expand and Integrate Global Characterization Capabilities**. Biological systems are composed of a wide array of differing molecular species. We need to "see it all" and be able to monitor dynamic changes at increased spatial resolution. Currently, we cannot probe many of the dynamic processes occurring within the cell at the required chemical, spatial, or temporal resolution nor can we measure the response of these processes to chemical and physical perturbations. A wide assortment of metabolites, lipids, carbohydrates, and other biochemicals simply cannot be identified or accurately measured. Understanding the interactions and fates of these materials is essential for understanding cell function. Combining global measurements and extending the ability to comprehensively characterize and manipulate any system component are needed advances. New technologies for completing the parts list of cellular components are essential.
- **Identify and Measure Important Molecular Species, Events, and Cells**. Biological systems are recognized as containing complex mixtures of different chemical and biological species. The relative importance of particular components to functional outcomes is difficult to assess. Even more difficult is associating rare events or minority components to functional outcomes. Current technologies have the ability to monitor single cells and detect single molecules. However, they are limited in their ability to do so in complex, heterogeneous environments, let alone in natural systems. Technologies are needed that can identify and detect single or small populations of molecules or cells amidst complex, heterogeneous backgrounds. These technologies will aid in understanding the effects of chemical and physical forces on the cell and the interactions among cells.
- **Simultaneously Measure Many Chemical and Biological Species Across Broad Spatial and Temporal Ranges.** Biological information is carried in a wide variety of molecules and is expressed across broad spatial and temporal ranges. Current tools often are appropriate for providing characterizations at only specific spatial and temporal scales, or they are limited in the number or type of species that can be measured. There is a strong need to bridge the discontinuities between different measurements and to address the gaps that occur as we span length and time scales. Multiple "dimensions" need to be added to biological measurements so that molecular events can be linked to cellular, multicellular, and environmental scales.
- **Integrate and Interpret Diverse Information and Technology Platforms.** Biological systems can be assessed at many levels ranging from the molecular to the cellular to the ecosystem scale. Beyond the challenges of *how* to measure and collect such biological information, critical challenges related to *what* to measure and *how* to interpret the information remain. Addressing these problems will depend on effective tools for integrating and interpreting the information. Useful databases and computational approaches are needed for integrating measurement information and for modeling systems at multiple scales. Currently, we do not know at which scale to measure or model biological system function. Effective focusing of measurement technologies will rely on an iterative relationship with computational modeling approaches. Models that are capable of dealing with the gradients and discontinuities of biological systems must be developed and integrated with experimental design.

Overcoming these technical challenges will facilitate basic understanding of biological processes, not just at a particular physical or temporal scale, but the linking and relating of such scales to genomic information. Focused advancements in characterization technologies will address critical knowledge gaps and support the realization of mission needs. These advancements will broadly impact the biological and environmental sciences in general and ultimately transform biology into a quantitative science.

Moving forward in these needed developments will require concerted efforts on several fronts. Key among these is investment in stimulating technology developments. These developments will need to proceed within the context of the primary biological challenges identified in this report. A priority should be the development of approaches for simultaneously

assessing multiple species at appropriate spatial and temporal resolution. This likely will proceed through combinations of different measurement techniques. Technology advancements also must commence with developing high-throughput parallel approaches for making sensitive measurements in heterogeneous environments. Small molecules within single cells and small populations must be tracked. Measurement techniques alone will not be sufficient. Rather, what is required to reveal function is the ability to manipulate relevant biological, chemical, and physical variables while tracking their effects on biological systems.

A second key focus should be on promoting analysis of the biological systems most relevant to DOE missions. We must move past the study of relatively simple model organisms and toward the study of organisms within their natural environmental setting (e.g., *in planta* and *in terra*). Initially, organism systems that are representative of the technological and environmental problems we wish to understand must be identified. Capabilities for culturing or studying organisms at the single-cell level need to improve along with the tools for manipulating and studying these organisms at the molecular level. Systems research will need to progress past the study of individual organisms in isolation and toward systems of increasing biological complexity, replete with the structuring and heterogeneity found in natural systems. Interrogating natural systems *in situ* should be a long-term goal.

A third focus should be on integrating biological and technological developments through computational tools. Large, disparate datasets must be combined and analyzed to yield new insights into the function of biological systems across diverse scales. Iterative cycles of experimentation and modeling in concert with new theory will be needed to define the appropriate scales for measuring, modeling, and functionally understanding biological processes. This integration will need to capitalize on DOE BER traditions and success in integrating scientific disciplines and in solving grand challenge problems. Multidisciplinary teaming should be promoted and facilitated through integrative training opportunities, incentives for collaborative science, and facilitated access to high-end technologies.

Understanding the connections that link the genome to events at different scales promises to unravel many of the challenges facing the biological and environmental sciences. Such insight will enable effective routes to sustainable energy solutions and responsible stewardship of the environment. Analytical technology developments are key to sustaining progress toward these goals and addressing the challenges and knowledge gaps that emerge. The complexity, emergent properties, and multiple scales of biological systems present substantial obstacles. The tremendous progress in characterizing whole genomes, which only a few decades ago was considered a nearly intractable problem, was enabled by the focused integration of a biological problem with technological advances in analytical measurements and computation. Similarly, the seemingly daunting challenges that we now face can be addressed through focused developments and bold advances in characterization technologies.

1. Introduction

In the environment, plants and microorganisms perform key functions that sustain the biosphere. These functions include capturing and converting sunlight to chemical energy, cycling carbon and nutrients, and detoxifying and immobilizing natural and man-made pollutants. Living organisms collectively possess an astonishing array of capabilities that, if characterized and understood, could be leveraged to revolutionize approaches for solving many of the critical problems currently facing the nation and planet. For example, photosynthetic organisms use energy from sunlight to fix carbon dioxide and produce a myriad of carbon-based molecules needed for cell growth and other metabolic functions. Some organisms produce enzymes that break down complex biopolymers such as lignocellulose into individual component molecules that other organisms, in turn, can employ. Understanding the basis of these complex biological processes for capturing carbon dioxide, immobilizing environmental contaminants, and developing biofuels is of keen interest to the Department of Energy (DOE). Similarly, understanding the dynamic interactions between living organisms and the environment is critical to predicting and mitigating the impacts of energy production on the environment and human health.

Biological systems are extremely complex and thus challenging to understand (see box, Characteristics of Biological Systems, this page). This inherent complexity makes it difficult to predict and manipulate their behavior, for example, to produce biofuels or mitigate environmental contamination. The advent of molecular biology steered science toward "fundamental reductionism" (Woese 2004), which focused research on determining the structure and function of individual cellular parts. To develop a predictive understanding of these complex systems, today's scientists are starting to build mechanistic models to learn how individual cellular components as well as whole cells and mixed-organism (heterogeneous) communities can give rise to emergent properties. The biological research paradigm thus is transitioning from understanding *individual parts* toward understanding biology at a *systems level*. The application of systems-based biological approaches will provide a basis for enhancing and manipulating "natural" processes for societal and environmental benefits.

Contributing to this rapid shift in approach are the availability of whole-genome sequences for an increasing number and variety of organisms and the technical advances that enable such high-throughput genome-

Characteristics of Biological Systems

Why Biological Systems Are Difficult To Predict

- **System components (often unknown, multifunctional, and heterogeneous) operate in crowded volumes.**
- **Seemingly minor system components (rare events and rare organisms) can have major functional influences.**
- **Functional systems involve multiple component levels (e.g., molecular, molecular assemblies, cells, tissues, communities, environments, and ecosystems).**
- **Functional systems can be hierarchically arranged, though connections between levels are poorly understood.**
- **Functional systems display emergent properties at all system levels, thus making extrapolations difficult.**
- **Functional systems exploit small thermodynamic gradients and stochastics.**
- **System operation**
  - **Is dynamic and occurs across multiple spatial and temporal scales**
  - **Often depends on spatial and temporal ordering**
  - **Can be nonlinear and adaptive across a range of conditions.**

Simplifying Features of Biological Systems

- **Diverse systems' sharing of a common molecular basis (e.g., nucleic acids, proteins, lipids, and carbohydrates) allows comparative analyses to elucidate their operation.**
- **Systems' frequent modularity allows approximation of particular aspects.**
- **Systems' frequent exploitation of similar feedback mechanisms allows motifs to be recognized and understood.**

scale measurements. As a result, biological science is evolving from being merely descriptive to providing quantitative and mechanistic explanations of cellular processes. Ultimately, such a predictive science—and the crucial capabilities for collecting and using large amounts of biological data—will be required to develop and evaluate models of biological function.

Understanding how different assemblies of molecules in cells give rise to functional outcomes is fundamental to systems and synthetic biology. The advancement of this new biology will require novel analytical and computational technologies and approaches to validate the predictions derived from system-level research. Varied methodologies will be needed to measure and track the temporal and spatial disposition of molecules that differ widely in structure and concentration in cells and to determine how networks of proteins and regulatory molecules are organized and give rise to specific activities at larger spatial scales. Moreover, such measurements are needed to characterize individual cells within populations as functions of space and time and varying environmental conditions.

To solve its mission challenges, the Office of Biological and Environmental Research (BER) within DOE's Office of Science promotes the development of systems biology approaches—the major thrust of its Genomic Science program (formerly Genomics:GTL; see Fig. 1.1. Genomic Science Program Goal and Objectives, below). In a 2006 review of this program, the National Research Council (NRC) endorsed DOE's systems biology research in plant and microbial biology. Furthermore, NRC stated that development of new technologies and methods was not only a logical pathway but an essential one for accomplishing the program's goals. The major aim of BER's Genomic Science program is to develop a predictive understanding of biological systems relevant to DOE mission challenges and to develop the commensurate tools (see Fig. 1.2. DOE Genomic Science Program, p. 3). To date, the program has emphasized system-level investigations of individual organisms and the generation of data for constructing and evaluating predictive mechanistic models. Longer-term goals of the program are to develop a predictive understanding of complex biological communities and to improve the current technical

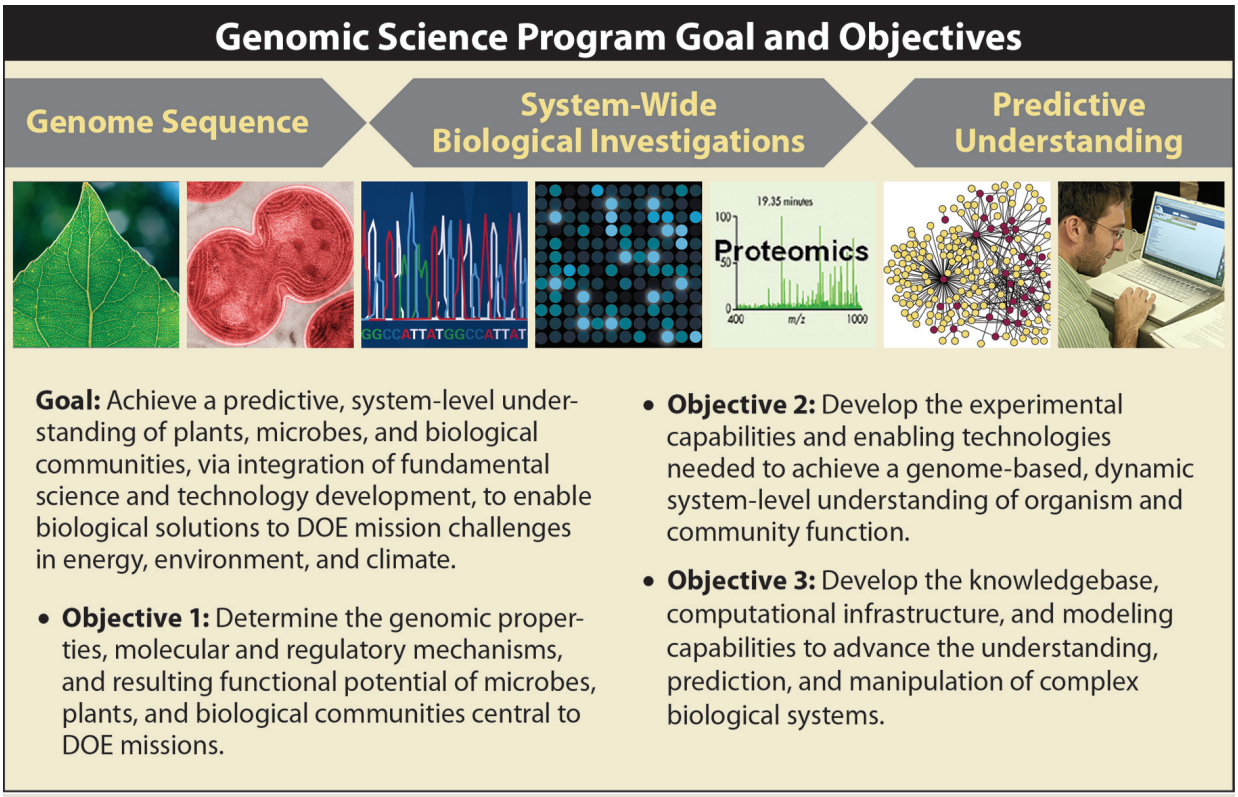


Fig. 1.1. Genomic Science Program Goal and Objectives.

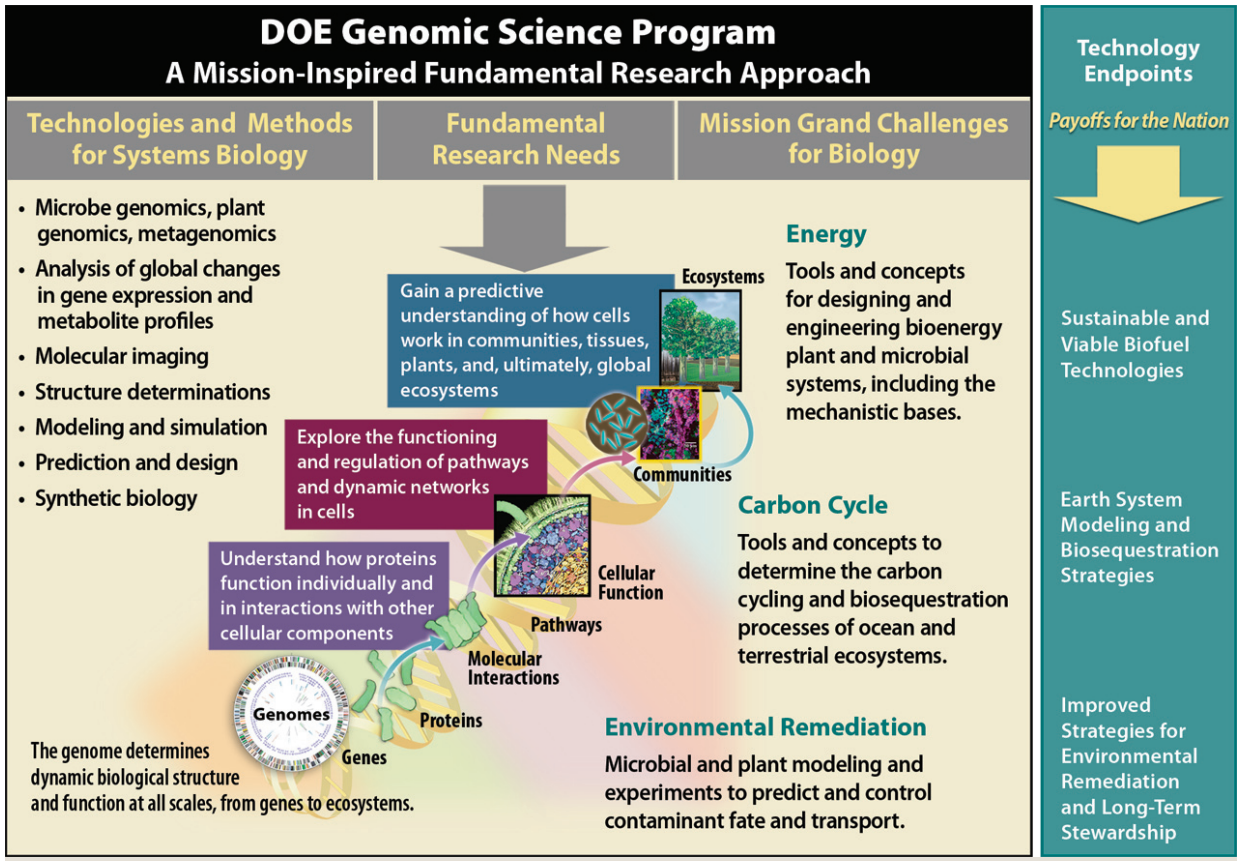


Fig. 1.2. DOE Genomic Science Program.

and scientific infrastructure to address these problems. These specific goals and infrastructure improvements have been described and reiterated in several reports (American Academy of Microbiology 2006) and Genomic Science roadmaps, including *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda* (U.S. DOE 2006), *Carbon Cycling and Biosequestration: Integrating Biology through Systems Science* (U.S. DOE 2008), and *U.S. Department of Energy Office of Science Systems Biology Knowledgebase for a New Era in Biology* (U.S. DOE 2009). These reports discuss the promise of a systems biology research approach in addressing critical DOE missions and complex problems of national and international interest, and they lay out the technical and scientific requirements for timely solution to such challenges.

As systems biology research is still at an early stage, emphasis in BER's Genomic Science program has been on the select tools and fundamental science required to conduct this research. Recent developments in several areas are bringing new opportunities to accelerate science. For example, maturing large-scale "omics" technologies are allowing cell responses to be monitored over increasing time scales. More targeted analytical techniques are being used to measure multiple responses at the single-cell level, and reconstituted multicellular systems are becoming more sophisticated. These and other advances are improving our understanding of organisms at a systems level, but many technological barriers still need to be overcome before significant progress is made in achieving a predictive, system-level understanding of plants and microbes.

Over the last several years, the Genomic Science program has made substantial investments in fundamental analytical technologies needed to support systems biology studies of plants and microbes, especially in the areas of proteomics and genome sequencing. Many hundreds of different organisms have been sequenced, and the metagenomes of numerous microbial communities have been analyzed. The program has laid the

foundation for understanding the regulatory dynamics of individual model microorganisms such as *Shewanella*, *Geobacter*, and *Desulfovibrio*, as well as for deducing the gene and protein networks that control basic microbial functions. Techniques are being developed to identify protein complexes (Pelletier et al. 2008; Dong et al. 2008; Hura et al. 2009) and to track the position of molecules in complex systems at the cellular level (Vinson and Chin 2007; Blow 2009; Xie et al. 2008). Progress in these areas has resulted largely from the DOE emphasis on developing new and innovative tools to enable such research. Although current technologies, such as genome sequencing and high-throughput methods for characterizing proteins and protein complexes, have been and will continue to be critical to systems biology research, DOE recognizes that new tools and capabilities are essential for analyzing and understanding complex biological systems and validating their individual components. To that end, the objective of DOE's New Frontiers in Characterizing Biological Systems workshop was to identify the most significant scientific challenges facing biology in areas relevant to DOE missions and to develop a vision for the new capabilities and tools needed to meet these challenges.

2. Biological Challenges

To develop a predictive understanding of biological systems, BER's Genomic Science program has emphasized systems-level investigations of individual organisms, primarily genome-based studies of microbes. Although current research approaches have not achieved the ultimate goal of measuring and predicting the response of individual cells within a complex environment, they have revealed the challenges involved. These challenges must be met to realize the program's longer-term goal of being able to predict the behavior of complex, natural communities of organisms from the behavior of their dynamic individual parts. To understand the connection between the genome and natural environmental systems, three interrelated grand challenges have been identified:

1. Understanding the cell and its response to chemical and physical perturbations.
2. Understanding interactions between cells.
3. Understanding dynamic biological systems across multiple scales of time and distance.

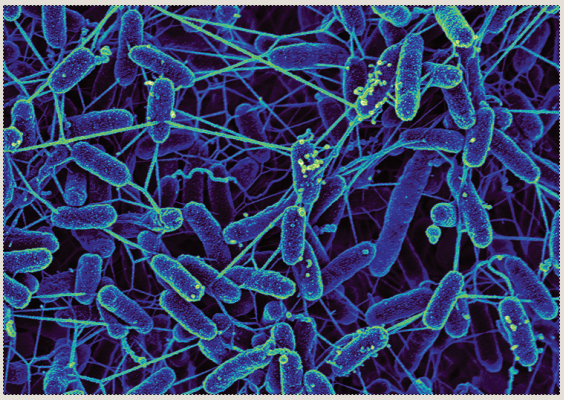
The key biological challenges described in this chapter are summarized on p. 17.

To achieve this understanding and address critical questions at the frontiers of biology, the scope and scale of current measurements must be expanded. In many cases, this involves combining systems-level physiological studies with more-focused measurements of individual components, an approach that requires greater spatial and temporal resolution. These studies will aid the construction and evaluation of mechanistic models that can predict the consequences of cellular and environmental changes to an organism's regulatory network.

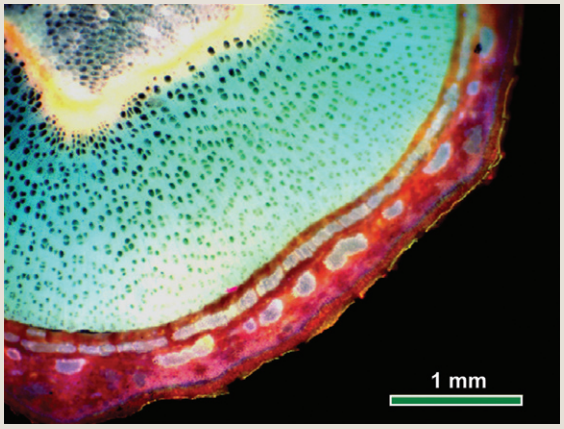
Overcoming these challenges will allow better use and realization of the information contained in the genome and will enable the biological and environmental science communities to move forward in addressing concerns in energy production, carbon cycling and biosequestration, and environmental remediation (see Fig. 2.1. Cellular Systems for Diverse National Needs, this page). Details of these challenges and the technological measurements needed to overcome them are highlighted in the following sections.

This is an exciting time in science; advances in biology and concomitant improvements in technology and information processing have led us to the point

Fig. 2.1. Cellular Systems for Diverse National Needs. *Guided by the biological information encoded within genome sequences, we can begin to identify, understand, re-engineer, and harness specific cellular systems for energy production, environmental remediation, and other national needs. [Image credits:* Shewanella oneidensis *image by Rizlan Bencheikh and Bruce Arey, Environmental Molecular Sciences Laboratory, DOE Pacific Northwest National Laboratory. Poplar image by DOE BioEnergy Science Center, DOE Oak Ridge National Laboratory.]*



Shewanella oneidensis *cells can reduce uranium and metals in contaminated environments.*



Different cell types make up the various plant tissues in Populus trichocarpa *(a poplar species), a model system for bioenergy and carbon cycling.*

where we can begin to address biological processes as they occur on a whole-systems scale. With appropriate coordination and efforts, the coming decade can see major advances in understanding how individual cellular properties give rise to community responses. In turn, this should lead to revolutionary advances in the application of biological solutions to societal needs.

2.1 Understanding the Cell and Its Response to Chemical and Physical Perturbations

The cell is the basic building block of life, and yet our understanding of systems' architecture and information flow in even the simplest cells is rudimentary. Methods to probe many of the dynamic processes occurring within the cell with the required chemical, temporal, and spatial resolutions simply do not exist. Additionally, the ability to assess cellular response to extrinsic chemical and physical perturbations is limited. Ultimately, understanding the full range of cellular phenomena will aid in interpreting a cell's niche within its community and its relationship to other organisms.

Even in the postgenomic age, the "parts list" of cellular components is incomplete. Prior efforts in the Genomic Science program have aided in defining genetic-based components and many mechanistic relationships. However, to date, our understanding of small molecules and the various modifications, interactions, and functions of proteins is far from complete (see Fig. 2.2. A Glimpse into the Cellular World, p. 7). Thus, there is a pressing need to develop new tools for defining the functional role of cellular constituents and to characterize the full suite of interactions occurring within model cells. These efforts will require innovative engineering and measurement platforms as well as knowledge about cellular function and fate. With completion of the dynamic parts list, an important goal will be to characterize and validate the subcellular and cellular networks that regulate energy and material flows. For example, the key pathways and genes for lignin and carbohydrate monomer synthesis are known, but the full suite of genes and pathways for polymerization, crosslinking, and structure is not. Ideally, such efforts will allow probing a cell within its environment, thereby gaining the information needed to influence microbial and plant systems for specific purposes such as maximizing usable energy output.

In the quest for predictive science, the question is: once we have the parts list of the cell and understand how it changes over time in response to environmental factors, can we build realistic models of cells that capture and simulate this information? The greater the number of responses that can be measured simultaneously, the greater the specificity of the models that can be constructed. However, not all cells within a population respond equally, so simultaneously measuring multiple activities in single cells will be required. In the near term, this capability is likely to be achieved by measuring the expression of multiple genes at the individual cell level and, with future advances, measuring the activity of multiple signaling pathways or mapping the locations and activities of proteins and protein complexes.

Performing nondestructive single-cell measurements presents a considerable challenge. Current labeling technologies for live cell imaging are limited to only a handful of reagents; expanding this reagent set will be essential for understanding how cellular parts assemble and interact over time. Advances are needed in the development of "nonperturbing" labels or tags that provide the means for directly detecting specific molecules. Measuring force, temperature, pressure, shape, and electrical potential within the cell is surprisingly difficult, and yet these parameters must be characterized and correlated to the genome and molecular characteristics, cellular activity, and cellular fate to understand how cells operate. Additionally, the effects of chemical and nonchemical changes, such as mechanical and electrical forces and physical parameters, must be understood. Such information is necessary to comprehend how cells respond to stress and environmental changes.

The greatest cellular measurement need is to develop methods that operate on wide spatial, chemical, and temporal scales and that provide data that are global in nature. There is a plethora of existing tools for characterizing well-defined individual parameters within cells, but most of these measurements are limited to a single dimension, either spatial or temporal. These current measurement approaches require careful selection and optimization of the probe for each parameter being measured. Moreover, they lack the ability to be adequately multiplexed to provide more

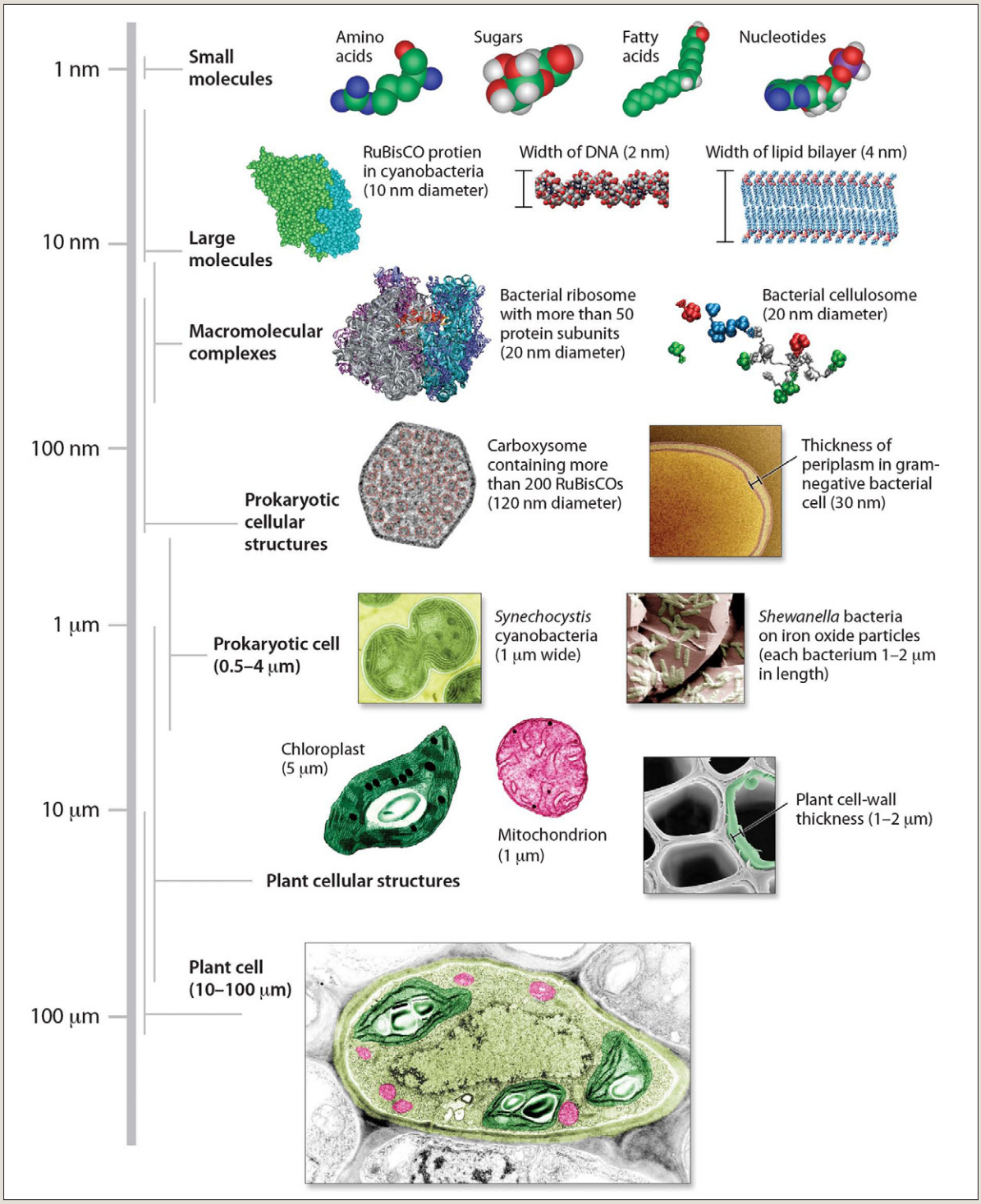


Fig. 2.2. A Glimpse into the Cellular World. *Molecules and structures within cells largely determine cellular behavior and characteristics. Although a single cell is one of the most chemically complicated systems known, just a few basic types of small molecules give rise to the extraordinary chemical diversity of life. Small molecules—typically containing 30 or fewer carbon atoms—are linked to assemble complexes and structures of increasing size and complexity. We know the genome provides instructions for building nucleic acids from nucleotides and proteins from amino acids, but much less is known about how the cell directs the assembly of macromolecular structures derived from sugars, fatty acids, and other small molecules or how the cell builds even larger cellular structures from macromolecules. The figure shows spatial scales associated with several molecules and structures in plant and microbial cells.*

global information. For example, although the interaction between a few specific molecules can be probed using microscopy approaches such as fluorescence resonance energy transfer, this technique does not support the simultaneous measurement of many interactions. While mass spectrometry does allow the global characterization of cellular contents, it is not yet possible to measure the levels of a large number of different proteins in a dynamic fashion at the single-cell level. Increases in the wavelength range and intensity of spectroscopic approaches could provide improvements in spatial and temporal resolution, but, without multiplexed probes, they will be unable to acquire the information most needed to understand complex cellular responses.

New tools to characterize and quantify the characteristics and molecular parts of individual cells and their regulatory networks as they respond to chemical and physical perturbations will enable the understanding, prediction, and manipulation not possible today. Such capabilities will illuminate the fate and transport of contaminants, cellular responses to climatic change, nutrient bioavailability, carbon biosequestration, how cells live in extreme environments, and processes for biomass degradation (see sidebar, Real-Time Chemical and Structural Analyses Are Needed To Improve Enzymatic Degradation of Lignocellulosic Biomass, p. 9). Many of these topic areas cannot be addressed with existing technologies and computational resources. Development and advancement of the next generation of imaging, spectroscopic, and high-throughput characterization tools will aid in understanding cellular processes and lead to capabilities for manipulating and optimizing a cell's output for desired applications. An improved understanding of cellular components, their spatial and temporal arrangements into networks, and their responses to chemical and physical changes will prove to be transformative for the biological sciences. Without a detailed understanding of cells—the building blocks of organisms and environmental communities—our ability to model and harvest their output will remain limited.

2.2 Understanding Interactions Between Cells: From One to Many

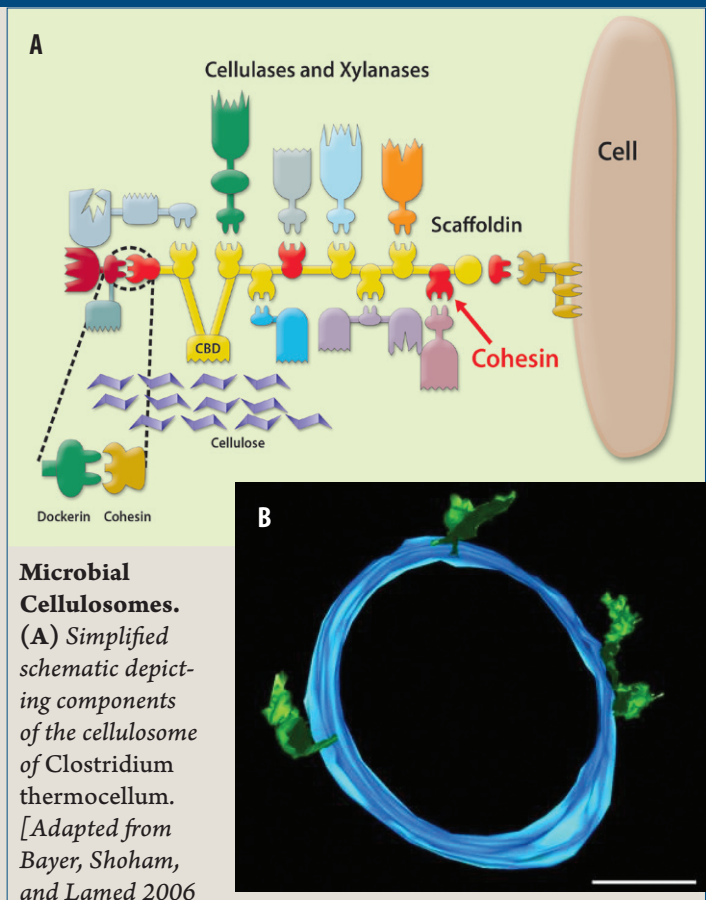
Whether found in a multicellular organism or surrounded by other cells within a heterogeneous community, a cell rarely acts independently from others. Rather, cells act as part of a community, tissue, or ecosystem. Biological function is critically dependent on such interactions. Elucidating how cells interact with one another is necessary for predicting how natural processes arise from the collective function of individual cells.

By understanding and manipulating populations of cells, cellular communities, and complex organisms, the research community stands ready to make the leap to understanding cellular function in the context of the surrounding physical and chemical environment. The activities of individual cells give rise to collective phenomena that have tremendous impact at macroscopic scales. However, even within populations considered to be clonal, there are extensive cell-to-cell variations in properties and behavior that may be due to stochastic variations in gene expression, exposure to variable microenvironmental conditions, or other factors. A well-known example is the differentiation in *Anabaena*, a type of cyanobacteria in which only some cells form nitrogen-fixing heterocysts. In heterogeneous populations, cell-to-cell variations typically are even more pronounced. This heterogeneity impacts the net function of a population of cells and is what makes functional measurements difficult. An example of collective phenomena arising from the activities of heterogeneous cells is the decomposition of structurally and chemically complex organic material by microbial communities, as seen in the degradation of lignocellulose in the termite hindgut by a microbial consortium (Warnecke et al. 2007). Another example is the cleanup of groundwater contaminants by the rare microbial species *Dehalococcoides* (see sidebar, Microbial Environmental Remediation of Chlorinated Solvents, p. 10). There is a critical need to develop global measurement approaches that do not average large populations of cells but characterize the composition and activity of the individual cells making up the population or community. Among other challenges, the global approaches for cell characterization described in Section 2.1, p. 6, need to be deployed in high-throughput mode.

Real-Time Chemical and Structural Analyses Are Needed To Improve Enzymatic Degradation of Lignocellulosic Biomass

Understanding functional cellular processes will be greatly facilitated by better insight into the component parts, their orchestration, and the multiple length and time scales in which they function. An illustrative example is microbial degradation of lignocellulosic biomass that occurs at microbial-plant interfaces. Understanding this process is critical to improving both plant biomass conversion to biofuels and the carbon cycling models used in climate projections.

Various extracellular enzymes, such as glycosyl hydrolases (i.e., cellulases), are tethered to the microbe's surface using a protein scaffold. The cellulases may be "sitting" from tens to a hundred nanometers from the cell surface while interfacing with their lignocellulosic substrate. A more complex example is the cellulosome of *Clostridium thermocellum*. This cellulosome, analogous to a Swiss Army knife, is a multienzyme complex with multiple glycosyl hydrolases bound to a common flexible scaffoldin protein. The cellulosome is bound to the biomass substrate by a carbohydrate binding domain (CBD) and is tethered to the host microbe (see figure, this page).



Microbial Cellulosomes. **(A)** *Simplified schematic depicting components of the cellulosome of* Clostridium thermocellum. *[Adapted from Bayer, Shoham, and Lamed 2006 and used with kind permission from Springer Science and Business Media.]* **(B)** *Electron tomography image showing a cross-section of a cellulosome (green) tethered to the cell surface of* C. cellulolyticum *(blue). [Image courtesy of Donohoe and Haas, National Renewable Energy Laboratory, 2009, unpublished data.]*

Despite years of study focused on glycosyl hydrolases, a more detailed understanding is needed of how the molecular parts are orchestrated in space and time. Cellulases are known to be slow degraders, and the reaction is often incomplete. Current questions whose answers may improve the action of hydrolytic enzymes on biomass include:

- Is the maximum potential turnover rate limited by the intrinsic kinetics of action on a solid substrate?
- Regarding the mode of enzyme binding and initiation of the reaction, how does the polysaccharide thread into the active site?
- What are the effects of the heterogeneity of the lignocellulosic biomass surface on the extent of the reaction? Does the presence of lignin cross-linking "stall" the hydrolysis progress?
- When using different glycosyl hydrolases, what are the role and mode of synergy that will improve degradation rate and extent? Do different enzymes help process different biomass cross-linkages?

To address these challenges, dynamic chemical and structural measurements need to be made at the enzyme-lignocellulose interface as enzyme catalysis proceeds. Essentially, nanometer-scale resolution of chemical and structural information needs to be coupled with dynamic information that spans from milliseconds to minutes.

Microbial Environmental Remediation of Chlorinated Solvents

Low-abundant *Dehalococcoides* species carry out the complete environmental remediation of the chlorinated solvents tetrachloroethene and trichloroethene (TCE) as well as other chlorinated ethenes. TCE compounds—the most abundant groundwater contaminants in the United States—are stepwise reductively dehalogenated to ethenes, via cis-dichloroethene and vinyl chloride (VC) as intermediates. These reductive dehalogenation reactions represent the sole energy-conserving pathway, known as organohalide respiration, in *Dehalococcoides*. Key enzymes in organohalide respiration are cobalamine- and Fe-S clusters containing reductive dehalogenases and hydrogenases. *Dehalococcoides* are strict anaerobic, H_2-consuming bacteria highly adapted to this organohalide respiration niche. Despite a streamlined genome of ~1.4 million base pairs (among the smallest of any free-living microorganism), *Dehalococcoides* show great diversity in reductive dehalogenases, 36 of which are found in strain VS. Interestingly, most of this important diversity in reductive dehalogenases—including those responsible for TCE and VC reduction—did not arise by gene duplication but were recently acquired by lateral gene transfer. In particular, transfer-messenger RNA—a bacterial RNA molecule with dual tRNA-like and mRNA-like properties—is emerging as a preferred integration site for the acquisition of new reductive dehalogenase genes.

Although a free-living microbe, the ecology of *Dehalococcoides* is firmly embedded into a complex metabolic network influenced by other community members. For example, the hydrogen substrate for organohalide respiration is provided *in situ* only via a complex anaerobic food web encompassing several other species. Moreover, *Dehalococcoides* cannot synthesize *de novo* the important cobalamine coenzyme of catabolic reductive dehalogenases, which thus must be acquired from members of the surrounding microbial community.

In noncontaminated environments, *Dehalococcoides* species are extremely rare and postulated to exist by respiration of low-abundant organohalides produced at low rates by innate organisms. This combination of high niche specialization of low-abundant compounds is believed to account for the natural low abundance and low growth rate of this population in complex communities. Yet the target of environmental remediation engineering efforts is this important, niche-specialized activity of rare *Dehalococcoides* sp.

Important components of global approaches are those that illuminate the interface between cells and their environment and how the interface facilitates the flow of chemical and physical information (see sidebar, Uncovering Key Interdependent Relationships Among Organisms To Leverage Beneficial Capabilities, p. 11). Learning how individual cells regulate these processes and how they sense and respond to their environment represents a particular challenge. In addition, there is the need to identify and characterize critical molecular interactions, including those that constitute communication pathways between cells, attachments between cells and the extracellular matrix, and autocrine signaling such as quorum sensing. The types of information that need to be elucidated include the substrate(s) for each cell type, the product(s) it produces, and the spatial relationship both between the various cells consuming and producing different materials and with the polymeric organic or inorganic materials to which the cells are interfaced. To follow material flow, concentrations of different substrates need to be mapped onto the cell population and the fluxes of various substrates through individual cells determined. Although such a dataset is important, acquiring it is well beyond our current technical capabilities.

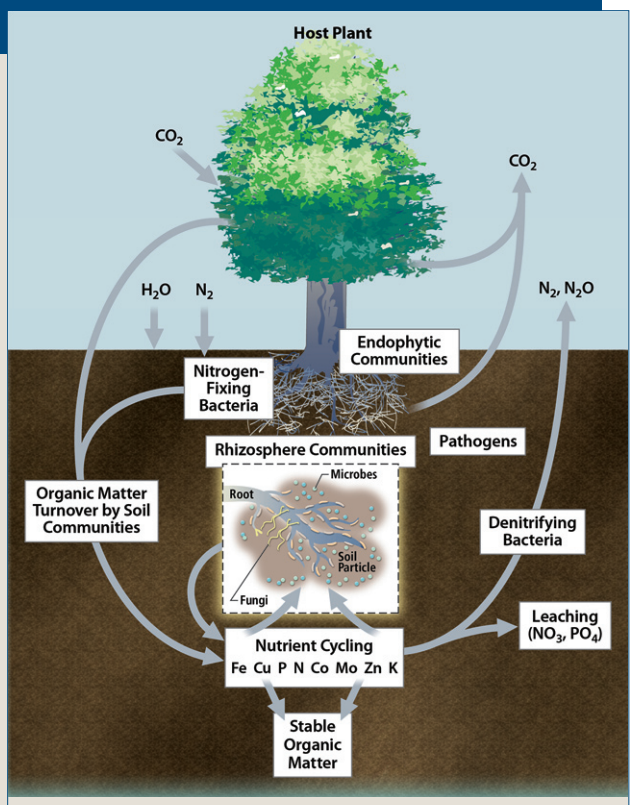
Even within the field of genomics, which uses analytical technologies that are extremely sensitive, most measurements require large populations of cells. This size is necessary regardless of whether the purpose is to determine genome sequence or measure the products of genome expression (i.e., transcriptome or proteome). However, building accurate models of the collective behavior of cells requires the capability to measure properties and responses of the individual cells within a community.

Uncovering Key Interdependent Relationships Among Organisms To Leverage Beneficial Capabilities

Prior efforts and investments in characterizing genome sequences and the functional activities of gene products have focused primarily on the role of these molecules in the context of individual organisms. Increasingly apparent, however, is that an organism's environment, including neighboring species, strongly influences how that organism uses its genetic-based information.

The relationship between plants and microbes exemplifies such a complex, multiorganism system, dependent on both the organisms involved and the environmental forces acting upon them (Bisseling, Dangl, and Schulze-Lefert 2009). Understanding this relationship is critical to exploiting natural routes to environmental remediation processes, the terrestrial cycling of carbon, and the development and management of renewable energy sources (see figure, this page). Microorganisms are intimately associated with various plant tissues, producing short- and long-term effects on plant growth and development. Short-term effects are especially important for improving plant establishment on marginal soils. Microbes can accelerate plant root development, thereby providing the plant with better access to nutrients and water and reducing the need for irrigation and soil amendments. These better-developed root systems also increase the belowground flux of carbon. Exploiting this property can improve belowground storage of recalcitrant carbon forms and enhance soil quality for increasing biomass production in which more labile carbon is released. Additionally, microbes can facilitate fast initial growth of a particular plant, allowing it to out-compete others for available resources. Over the long term, microbes improve plant growth, health, and survival. They also can counteract stress responses caused by drought or the presence of contamination when plants are grown on marginal soils or are used for managing and remediating contaminated sites. Furthermore, microbes can protect a plant against pathogens and directly assist the host plant by producing antimicrobial compounds.

Effective extrapolation of such beneficial interactions requires understanding the relationships between the genetic characteristics of the organisms involved and how they proportion and exchange information, energy, and materials among themselves. Specific knowledge gaps that need to be addressed include:



Plant-Microbe Interactions.

- Determining the mechanisms for recognition and colonization of plants by beneficial microorganisms.
- Understanding the temporal and spatial variations in plant colonization by beneficial microorganisms.
- Determining microbial community structure in the rhizosphere.
- Establishing how important endophytic and rhizospheric microbes are in fostering carbon and nutrient sinks, and how the size and recalcitrance of these sinks might be influenced to affect belowground carbon biosequestration and aboveground biomass production.
- Determining how the observed microbial stimulation in plant growth and development might be controlled to tailor biomass composition to less-recalcitrant forms better suited for biofuel production.

This difference between the level at which we can make measurements and the level at which we need understanding is both a conceptual and a technical challenge.

Another particular challenge is the ability to dynamically monitor a selected individual living cell in a complex environment in the presence of other live cells. Electron microscopy and other current technologies exhibiting excellent spatial resolution often damage cells because they require special preparative techniques that "fix" and stain cells to maintain their structure under high vacuum and to provide contrast. To truly understand the fundamental mechanisms of cellular responses, noninvasive (or minimally invasive) methods such as fluorescence microscopy will be needed to measure metabolism and other key cellular properties in these selected cells within the population. These properties include expression of specific genes or networks of genes or proteins in a specific metabolic pathway. Because many cellular signaling and metabolic pathways act on time scales of milliseconds to seconds, there also is a critical need for single-cell measurement techniques with similar time resolution. Thus, techniques are needed not only for spatially resolving measurements in specific cells, but also for temporally resolving them.

Biologists are beginning to address the question of "who is present" using rapid and sensitive nucleic acid technologies to determine phylogeny, but measuring function is much more difficult and represents an emerging challenge. Current capabilities can provide insight into microbial community structure but cannot detect activity at the single-cell level. The traditional approach of measuring microbial activities under a narrow set of optimal abiotic conditions in the laboratory does not provide a sufficient basis for predicting community responses under the wide spatiotemporal variability of abiotic conditions in natural environments. Thus, approaches such as stable isotope probing are needed for detecting the microbes that are active (e.g., respiring, performing maintenance, producing proteins or other cell constituents, or replicating) under a wide range of conditions and revealing their functions in real time (Dumont and Murrell 2005). We also need to be able to enrich and isolate microbes involved in key processes in the environment, but the vast majority of microorganisms living there are perhaps not cultivatable. Thus, different strategies are needed for performing manipulative experiments. Along with patient and careful "conventional" approaches, new techniques are needed to capture as-yet unidentified metabolic capabilities for study under carefully defined conditions in the lab.

Other challenges include linking different types of data. As one example, how do we link genomic-based information on individual cells to different levels of molecular information, such as the transcriptome, proteome, metabolome, regulatory networks, and even the environmental niche? Another critical unanswered question is whether or how a cell's genetic program is coordinated with that of neighboring organisms. We also do not know how the phenotype of a population arises from the functions of individual cell types. Some methods show promise for monitoring activity at the level of single microbial cells within complex mixtures; these methods include the use of stable isotope–labeled compounds in combination with nanoSIMS, microautoradiography–fluorescence in situ hybridization (MAR-FISH), and Raman-FISH (Neufeld, Wagner, and Murrell 2007). New methods are needed to accurately measure a greater variety of biomolecules—ranging from proteins, nucleic acids, carbohydrates, and lipids to low-molecular-weight metabolites for selected cells in the community or population. Single-cell transcriptomic, proteomic, and metabolomic measurements are needed, but whether such measurements are possible for most molecules is unclear. Nevertheless, it is essential to understand the flux of metabolites through particular metabolic pathways and the responses of regulatory networks in individual cells under relevant conditions. This information would move us away from the current approach of modeling populations of cells based on the potential of their genomes and into the area of modeling the "configuration" of individual cells within a given environment. Such capabilities would significantly advance the ability to translate genomic data into phenotype. In combination with improved capabilities for data integration, this would greatly enhance our ability to predict the response of heterogeneous biological communities.

2.3 Understanding Dynamic Biological Systems Across Multiple Scales of Time and Distance

Understanding and modeling cell dynamics represent two of the most significant challenges for the postgenomic era. A current limitation to understanding biology is the tendency to model bioprocesses as hierarchical, linear responses rather than combinatorial, recursive networks. A major challenge associated with the increased throughput of molecular analysis is managing large sets of data and translating them into predictive models. New modeling approaches will be required to represent the nonlinearity and adaptive behavior of biological systems. These approaches could involve modeling the responses as assets of a biological network rather than simply the programmed responses of cells. To advance these efforts, new approaches should bring together tools from such diverse disciplines as engineering, biology, and computational sciences. Systems biology approaches could be used to address the multiscale nature of a phenotype and then be applied to understand how the behavior of multiple cell types is coordinated physiologically or in response to external stimuli (see sidebar, Working Toward Real-Time Measurements of Biological Responses: The Low Dose Radiation Challenge, this page).

Most biological problems are multiscale and thus complex. Understanding the nature of this complexity is critical because microscale biological and chemical interactions in large part are moderated by signaling between cells and organisms that, in turn, influence macroscale ecosystem compositional dynamics. There also is a need to adequately define "complexity," especially as it pertains to microbes and interfaces or eukaryotic cells in their microenvironments (see sidebar, Targeting the Interfacial Region Between Microbial Cells and the Geosphere, pp. 14–15).

Working Toward Real-Time Measurements of Biological Responses: The Low Dose Radiation Challenge

Low dose radiation research seeks to determine when and why the physical attributes of radiation (i.e., dose and dose rate, volume, and quality) impact human health. The physical interactions of radiation energy with biological macromolecules are very well defined on the femto to millisecond time scales, and in the last decade so too are the immediate biological responses (i.e., damage recognition, repair, and cell-fate determination). However, the ensuing biological changes lose *definition* over the course of days. How can detailed mechanistic information about energy deposition be used to identify significant cellular or molecular events? How can this information in populations be linked across cell types and long time scales? These low dose radiation problems reflect common issues of data integration and analysis that span Department of Energy research problems.

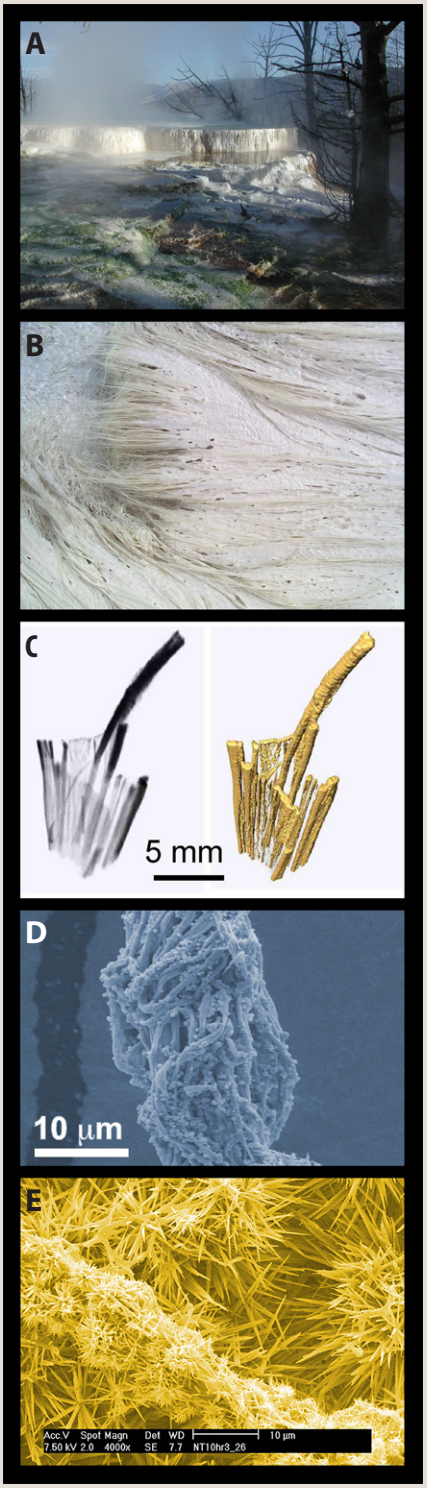
Understanding the biological bases for health effects caused by exposure to low doses of radiation requires interrogating multiple scales of information (from micro to macro) over sometimes very long time scales (from minutes to years). Non- or minimally invasive interrogation of multiple critical events *in vivo* is needed. A research framework might be initiated to a first approximation by analyzing fixed specimens across relevant time scales to generate a statistical model of the interactions between irradiated cells, tissues, and organs. The second generation might use complex *in vitro* models, but the ultimate goal is to analyze the evolution of interactions in an organism in real time.

Radiation exposure has a finite, dose-dependent probability of generating stochastic genetic change (e.g., mutation). However, the biological response to radiation is both nonlinear and deterministic, and it is the interaction of these two types of processes that can produce cancer. Thus, the fundamental problem is to identify and track rare events (aberrant cells) and characterize their dynamic interactions within a tissue environment. Besides the technical breakthroughs necessary for multiplex analyses at multiple scales, a critical need is to develop the computational models that identify watershed events, pathways, hubs, or processes and to mechanistically link them across time scales.

Targeting the Interfacial Region Between Microbial Cells and the Geosphere

Microbial life has been closely intertwined with the geosphere throughout Earth's history. Microorganisms, because of their small size, high ratio of surface area to volume, and incredibly diverse metabolism, have a tremendous influence on their environment through the transfer of energy and materials across complex biologic-solvent-solid interfaces. These microbes are agents of rock and mineral weathering as well as catalysts for their formation. Although the products of such microbial "sculpting" of the geosphere are often evident at large scales over the surface of the planet, the interplay between microbes and geological materials is dominated by processes at the molecular and microscopic scales (see figure, Interfaces Between Microbial Cells and Reactive Solids, from Landscape to Microscopic Scales, next page). The microbe-mineral interface is a prime example of this interplay and represents a complex, relatively unexplored domain that has important implications for carbon cycling and biosequestration, environmental remediation, and bioenergy. This interfacial region between microorganisms and minerals is dynamic, with chemistry and structure determined by interplay and response. The properties of reactive surfaces, cells, and adjacent regions alike are difficult to characterize because they are dynamic and occur at small (i.e., nm) scales. In 2000, the American Academy of Microbiology sponsored a colloquium, Geobiology: Exploring the Interface between the Biosphere and the Geosphere, and participants concluded that significant and critical events in geobiology happen at the level of individual cells or groups of cells. The ensuing report (Nealson and Ghiorse 2001) emphasized that the details of such processes would be revealed only by observations and measurements made at small scales—that of individual cells and their surrounding environment. This cell-scale research would seek to answer, for example, the following questions:

- How do microbial cells adjust their cell-surface molecular properties to facilitate the formation, dissolution, and weathering of minerals? Do cells similarly adjust their metabolism to alter the local chemical environment and thereby influence geobiological reactions? Do cells sense and respond to the precipitation of solids in association with their cell surfaces? If so, what are the environmental signals and the mechanisms of signal transduction?
- What are the mechanisms by which microbial cells catalyze $CaCO_3$ precipitation in advection-dominated systems?
- How do microorganisms engage with mineral surfaces and exchange electrons? How are electrons transferred across the outer membrane of gram-negative bacteria to minerals, metal ions, and electrodes?
- How do microscale biological and chemical interactions control and influence macroscale processes such as the formation and dissolution of minerals?



Interfaces Between Microbial Cells and Reactive Solids, from Landscape to Microscopic Scales. *[Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA. Images courtesy of Bruce Fouke.]*

(A) *A travertine terracette formed as groundwater emerges at 73°C and 6 pH, rapidly cooling and then degassing CO_2 to produce distinct covarying assemblages of microbial communities and $CaCO_3$ mineral precipitates. Aragonite (a polymorph of $CaCO_3$) crystal growths composing the travertine grew at a rate of 5 mm/day to create the 1-m-thick travertine terracette in less than 9 months. Controlled field experiments have demonstrated that microbial communities catalyze precipitation of the travertine $CaCO_3$ (Kandianis et al. 2008). [Image used with permission from the Society for Sedimentary Geology. From Veysey et al. 2008.]*

(B) *Field photograph of filamentous thermophilic microbes encrusted by $CaCO_3$ travertine precipitation. The spring water moves from right to left as a shallow (< 3 cm deep) turbulent sheet flow that drives rapid CO_2 degassing. The large (2 mm in diameter and 10 cm in length) individual microbial filaments are widely spaced and excrete draped sheets of extracellular polymeric substance (EPS) that contain abundant elliptical holes caused by water turbulence, gravity stretching, and gas-bubble release. The travertine directly encrusts and thus mimics the morphology of these filamentous microbial mats that then become well preserved in the geological record (Veysey et al. 2008). The 16S rRNA clone libraries, T-RFLP*, and metagenomic analyses indicate that these large filaments are* Sulfurihydrogenibium.

(C) *A microcomputed tomography scan (left) and a three-dimensional X-radiograph rendering of* Sulfurihydrogenibium *filaments (right).*

(D) *A* Sulfurihydrogenibium *filament, captured by an environmental scanning electron microscope (ESEM), showing that it is composed of small microbial filaments interwoven to create a larger filament.*

(E) *An ESEM micrograph showing aragonite ($CaCO_3$) precipitation growing on a* Sulfurihydrogenibium *filament at 71°C and 6.2 pH in a field experiment. Larger aragonite crystals surround densely packed smaller crystals coating a bacterial filament; significantly larger crystals are precipitating on EPS, filling the void spaced between microbial filaments. These crystal fabrics are evidence of microbial catalysis of aragonite crystal growth. [Image used with permission from the Geological Society of America. From Kandianis et al. 2008.]*

See also Fouke et al. 2000 and Fouke et al. 2003.

*Terminal restriction fragment length polymorphism.

Related to these challenges are questions concerning the "reducibility" of complex systems. For example, can the complex behavior of an observed system (e.g., a microbial interface) be subdivided into functional modules that can be fully understood in terms of their interactions? Or are there synergistic effects that make the system irreducible? The reducibility of a system has direct implications for determining how to treat it (i.e., microscopically or macroscopically) and at which scale to measure it. Another question related to complex systems involves "downward causation." Can global indicators and objectives (e.g., climate change, contaminant remediation, and biofuel production rate) influence the functional level from which the complex system must be understood? Knowing this is important for defining the level of granularity needed for constructing predictive systems-level models.

The success of new research to address these challenges will require integrated application of both reductionism and systems approaches. As a first step, meaningful systems-level interfaces need to be identified for study. One approach is to model cells as "input-output" systems in which the input of one or more pieces of information is translated into an appropriate cellular response. Currently, the most powerful (e.g., sensitive, specific, and diverse) methods to measure cellular responses are biochemical and molecular assays, whose use typically is practical only with relatively homogeneous populations of cells. Nevertheless, as technologies improve, data eventually will be available on systematic rather than representative responses of individual cells within a community. Incorporating these data will require building multiscale community models that will provide a framework for mapping the information derived from individual cells to the population's behavior. These models should encompass dynamic modeling approaches that include multidimensional chemical and physical data generated by a variety of techniques—the various "omics," microscopy, detection of activity, and detection of metabolic products. Community models should be capable of representing both extracellular and intracellular networks; predicting system stability under a range of conditions; highlighting areas of uncertainty; and elucidating feedbacks, thresholds, and nonlinearities.

Successful advancement of multiscale models for understanding biological systems will be facilitated by sophisticated new approaches to create model systems. Experimental design should include (1) observing, mapping, and characterizing molecular events across interfaces in the relevant natural environment; (2) controlled experimentation in the laboratory; and (3) controlled experimentation in the natural environment to test hypotheses derived from field mapping and lab experimentation. In addition, these controlled experimental systems must be able to incorporate advanced characterization technologies. Coupling these technologies with experimental systems typically has been achieved by developing some type of controlled experimental chamber. Such approaches involve moving the system and environment of interest to the analytical device. Determining how to conduct analyses *in situ* would be a significant advantage. Likewise, if data synthesis and modeling activities could be conducted simultaneously with experimentation and data collection, computational simulations could become a mechanism for informing and altering the ongoing experimentation rather than an exercise performed only after data have been collected and the experiment completed. For example, metabolic network modeling could exploit early metatranscriptomic experimental results showing the enzymes that are actually expressed by the system. This would narrow the scope of metabolic possibilities that otherwise would have to be established through labor-intensive experiments. However, certain aspects of biological systems and interfaces are currently only addressable by simulation and modeling. For example, water activity near a surface with cellulose or metal oxide crystals can be estimated best with molecular dynamic simulations. Thus, improved computational models and strategies are greatly needed for increasing the usability of this process for different problems.

Parameter-driven mathematical and physical models can describe many systems and processes, both natural and artificial. These models typically are used in a forward fashion, meaning that a certain set of values for model parameters yields a specific outcome. Defining an optimal outcome of the model is usually straightforward. A far more difficult, if not impossible, task is determining the corresponding parameter values

that, when applied to the model, yield the appropriate outcome. Frequently, more than one set of parameter values yields the same desired outcome, indicating that the model is degenerate. One way to determine optimal parameter values is analytically inverting the models. In many cases, this is analytically or practically infeasible because of model-related complexity and a high degree of nonlinearity (Fink 2008). It is unclear whether the optimal structure of a biological model can be defined for a given output, but such a development would transform this area of research.

Understanding and modeling dynamic biological systems across multiple scales of time and distance will require the simultaneous development of new experimental approaches for real-time measurements of individual cells within their natural microenvironments as well as new modeling approaches that can handle the nonlinear and adaptive nature of biological systems. It will be particularly important to develop realistic models of biological pathways as an important link between molecular data and higher-level biological functions.

Summary of Key Biological Challenges

2.1 Understanding the Cell and Its Response to Chemical and Physical Perturbations

- Complete the cellular "parts list." Identify and functionally characterize the full suite of cellular constituents and determine how specific subsets mediate the energy and material flows within cells.
- Measure cellular and subcellular environments. Comprehensively determine the physical and chemical factors that define the internal and external environments of individual cells.
- Expand the functionality and availability of tools and reagents for live-cell imaging. Advance the development of nonperturbing tools and methods for simultaneous measurements of multiple factors affecting cellular activity.

2.2 Understanding Interactions Between Cells: From One to Many

- Characterize the activities of individual cells that collectively impact macroscopic phenomena. Identify cellular characteristics and microenvironmental conditions responsible for cell-to-cell variations observed in clonal and heterogeneous populations.
- Explore the interface between cells and their microenvironments. Investigate the spatial relationships, physical connections, and chemical exchanges that facilitate the flow of information and materials between cells.
- Dynamically monitor selected individual cells within a population. Advance beyond average measurements of large populations of cells to targeted, real-time measurements of genome expression, metabolism, and other activities for selected cells under a wide range of conditions.

2.3 Understanding Dynamic Biological Systems Across Multiple Scales of Time and Distance

- Develop new modeling approaches that address the multiscale nature of phenotype. Establish new methods for modeling the emergent properties of nonlinear biological systems that are influenced by interactions among smaller-scale system components as well as larger-scale external phenomena.
- Support new model-driven approaches to experimental design that integrate reductionist and whole-system investigations. Insights from global analyses of environmental samples could drive laboratory-based experimentation and modeling that generate hypotheses for controlled field-based studies in natural environments.

3. Technological and Capability Needs

Technological developments are a hallmark of BER's Genomic Science program (see Fig. 1.2. DOE Genomic Science Program, p. 3). Progress in developing analytical measurement techniques and imaging capabilities has enabled genomic-scale characterizations and detailed assessments of biological systems at the organism level. Continued advancements, however, are essential for addressing the biological challenges outlined in this report. To gain a predictive understanding of systems, information on a large number of different chemical and biological species must be measured and understood across diverse length and time scales. This challenge is further complicated by the need to conduct these measurements on heterogeneous mixtures of biological components in natural settings. Currently, many tools are unable to obtain the required chemical, physical, and biological measurements occurring within complex, heterogeneous environments. Other needed capabilities simply do not exist. Several significant technological advancements, therefore, are necessary to measure, mechanistically understand, and predict the response of individual cells and their ultimate impact on environmental systems. These advancements include the following:

1. Expanding and integrating global characterization capabilities.
2. Identifying and measuring rare events and rare molecular species and cells, such as those within complex, heterogeneous biological systems.
3. Simultaneously measuring many chemical and biological species across broad spatial and temporal ranges.
4. Integrating and interpreting diverse information to create mechanistic models.

The key technological capabilities described in this chapter are summarized on p. 38.

Developing these capabilities, in the context of the biological challenges described previously, will facilitate our understanding of the connection between the genome and natural environmental systems. In turn, this understanding will allow us to address pressing questions concerning energy production, carbon cycling and biosequestration, and environmental remediation. Continued engagement among the physical, computational, and biological science communities—as facilitated by DOE's Genomic Science program—will place these measurement challenges within reach and result in forthcoming revolutionary advances.

3.1 Expanding Global Characterization Capabilities

Genomics, transcriptomics, proteomics, and metabolomics (collectively described as "omics" analyses) are approaches used to identify and comprehensively measure the molecular species produced by an organism or community under defined environmental conditions at specific points in time. Global analyses of RNA transcripts, proteins, metabolites, and other cellular constituents provide insights into the functions and physiological status of a biological system and indicate which parts of the genome are activated and translated into functional molecules as organisms and communities develop or respond to their environments.

3.1.1 Genomics

Sequencing the entire genome of an organism—once only a dream among scientists—is now routine. By enabling us to sequence organisms at an astonishing rate and for an affordable cost, current technologies have transformed genomics into the most accessible global characterization modality. This accessibility has been achieved by the decreasing cost and increasing throughput of sequencing over the past decade, as highlighted in Table 3.1. Advances in Sequencing, p. 20. This trend does not appear to be slowing. In fact, new technologies are emerging from efforts to lower costs for sequencing the human genome to $1000 (Pushkarev, Neff, and Quake 2009).

High-throughput sequencing also allows microbial communities to be investigated en masse, thus giving rise to metagenomic analyses. DOE has played a critical and pioneering role in these studies, which are far beyond what, until only recently, was deemed possible. The newest generation of

technologies enables the sequencing of a significant fraction of genomes in a simple microbial community. However, more-complex communities and rare genomes within a community still are not amenable to systematic characterization. In addition, capturing metagenomic profiles at multiple time points during environmental transitions is critical for adequately monitoring the genomic changes associated with an environmental perturbation. As the wealth of omics data grows, better tools are needed for data management, analysis, and integration (see Fig. 3.1. Categories of Global Omics Measurements, below, and sidebar, Whole-Genome Amplification Holds Promise for Single-Cell Genome Drafting, p. 21). Although improvements in sequencing and analytical technologies can reveal the presence of low-abundance organisms, these capabilities need to be extended to the single-cell level.

3.1.2 Transcriptomics

Although advances in DNA sequencing are impressive, they represent only the first step in global characterization of biological communities. A widely recognized need involves moving beyond potential functionality (DNA) to the world of actuation (RNAs, proteins, and small molecules). Global, high-throughput analyses of these molecules are conducted using ancillary techniques (metatranscriptomics, metaproteomics, and metabolomics) having well-known terminologies and concepts. None of these methods, however, is as universally applied and unbiased as genomics. Metatranscriptomics by *de novo* cDNA sequencing recently has been shown to be a viable approach to map both protein (Frias-Lopez et al. 2008) and functional RNA (Shi, Tyson, and DeLong 2009) expression. Another valuable advance would be single-cell transcriptomics, a method that could greatly improve our understanding of biological variability. Improvements in microfluidics-based sample processing eventually could lead to such measurements. As costs continue to fall, *de novo* sequencing likely will replace microarray approaches because of its high reproducibility, discovery potential, and elimination of the need to generate organism-specific microarrays. Regardless of approach, these applications currently face several challenges, including the development of an amplification method, the ability to detect

Table 3.1. Advances in Sequencing*

Platform	Million Base Pairs per Run	Cost per Base (U.S. ¢)	Average Read Length (Base Pairs)
Dye-terminator (ABI 3730xl)	0.07	0.1	700
454-Roche pyrosequencing (GS FLX titanium)	400	0.003	400
Illumina sequencing (GAii)	2000	0.0007	35

*From Hugenholtz and Tyson 2008. Reprinted with permission from Macmillan Publishers Ltd: *Nature*.

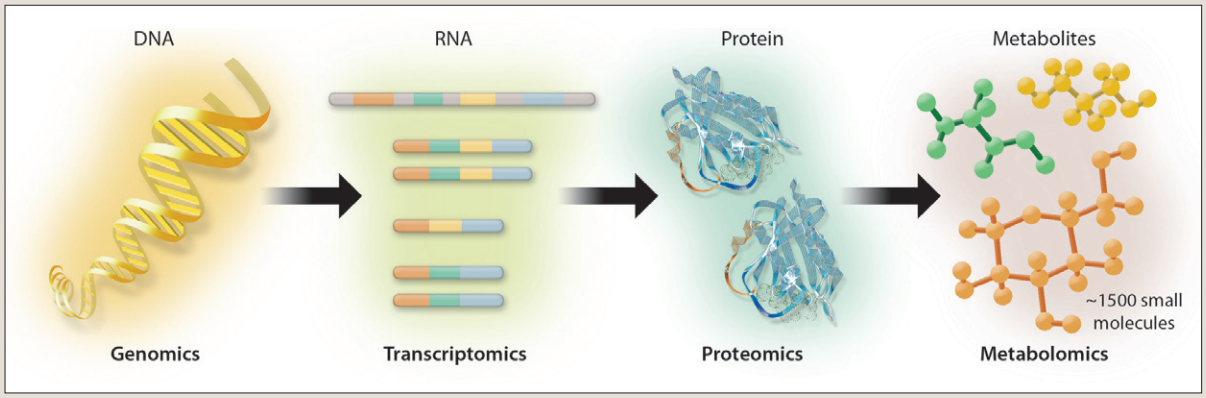


Fig. 3.1. Categories of Global Omics Measurements. *A cell has one genome and a dynamic transcriptome and proteome that all impact its metabolome. Metagenomic studies are global characterizations of heterogeneous microbial communities. Currently, our characterization capabilities decrease during the transition from genome to metabolome.*

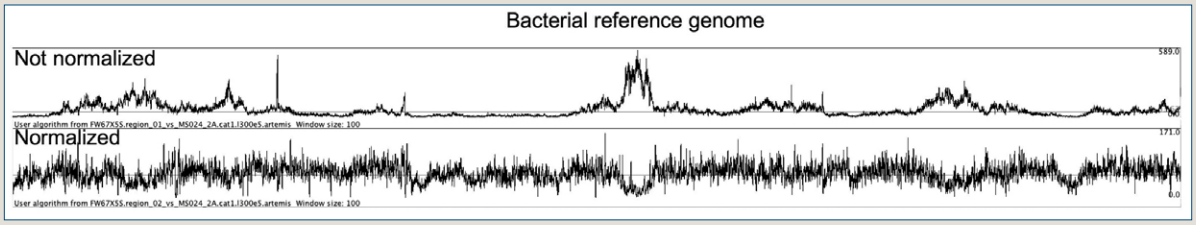
small regulatory RNAs such as miRNAs or nanoRNAs, and finer-scale measurements (see Section 3.1.6, From Bulk to Finer-Scale Characterization, p. 23).

3.1.3 Proteomics

Significant technical advancements in proteomic measurements have been achieved during the last decade. New generations of mass spectrometers and advanced separation technologies now allow the measuring of thousands of proteins (the product of mRNAs) in a population of cells and hundreds of post-translational modifications. Applying these techniques to microbial communities (metaproteomics) has provided promising and unexpectedly high resolution results, such as the ability to discriminate proteins from orthologous genes belonging to closely related strains (VerBerkmoes et al. 2009). However, current technologies are resource intensive and still do not offer a complete characterization of the proteome. Membrane proteins are particularly difficult to measure, in part because of their resistance to most proteases used in sample preparation. Until rare proteins can be measured, proteomics will continue to provide an incomplete picture of the cellular proteome. Proteomic analyses of environmental samples also have been constrained by challenges in extracting and purifying proteins from their environmental matrices. Peptide identification for proteins from complex mixed microbial communities presents another challenge because most members of such communities have yet to be identified and included in genomic databases. Further development of informatics approaches thus is required for identifying the function and origin of

Whole-Genome Amplification Holds Promise for Single-Cell Genome Drafting

Finer-scale sampling of microbial communities is essential to bring sample size in line with the size of the organisms being studied. However, reducing sample size unavoidably limits the amount of macromolecules available for analysis. In the case of nucleic acids, whole-genome amplification (WGA) methods provide the microgram quantities of DNA needed for sequencing from as little as femtogram quantities of starting material (the amount of DNA in an average microbial cell). There are several WGA approaches based on polymerase chain reaction (PCR), but unfortunately all result in incomplete coverage of the genomic template and are susceptible to biases. Ideally, WGA products should be random, complete representations of the starting material. Approaching this ideal are methods not based on PCR, including multiple displacement amplification (MDA). In practice, MDA produces a skewed coverage of DNA templates, particularly from very small amounts of starting template. Postamplification normalization using duplex-specific nuclease (DSN) appears to be the most promising approach to correcting biased coverage. Denatured MDAs are allowed to partially renature, with the most common (over-represented) regions of the genome renaturing first. DSN then is used to destroy the common double-stranded DNA, thus normalizing the MDA. The net result is a more complete coverage of the target genome, even from single-cell templates. With such capabilities, we may be on the verge of producing cost-effective, high-quality draft genomes from single cells.



The Great Flat Hope: Effect of Normalization of MDA Product from a Single Bacterial Cell. *Without normalization, genome coverage is highly skewed with peaks having >500× coverage and the average coverage being <1. After normalization, peaks are suppressed and the average coverage rises to ~60× for the same amount of sequencing. This greatly facilitates* de novo *assembly of the sequence data. [Plots courtesy of Tanja Woyke, DOE Joint Genome Institute.]*

proteins in uncultured microbial taxa. Other urgent needs include technology developments to increase the speed, sensitivity, quantification, and dynamic range of proteomic measurements and approaches to decrease the bias in protein coverage. Also needed are new innovative techniques that can provide additional insight into function, such as activity-based proteomic profiling.

3.1.4 Metabolomics

Although perhaps the oldest of the global omics technologies, metabolomics is the least developed, partially because of the large diversity of small molecules present in cells. Other challenges are the inability of current technologies to provide information on the spatial and temporal dynamics of metabolites and the problems in interpreting metabolomic data. For example, significant differences have been observed in the metabolomes of identical-appearing cells following specific environmental perturbations. The source of this variation is unclear. Further complicating efforts is the fact that only a small fraction of the metabolome can be characterized despite advances in mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy over the past 20 years. This limitation largely results from the lack of robust techniques for structurally characterizing metabolites identified by MS and other technologies. However, inadequate measurement sensitivity also constrains metabolomic characterization. For example, as a sample size decreases from tens of cells to even a single cell, the metabolomic depth of coverage decreases to only a few of the most abundant metabolites. However, the power of metabolomic measurements could be increased significantly by new approaches for combining multiple characterization techniques (e.g., NMR and MS) and the construction of specific metabolite libraries.

Measuring metabolic flux networks is critical for understanding and manipulating metabolic phenotypes in plants (Ratcliffe and Shachar-Hill 2006) and microbes (Feng et al. 2009). By analyzing the mass spectra of different fragmentations in proteogenic amino acids derived from various pathways, the active fluxes of intracellular biosynthetic pathways can be determined. Similar approaches allow estimates for small metabolites but are based on static measurements of low concentrations. While this approach is reasonably well established for microorganisms, several challenges must be addressed in applications for plants. These include (1) improving data collection by conducting more-sensitive NMR and MS analyses in conjunction with new subcellular fractionation approaches, (2) accelerating flux analysis so the method can be high throughput, (3) dealing with dynamic range issues, and (4) developing fully predictive metabolic network models rather than simply generating flux maps for the specific system and condition of interest (Kruger and Ratcliffe 2009).

Global characterization techniques are poorly developed for several molecular classes. For example, globally elucidating the identity and structure of unknown carbohydrates and glycosylated proteins (glycomics) is currently impossible, and even though technologies now allow lipid characterizations, they do not provide comprehensive information on lipid content. Therefore, developing global characterization technologies for these classes of molecules is important. Lipids and carbohydrates play crucial roles in cell-environment interactions and are essential components of membranes, driving the complex chemistry occurring within and near them. Without understanding the changing lipids and carbohydrates in a membrane, our knowledge and progress will lag.

3.1.5 From Snapshot to Spatial and Temporal Sample Series

Due to the high cost of molecular characterization, the majority of functional genomic studies to date have been snapshots of microbial communities. Just as watching a film is more informative than looking at a single frame from that film, kinetic series of molecular data hold great promise for advancing the understanding of biological systems. Because of the falling cost and increasing throughput in sequencing, the production of temporal and spatial molecular data series is now becoming possible. An example of research being conducted in this area is community profiling that allows comparison of hundreds of microbial samples using conserved marker genes, notably ribosomal RNA genes (Tringe and Hugenholtz 2008). Kinetic data are particularly informative if the sampling time is short relative to the characteristic time of the process of interest. For fast processes, such as stress responses or metabolite turnover, producing a data series could require very rapid sampling, which necessitates technological

advances in sample processing and increases in sample throughput. Multiple samples also can provide information on spatial patterns of gene and protein expression, particularly if taken across conspicuous physicochemical gradients. Since increasing the resolution of spatial sampling decreases sample size, greater sensitivity is required for the global measurement technology being used. A significant advantage of genomics technologies for characterizing multiple samples is that the changing composition of a cell community can be inferred from the resulting data.

3.1.6 From Bulk to Finer-Scale Characterization

For all omic methods, a major challenge is fine-scale resolution. In characterizations of microbial communities, gram or milliliter quantities of an environmental sample typically are used as starting points. At the organism scale, this is akin to trying to determine the global function of a rainforest community by blending together species occupying the forest floor, understory, canopy, and emergent layers and then trying to piece together functionality based on the combined data. Analysis and interpretation could be greatly enhanced if cell populations could be separated into more-refined spatial niches (Bergeron et al. 2007), particularly across distinct physicochemical gradients (e.g., Kunin et al. 2008). Although microscopic imaging can help delineate meaningful spatial niches in microbial communities, it thus far has not been used extensively to guide molecular characterization.

The finest scale for community characterization is at the level of individual cells. Significant inroads have been made in applying molecular technologies at this scale, largely because of whole-genome amplification technologies. Unfortunately, this technique requires physically separating a cell from its ecosystem and destroying it to obtain its molecular information. Though these are acceptable technical caveats for DNA-level characterization, for molecules strongly modulated by the cellular environment (RNA, protein, and metabolites), the isolation and destruction processes could alter the observed molecules significantly. Changes in expression profiles can occur very quickly (often on the millisecond scale), and these fluxes thus are unlikely to be representative of the cell in its native setting. Although currently intractable, the ability to characterize and relate transcriptome, proteome, and metabolome fluxes in selected single cells *in terra* remains an extremely important area for technology development in the coming decade.

3.2 Identifying and Measuring Rare Events, Molecules, and Cells Within Complex, Heterogeneous Biological Systems

Many of the biological issues relevant to DOE science missions—such as contaminant fate and transport, bioenergy, carbon cycling and biosequestration, and low dose radiation responses—require understanding a complex web of interactions. These interactions and responses occur between biomolecules within cells, between individual cells, between populations of cells, and between communities of cells and their environment. In many instances, the interactions are associated with rare events or with numerically rare (low-abundance) molecules and cell types (see sidebar, The Value of Understanding Rare Events and Cells, p. 24). In cell biology, single-cell dynamic measurements have been used to detect rare signaling events with important biological consequences that likely would have been missed or averaged out using bulk measurements (Faley et al. 2008).

A current challenge involves determining how to investigate the physiology of rare microbes within populations that cannot be readily cultivated. New technologies are needed to quantitatively analyze the physiology and genomics of low-abundance cells. For example, state-of-the-art cell-sorting capabilities, using innovative microfluidics platforms, could be coupled with sensitive instrumentation to measure individual rare cells. Although in their infancy, single-cell measurements already are making significant contributions to microbial biology (Behrens et al. 2008; Chan et al. 2007; Yu et al. 2006; Elf, Li, and Xie 2007; Huang et al. 2007). Additional technical advances in this area undoubtedly would have tremendous scientific impact by enabling scientists to simultaneously determine the identity of cells and measure metabolites, transcripts, proteins, and signaling molecules at sufficient temporal resolution to capture their dynamics. For example, EL-FISH nanoSIMS can be used to simultaneously image physiological processes and phylogeny at single-cell resolution in microbial communities (see sidebar, Investigating Rare Microbes' Function and Phylogeny at the

The Value of Understanding Rare Events and Cells

Below are two examples of situations involving rare events or organisms with significant biological implications. The first is a rare event leading to DNA mutations and cancer, and the second is a low-abundance (rare) population of microbial cells. Though rare, the importance of both warrants further characterization.

In the first case, given the radiation biology paradigm of mutation driving malignancy, radiation exposure has a finite, dose-dependent probability of generating stochastic genetic change. The biological response to radiation, however, is both nonlinear and deterministic, and it is the interaction of these two types of responses that can produce cancer. Thus, a fundamental challenge involves identifying and tracking rare events (mutant cells) and characterizing their dynamic interactions with a heterogeneous environment. Breakthroughs could be achieved with new tools and technologies that would enable non- or minimally invasive interrogation of multiple critical events *in vivo,* leading to the ultimate goal of analyzing the evolution of interactions within an organism in real time. Also critical is the development of computational models that identify "watershed events" and mechanistically predict their functional outcome over long time scales.

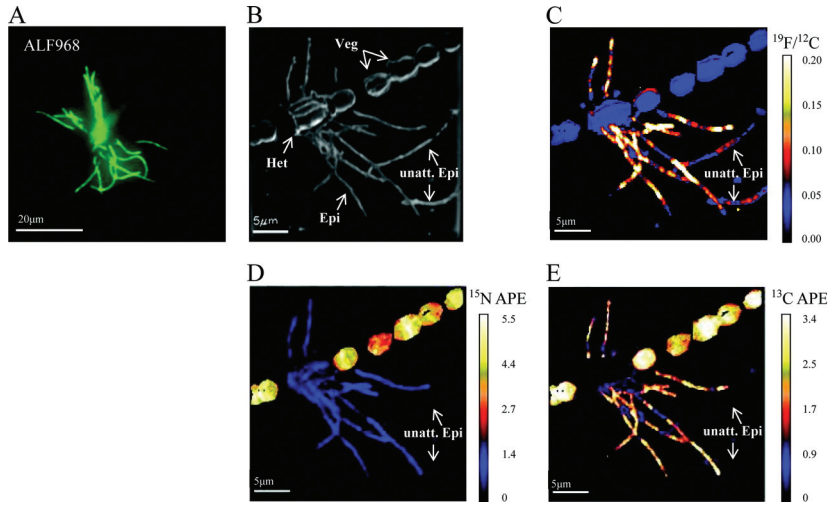
Second, low-abundance (rare) populations can account for significant amounts of phylogenetic diversity of microorganisms observed in many environments. By using massively parallel sequencing approaches, such populations have now been identified in a variety of environments, but their physiological and genomic characterization remains challenging. Although cell culture has enabled the characterization of some rare microbes and metagenomics is providing glimpses into their possible functions, the majority of these organisms are yet uncharacterized. New tools and approaches are needed to identify which populations are contributing to a particular process of interest. For environmentally relevant microorganisms, determining the distinctiveness of "isogenic" populations is important. Also critical is establishing the extent of variation (e.g., epigenetics, bistability, and persister cells) within the population to determine the mechanism of phenotypic heterogeneity within it. A poor understanding of the extent of genetic and physiologic diversity within populations prevents us from determining the number of individual cells that must be sampled and analyzed to gain insights into the functional diversity of complex populations, communities, and ecosystems. Current data suggest that such diversity is extensive in many environments.

Single-Cell Level Using EL-FISH NanoSIMs, p. 25). Ambient ionization mass spectrometry that involves ionization of untreated samples, such as microorganisms or plant and animal tissue in the ambient environment, also may be useful for resolving molecules and processes at the single-cell level. Moreover, identifying rare events at the single-cell level will require developing new tags and affinity reagents with high specificity and more-sensitive detection methods. Functionalized nanoparticles (e.g., TiO_2) show considerable promise for such applications (Rozhkova et al. 2009). In combination with new high-resolution imaging techniques like computed tomography, X-ray nanoprobes, and spin imaging, detection of rare events at the nanoscale can be achieved.

The effectiveness of environmental remediation can be dependent on the activities of rare, low-abundance organisms. For example, *Dehalococcoides* sp.—a low-abundance, slow-growing microbial species—is a primary agent in the remediation of chlorinated solvents. To enable the long-term management of slow-growing, rare populations in complex communities, we need the ability to measure their physiology and functional genomics in natural systems. We do not know whether the rare organisms found at environmental remediation sites are present because of recent adaptive evolution or because of sudden, favorable growth conditions due to the release of the environmental contaminant. To what extent is evolution (horizontal gene transfer, mutations, recombination) occurring *in situ* and resembling the

Investigating Rare Microbes' Function and Phylogeny at the Single-Cell Level Using EL-FISH NanoSIMS

Two fundamental technological challenges in biological characterization are identifying rare, uncultured microorganisms and determining their ecological function and their adaptive and evolutionary potential. Because they are rare, physiological and phylogenetic information for such microbes likely is limited. A crucial first step, therefore, is obtaining some insights into these microbes' ecological function in their current environment.



EL-FISH nanoSIMS Analysis of a Microbial Consortium. *Fluorescence and nanoSIMS images show a microbial consortium consisting of filamentous cyanobacteria (*Anabaena *sp. strain SSM-00) and α-proteobacteria (*Rhizobium *sp. strain WH2K) attached to heterocysts. Images taken after a 24-hour incubation with ^{13}C-bicarbonate and ^{15}N-dinitrogen.* (**A**) *Fluorescence image of the microbial consortium after EL-FISH with probe ALF968.* (**B**) *NanoSIMS secondary-electron image corresponding to panels* **C** *to* **E**. (**C**) *Localization of fluorine relative to carbon after EL-FISH with ALF968.* (**D**) *Distribution of ^{15}N-nitrogen enrichment.* (**E**) *Distribution of ^{13}C-carbon enrichment. Color bars indicate relative fluorine abundance* (**C**) *and isotope enrichment* (**D** *and* **E**) *in the image. [Key: Het, heterocyst; Veg, vegetative cell; Epi, epibiont; unatt Epi, Epibiont cells not attached to heterocysts. Image used with permission from the American Society of Microbiology. From Behrens et al. 2008.]*

Single-cell sorting by microfluidics and subsequent single-cell genome amplification and sequencing comprise a promising technology platform to study uncultured microbes. Currently, this method relies predominantly on visual selection of the cell of interest, which can be accomplished with fluorescence in situ hybridization (FISH). However, at present, no physiological or metabolic property can be considered in the cell-sorting and isolation process. Recent progress in nanoSIMS technology has enabled scientists to link microbial phylogeny to function while maintaining a structured environment. This technique, which combines nanoSIMS and elemental labeling–FISH (EL-FISH), might provide a basis for one path forward. NanoSIMS is emerging as a powerful technique to image biological systems' elemental composition, including stable isotope distribution, at 50-nanometer resolution. EL-FISH is an advancement of catalyzed reporter deposition (CARD)-FISH and exploits the chemical coupling of a specific element or isotope to a specific phylogenetic oligonucleotide probe; the isotope then can be subsequently visualized by nanoSIMS at high spatial resolution.

Thus, EL-FISH nanoSIMS simultaneously images physiological processes and phylogenetic identity at single-cell resolution within microbial communities while maintaining a structured environment. This technique has been used to identify the metabolic basis of a specific interaction of an α-proteobacterial epibiont with the heterocyst of an *Anabaena* species (see figure above). To functionally examine rare, individual microorganisms in a complex environment, advances are needed in nanoSIMS, EL-FISH, and EL-FISH nanoSIMS. Such advances should enable high-throughput, semiautomated analyses of microbial samples from diverse environments.

rate-limiting step for biotransformation? To answer these and other questions, new capabilities are required for detecting low-abundance microbes and genomes and for measuring their *in situ* activities—analogous to finding a needle in a hay stack. Tools also are needed for studying the biology of slow-growing microbes at the single-cell level and for taking rapid metabolomic and genomic measurements of microbial communities, ideally with high spatial and temporal resolutions. Imaging methodologies are the most powerful current technologies for identifying rare cell types *in situ*, but they still are limited in throughput and the types of parameters they can detect. Improvements in the speed, sensitivity, and automation of cell imaging are urgently needed, as are significant advancements in detection technologies. If available, these tools would greatly facilitate the monitoring of *in situ* ecological and evolutionary dynamics and provide essential data for constructing predictive models.

To develop a predictive, mechanistic understanding of plant and microbial processes using systems biology approaches, performing noninvasive or minimally invasive molecular measurements within individual, intact cells is critical. Even within clonal populations of *Escherichia coli* grown under highly controlled cultivation conditions, there is considerable cell-to-cell heterogeneity in gene expression because of both the inherent stochasticity in expression and the fluctuations in other cell components (Elowitz et al. 2002; Choi et al. 2008; see sidebar, Single-Molecule Measurements Reveal Phenotype Switching in Genetically Identical Cells, p. 27). High-throughput analytical procedures currently use extracts from large populations of cells to measure proteins, mRNA, metabolites, signaling molecules, and other types of biomolecules in a population of cells. These approaches provide, at best, an average over the entire population and thus are unlikely to detect rare events within a population or different responses of subpopulations of cells. Nevertheless, understanding the role of stochastic events is a challenge that must be addressed.

Another related challenge is the observation of low-level events within individual cells. For example, receptors within many cell types operate effectively when only a small fraction of them are activated. This sensitivity allows cells to detect changes in only a few molecules in the environment. Although initial detection events might generate extremely small biochemical signals, they eventually are amplified by the cellular machinery into physiological responses. However, unless initial detection events are observable, understanding the resultant response mechanisms can be extremely difficult. The biochemical processes regulating many cellular responses, such as transcription, chemoattraction, and stress adaptation, are usually below the limit of detectability using current technologies. Urgent needs, therefore, include novel approaches for amplifying the biochemical signature of these processes *in situ*, coupled to more-sensitive detection technologies.

3.3 Seeing It All: Simultaneous Measurements Across Multiple Dimensions

Addressing the critical challenges faced by genomics-based science will require continued progress in and new approaches to applying and developing analytical technologies. As highlighted earlier, identifying and tracking molecular species and processes within and between cells present unique measurement challenges because of the multitude of complex interactions and dimensions involved. Not only is there a large diversity in the types of molecules present, the spatial and temporal dynamics of these molecules often dictate their function.

Molecular interactions can extend over many orders of magnitude in length scales—from the subcellular (nanometers) to the cellular (microns) to biofilms and plant leaves and roots (milli- to centimeter) and even to larger organisms and ecological communities (meter to tens of meters). The relevant time scales of these interactions also can extend from milliseconds for enzymatic reactions to the multiyear responses seen in carcinogenesis or microbial community evolution. Investigating systems-level processes requires decisions on the types of information to be collected as well as the granularity of sampling in time and space. Enormous difficulties lie in extending the global characterization approaches described in Section 3.1, p. 19, to the tracking of molecules across many cells, across interfaces to disparate materials and environments, and throughout an ecosystem. However, an important objective is to understand how microscale

Single-Molecule Measurements Reveal Phenotype Switching in Genetically Identical Cells*

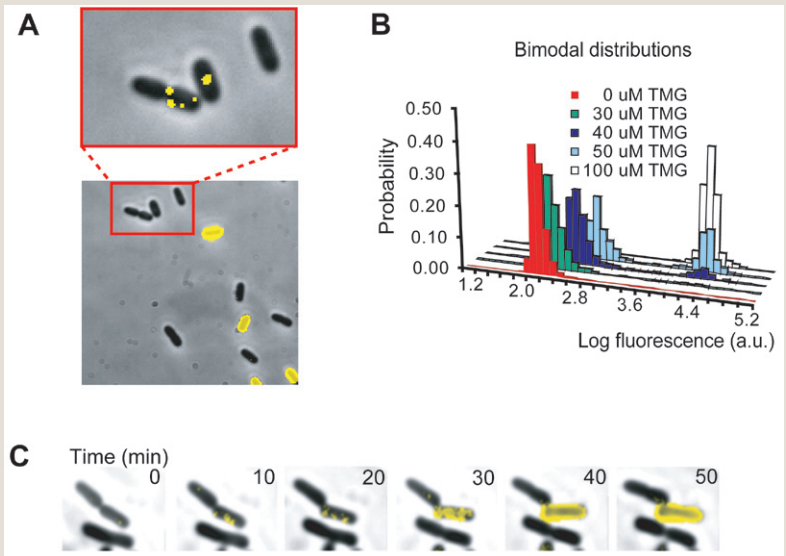
Genetically identical cells in the same environment can exhibit different phenotypes, and a single cell can switch between distinct phenotypes in a stochastic manner. In the classic example of lactose metabolism in *Escherichia coli*, the lac genes are fully expressed for every cell in a population under high extracellular concentrations of inducers, such as the lactose analog TMG. However, at moderate inducer concentrations, the lac genes are highly expressed in only a fraction of a population, which may confer a fitness advantage for the entire population. Recent advances in single-molecule imaging have revealed the detailed molecular mechanism controlling the stochastic phenotype switching of a single cell.

Lactose metabolism is controlled by the lac operon, which consists of three genes including lactose permease. Expression of the operon is regulated by the lactose repressor that dissociates from its specific binding sequences of DNA, the lac operators, in the presence of the inducer to allow transcription. The production of the permease increases inducer influx, resulting in positive feedback on expression of the lac operon. When the permease is labeled with a yellow fluorescent protein, two different phenotypes are observed in an isogenic population of cells, with the fluorescence intensity histogram of the cells exhibiting a bimodal distribution (see A and B portions of figure at right). Interestingly, the low-fluorescence cells have a few individual permease molecules, suggesting that one permease molecule is not enough to induce the transition from the low- to high-fluorescence phenotype. Scientists have determined that 200 to 300 permease molecules, corresponding to a big burst of gene expression, are needed for this transition.

The tetrameric lactose repressor simultaneously can bind to two operators to form a DNA loop. Under low inducer concentrations, the repressor cannot be pulled off the DNA by the inducer. Rather, spontaneous, partial dissociations of the repressor result in transcription of one mRNA and a small burst of proteins. However, infrequent events of complete dissociation of the repressor result in large bursts of permease expression (200 to 300 copies) that trigger induction of the lac operon. This change in phenotype is the result of the stochastic, full dissociation of the tetrameric repressor from all of its binding sites. The time-lapse sequence (see C in figure) captures such a phenotype-switching event. This illustrates that a stochastic, single-molecule event of the complete dissociation of the tetrameric repressor determines a cell's phenotype, and that a DNA loop is crucial in maintaining phenotype stability. This is a clear example that a rare event of a single molecule can have significant biological consequences.

---

*From Choi et al. 2008. Portions of text and figure reprinted with permission from AAAS.



Expression of Lactose Permease in *E. coli*. **(A)** *In the presence of moderate amounts of inducer (lactose analog TMG), an* E. coli *strain expressing lactose permease fused to yellow fluorescent protein (LacY-YFP) exhibits two phenotypes, all-or-none fluorescence in a fluorescence-phase contrast overlay image. Fluorescence imaging with high sensitivity reveals single molecules of LacY-YFP in the uninduced cells.* **(B)** *Bimodal fluorescence distributions show that the cells exist either in an uninduced or induced state, with the relative fractions depending on the inducer concentration.* **(C)** *A time-lapse sequence captures a phenotype-switching event. One cell switches phenotype to express many LacY-YFP molecules (yellow fluorescence overlay), but the other cells do not.*

biological and chemical interactions control and influence the dynamics of ecosystem composition on the macroscale. To realize this goal, characterization tools must be able to measure and monitor an array of molecular details across diverse length and time scales.

Approaches are needed that combine multiple techniques to map structural, chemical, and physiological changes in biological systems as functions of external stress and environmental conditions. For example, coupling multiple spectroscopic approaches can enable characterization of responses to environmental stress, including elevated temperatures, pressures, and chemical treatment. Measurements using optical, X-ray, infrared, and neutron scattering can enable structural, functional, and chemical analysis of molecular complexes and assemblies under differing environmental conditions. Such a multiplexed measurement platform could allow the tracking, visualization, and interpretation of real-time changes in microstructure assembly and organization—initially *in vitro*, then eventually *in situ* and *in vivo*. Also needed are new techniques for interrogating biological interactions on time and distance scales ranging from the level of biomolecular interactions (nanometer and picosecond scales) to intercellular communication (millimeter to centimeter and second to minute scales). To achieve these ambitious goals, advances are needed in imaging techniques, chemical probes (contrasting agents), and computational methods.

The ability to *see it all* requires characterizing the distribution of specific proteins, nucleotide sequences, lipids, polysaccharides, and other cellular components (see Fig. 3.2. Examples of Characterization Techniques and Their Ranges of Resolving Power, p. 29). Several promising advances include recent light microscopy techniques allowing the identification of microbial proteins (see sidebar, Super-Resolution Light Microscopies Enable the Imaging, Counting, and Localization of Single Proteins in Microbes, p. 30). At the most fundamental level, knowing the *elemental composition* of individual cells is important. Also critical is knowing the relevant chemical form over time. These types of information remain difficult to obtain even in controlled laboratory situations. The most general ways to assess elemental distributions in samples are X-ray absorption or fluorescence and in the case of imaging modalities, X-ray microscopy or tomography. These techniques can be used to characterize the distribution of nearly every element, from beryllium to plutonium. If the chemistry of particular elements is important, then different spectral techniques can provide this information. These technologies typically require the brightest sources of tunable X-rays, and thus synchrotron radiation sources are where advances in these approaches are likely to be made. Extending such measurements to field studies presents a considerable challenge.

One of the most powerful ways to characterize a complex biological system is to acquire multiple types of information about cell functioning at different spatial and temporal scales. Biological systems have evolved elegant pathways, molecular structures, and complex machinery at the nanoscale to power sophisticated cell functioning. However, current tools usually do not measure these entities directly but infer their presence using prior knowledge of their presence. Manipulating a living organism's biomolecules, supramolecular entities, and cell metabolites can provide important insights into their complex functioning. These types of manipulations also require extensive prior knowledge about the physical and chemical properties of molecules and assemblies within the cell. Thus, investigations of complex systems usually need to be preceded by extensive characterization. Development of unbiased analytical approaches that do not require prior knowledge of cell composition or function would greatly accelerate the investigation of previously unknown and uncharacterized biological systems.

The most important aspects of analytical measurement systems are those related to information content and resolution (whether spatial, chemical, or temporal) and those related to sensitivity, detection limit, and dynamic range. Unfortunately, these two important categories usually are in opposition to each other. For example, spatial resolution comes at the expense of temporal range and vice versa. As a result, imaging tools that are modified to work at smaller-length scales gather less chemical information or sacrifice dynamic range. This frustrates the study of important events, such as the cellular interface, which require sensitivity and resolution at the molecular scale but also a spatial range that allows coverage within a large

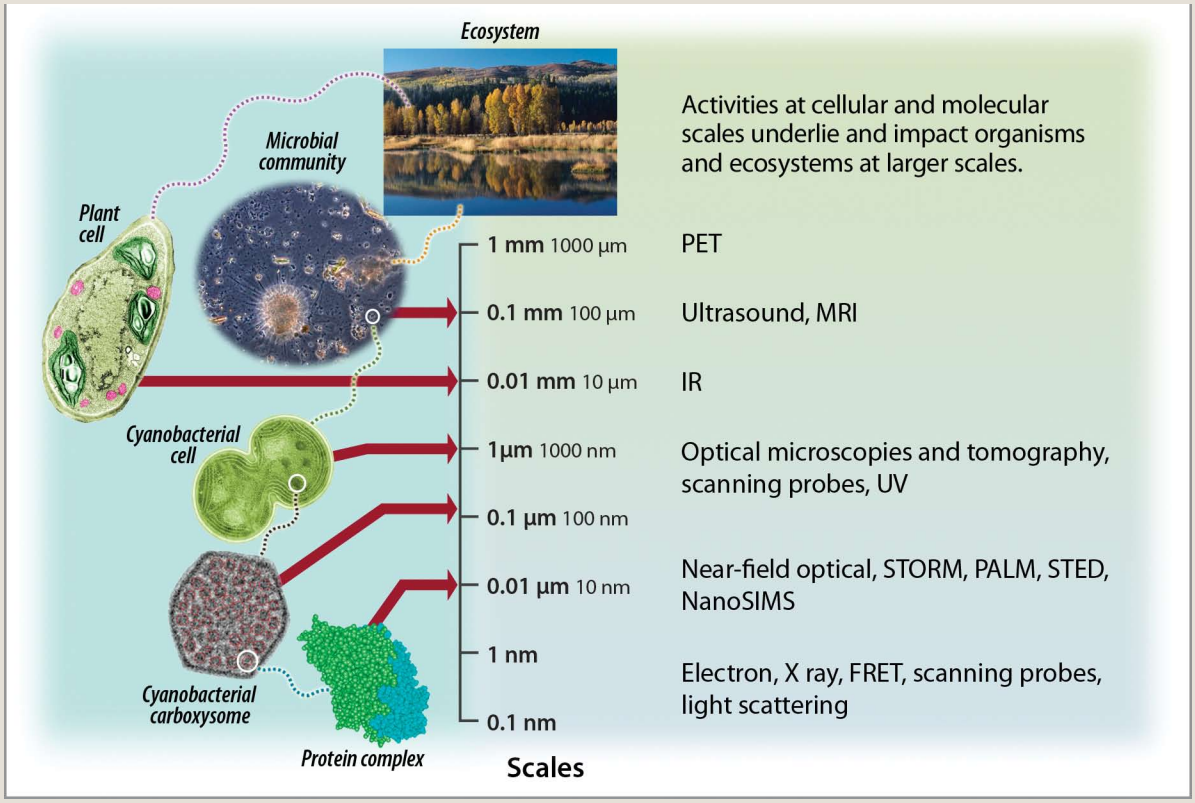


Fig. 3.2. Examples of Characterization Techniques and Their Ranges of Resolving Power. *The interactions of different chemical and biological species must be measured and understood across diverse spatial scales—from the environment and community levels to the cellular, subcellular, and molecular levels. Shown in the figure is a sample of different functional levels and characteristic sizes of representative components. Also listed are some of the analytical measurement tools that operate at these different length scales (also see sidebar, p. 30). A major need is to connect measurements and information at these different length scales. The types and variety of information involved complicate facile collection of relevant data, as does the need to monitor temporal changes in these systems. Typically, current tools are appropriate for providing characterizations at only specific spatial and temporal scales, or they are limited in the number or type of information that can be measured. Multiple "dimensions" need to be added to biological measurements so that molecular events can be linked to cellular, multicellular, and environmental scales.*

area. Illustrating this challenge is the light microscope through which observing a wide area requires a low-resolution objective with a considerable field of view. Detailed studies require a high-power objective, which intrinsically has a narrow field of view. Although one cannot simultaneously have it both ways, multiple objectives can be used sequentially, followed by multiscale integration. Alternatively, a plethora of measurements can be made using a high-powered objective followed by reconstruction of a wide field of view by image fusion. However, analyzing complex systems by sequential analysis creates an additional challenge with respect to observing living systems in real time. Perturbations to the system by measurement technologies are potentially disruptive to system functions and can generate unanticipated outcomes and prevent sequential analyses. Therefore, measurement technologies should be noninvasive or minimally invasive, allowing the system under study to progress and then be observed over time. For example, such techniques would benefit investigations of molecule and solute transport in plant cells, which are extremely sensitive to environmental perturbations (see sidebar, Challenges for Imaging Live Plant Cells, p. 31).

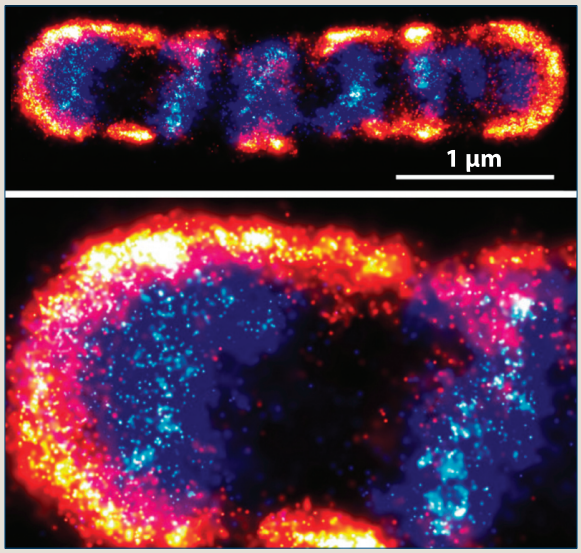
Super-Resolution Light Microscopies Enable the Imaging, Counting, and Localization of Single Proteins in Microbes

Over the past 3 years, a revolution in light microscopy has emerged with the potential to transform our understanding of microbial function. Prior to 2006, light microscopy largely was restricted to resolutions of about 200 nanometers by the diffraction limit, although some technologies already had begun to breach this limit, including stimulated emission depletion (STED; Hell 2003; Shaner, Patterson, and Davidson 2007) and structured-illumination microscopy (Gustafsson 2005). This meant the main tools for studying the mesoscopic (2 to 200 nanometers) organization of microbes were electron microscopy (EM) and atomic force microscopy (AFM). In 2006, three new techniques were reported: photoactivated localization microscopy (PALM; Betzing et al. 2006), fluorescence PALM (Hess, Girirajan, and Mason 2006), and stochastic optical reconstruction microscopy (STORM; Rust, Bates, and Zhuang 2006). These techniques allow single proteins to be visualized, counted, and localized with nanometer precision even when they are very densely packed (tens of thousands of tagged proteins per square micron). PALM and STORM—especially when used in conjunction with STED, EM, and AFM—may allow creation of comprehensive, nanometer-precision atlases of prokaryotic proteins and protein complexes.

In PALM, target proteins are genetically labeled with photoactivatable proteins, thus rendering them non-fluorescent until activated by near-UV light. By using near-UV light of sufficiently low intensity, only one protein per diffraction-limited region (~250 nanometers) is activated at a time. Following activation, each protein is then excited and imaged. Since only one protein is imaged at a time in each diffraction-limited region, the center of each molecular point spread function indicates the location of each protein. Serial cycles of activation and excitation are repeated until all fusion proteins are bleached. Because individual proteins are imaged, these techniques enable researchers to count the proteins and computationally assemble their locations into a composite, high-resolution image. The location of each protein can be determined to a precision of 2 to 25 nanometers, or ~10 to 100 times better than the diffraction limit. The localization error for each protein depends on the number of photons collected for that protein as well as background noise, pixel size, sample drift, and whether cells are live or chemically fixed. The highest optical resolution is achieved with chemically fixed cells, such as was done to obtain the figure below, which shows two views of a single *Escherichia coli* bacterium, wherein thousands of single chemotaxis receptor proteins have been "magnified" about one million times. PALM also can be used to follow the spatial and temporal dynamics of proteins in living cells (e.g., Shroff et al. 2008), although with very limited temporal resolution. Finally, PALM and STORM are able to image relatively deep inside tissues (Huang et al. 2008) using methods such as two-photon temporal focusing (Vaziri et al. 2008).

PALM combines high precision with the high specificity of genetically encoded fluorophores, whereas STORM allows the use of conventional organic fluorophores coupled to antibodies or to synaptosome-associated protein moieties. Such capabilities are important in systems that are either genetically intractable or anaerobic, characteristics that typically prevent genetically encoded fluorophores from fluorescing.

See also: Bahatyrova et al. 2004; Sener et al. 2007; Kirchhoff, Mukherjee, and Galla 2002.



Comprehensive Imaging of Densely Packed Transmembrane Proteins Using PALM. *Tar chemotaxis receptors fused to a photoactivatable protein, monomer Eos, expressed in* E. coli. *Proteins displayed in blue represent those studied in a total internal reflection fluorescence (TIRF) geometry, and those displayed in red were studied in an epi geometry. [Images from Greenfield et al. 2009.]*

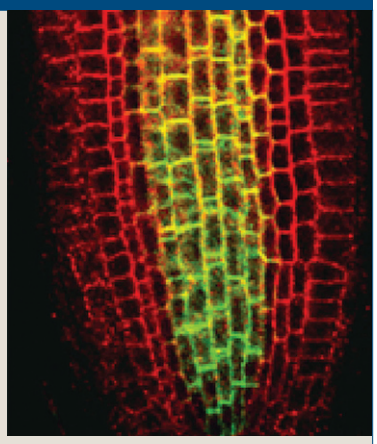
Challenges for Imaging Live Plant Cells

Transport in plant cells encompasses various types of cellular activity—from the shifting of molecules and solutes to the movement of proteins, lipids, and carbohydrates to the constant motion of intact organelles caused by several forces, including cytoplasmic streaming. Imaging this transport is a fascinating yet challenging endeavor requiring advancements in existing techniques and the development of new capabilities.

Technologies for visualizing both intracellular and intercellular transport are needed because plant cells are connected by specialized channels called plasmodesmata that pass through cell walls. To perceive and respond to external signals and to regulate growth, plants use hormones and solutes that require finely tuned, short-range transport into and out of cells and long-range transport among cells of several tissues. As sessile organisms, plants have evolved an extreme sensitivity to environmental cues and, in response, can subtly manipulate signaling networks and other cellular activities. Imaging technologies thus need to be noninvasive and nonperturbing to the cellular environment.

Although advances in X-ray imaging techniques have aided the study of ion and metabolite transport in plants, *in vivo* analyses could be facilitated by the development of protein-based nanosensors. Such sensors may enable us to determine cytosolic and subcellular metabolite levels in real time using fluorescence-based microscopy. Other techniques for examining metabolite transport and resource distribution throughout plants include positron emission tomography (PET) and nuclear magnetic resonance imaging (MRI).

Another major challenge is imaging fluid transport in intact plants, which is important for understanding the regulation of plant growth and stress response. Caged probes present a promising approach that could address this challenge and enhance the study of plant cell-to-cell communication. Using diffusion rather than microinjection, caged probes can be noninvasively preloaded into tissues where the caged molecules are activated in a spatially controlled manner. After photoactivation, movement of the uncaged fluorochrome can be followed in time and direction by fluorescence microscopy.



Fluorescent Imaging of Transport Proteins in *Arabidopsis* Root Tip. *Red and green indicate locations of two different types of auxin-transporting proteins. Yellow indicates where the two proteins work together. Auxin is a hormone that regulates the development of various plant tissues and structures important to bioenergy crop production and carbon partitioning in plants. [Image reproduced with permission of American Society of Plant Biologists © 2007. From Blakeslee et al. 2007.]*

The widespread availability of digital cameras and digital and optical imaging has improved the imaging of proteins and other important macromolecules, enabling the visualization of organelles or cellular domains that contain such molecules. Although plants present several unique challenges to imaging—background fluorescence and the effects of other parameters on the fluorescence signal (i.e., ionic strength, pH, and redox potential of organelles)—advanced imaging methods based on fluorescent protein technology have greatly enhanced study of the dynamic structure of plant cells (see figure, this page). This technology has overcome fundamental problems in plant imaging, yet challenges remain, especially for nonprotein molecules such as lipids. Because of the variation in cell membrane composition among kingdoms, probes developed for imaging lipids in nonplant species often are inadequate for plant-based investigations.

One of the major technological challenges in the next decade will be increasing the resolution of imaging techniques in a physiological context. Resolution is a critical factor for studies that must separate cell constituents in an unperturbed environment and determine the *in vivo* concentration of imaged molecules. A solution to such challenges may involve a combination of adequate statistics in single cells for analyzing noisy and partially correlated imaging data.

The problem of interrogating multicellular processes and systems is, *at least in part,* one of *molecular analysis of complex mixtures* (e.g., organics and biologicals). Next-generation analytical tools will tackle increasingly complex systems, enabling characterization of whole molecular machines and interacting protein networks. Most critically, these tools should push capabilities into the dynamic and kinetic regimes, allowing real-time analysis of interacting protein and macromolecular assemblies in solution, at interfaces, and across cellular and subcellular domains and membranes.

How do we achieve these capabilities? Significant progress could be made by developing approaches to measure selected cellular responses. The greater the number of cellular responses and dimensions that can be measured simultaneously, the greater the specificity of the model that can be built. Currently, the most powerful (e.g., sensitive, specific, and diverse) methods to measure cellular responses are biochemical and molecular assays. However, these types of assays typically are practical to use only with populations of cells, and the specific parameters being measured need to be known in advance, characteristics that prevent them from being global or unbiased in nature. Assuming that all cells in homogeneous populations respond identically is inherently problematic but is unsupportable for heterogeneous populations. Building realistic models of cell interactions thus requires the ability to either measure cellular responses at the single-cell level in communities or to know how to interpret population-based measurements in terms of the responses of individual cell types.

For example, understanding how a microbial community breaks down complex organic material requires knowing (1) the effective substrate for each cell type, (2) the products each produces, (3) which cell in turn uses those products, (4) and the spatial relationship among the cells. To follow this material flow, the concentrations and fluxes of different substrates on the cell population must be mapped and the ability of each cell to both use and produce different substrates must be identified. To accomplish this, we need to know how to build multiscale models of cellular communities that provide frameworks for mapping information derived from the population. The model should be able to represent both intercellular and intracellular networks.

3.4 Integrating Information

Living systems are extremely complex, and thus investigating their properties requires multiple experimental tools. Integrating diverse types of data generated by these different tools represents a major challenge in efforts to understand and predict biological function. Further complicating this challenge is the advent of high-throughput technologies, such as transcriptional profiling and MS–based proteomics, that can generate enormous amounts of data. Storing and managing these high-throughput data continue to be significant issues, but a more pressing and difficult problem involves interpreting and integrating the data with current biological knowledge. Much of this problem arises because knowledge in a given field is dispersed among individual investigators or found within a generally unstructured literature. Biology, unlike mathematics, lacks a formal language to describe and codify biological understanding and relationships. Therefore, most of our knowledge of biological systems is descriptive and qualitative in nature—a level of understanding insufficient for building predictive models of biological processes. Effective approaches for data integration and analysis need to be developed to successfully exploit biological systems to address energy and environmental challenges. DOE's Genomic Science program is beginning to address these needs with its Systems Biology Knowledgebase (U.S. DOE 2009; see Fig. 3.3. Phases in Knowledgebase Development and Functionality, p. 33).

Future analytical technologies for characterizing biological systems will generate additional data requiring further integration and analysis. Moreover, because biology is becoming an increasingly distributed science in which collaborative studies are conducted at universities and institutions across the world, data standards are needed to facilitate information exchange and integration. As the volume of research data continues to explode, bringing together the appropriate experts to solve a systems-level problem will become more difficult. For example, large-scale projects might involve dozens or hundreds of laboratories. Making the data of interest available to each lab is impractical. Instead, data storage, retrieval, and processing likely will occur at a few sites, with domain

specialists interacting with and interpreting data in a distributed fashion. As an example, genomic and very large scale proteomic datasets from many individual projects were effectively integrated at DOE's Environmental Molecular Sciences Laboratory to provide insight into the "core proteome" of bacterial species (see sidebar, Integration of Proteomics and Genomics Reveals a Core Proteome, p. 34).

To facilitate distributed biological research, several infrastructural problems must be solved, such as capturing experimental data and metadata and storing them in standard formats. Unfortunately, universal standards for biological data are unlikely to be implemented in the near future, so establishing general methods for translating data among multiple sources will be essential. Also needed are biologist-friendly software tools that handle heterogeneous datasets and support data analysis, simulation, concept building, and multi-investigator collaborations. A few such programs (e.g., Cytoscape) already are available on the web for analyzing genomic and protein-interaction data, but tools for integrating multiple data types are lacking. In particular, there is an urgent need for software that will integrate molecular, chemical, and structural information with spatial, temporal, and functional information. Providing scientists with training and support for any newly developed software tools also will be important.

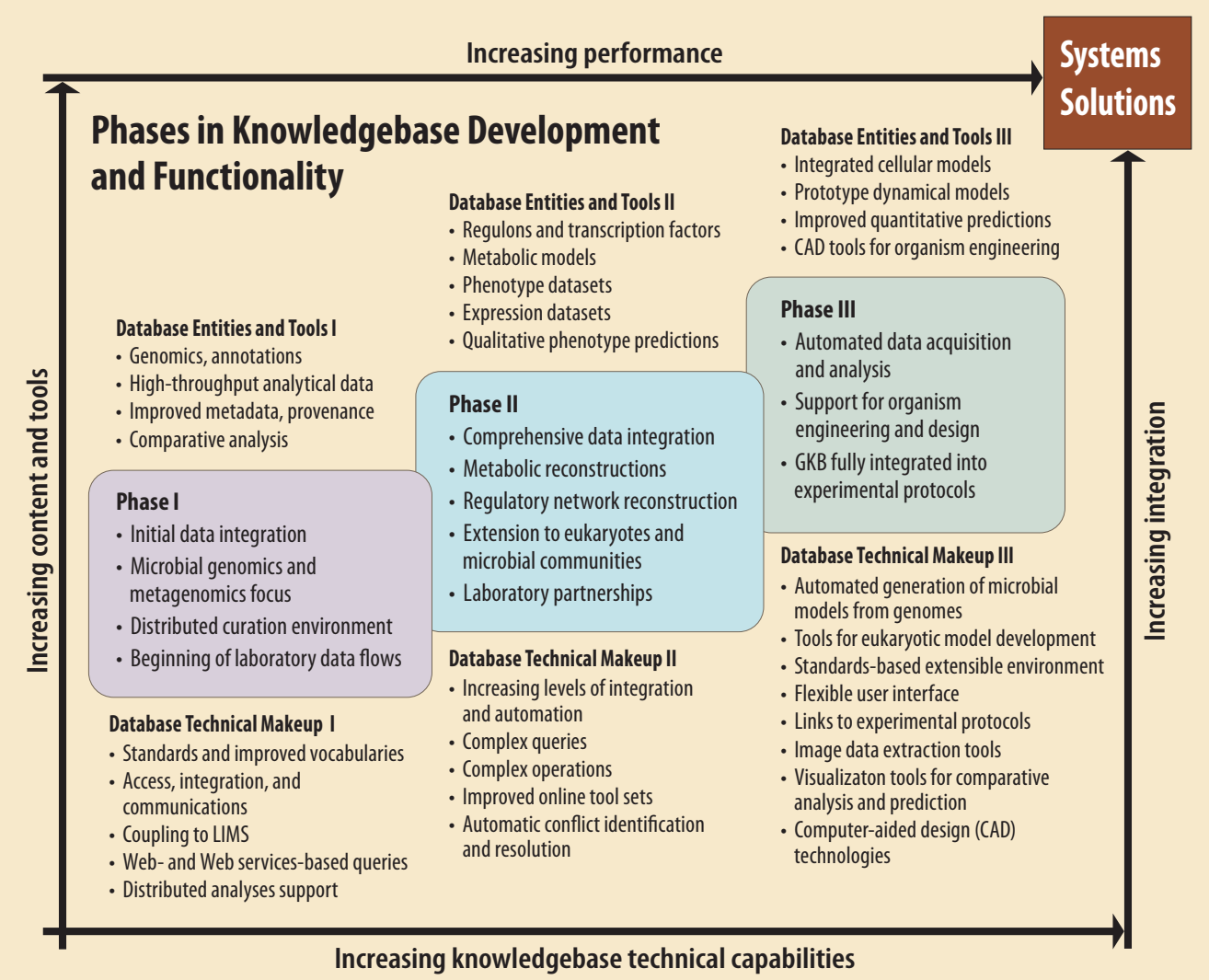
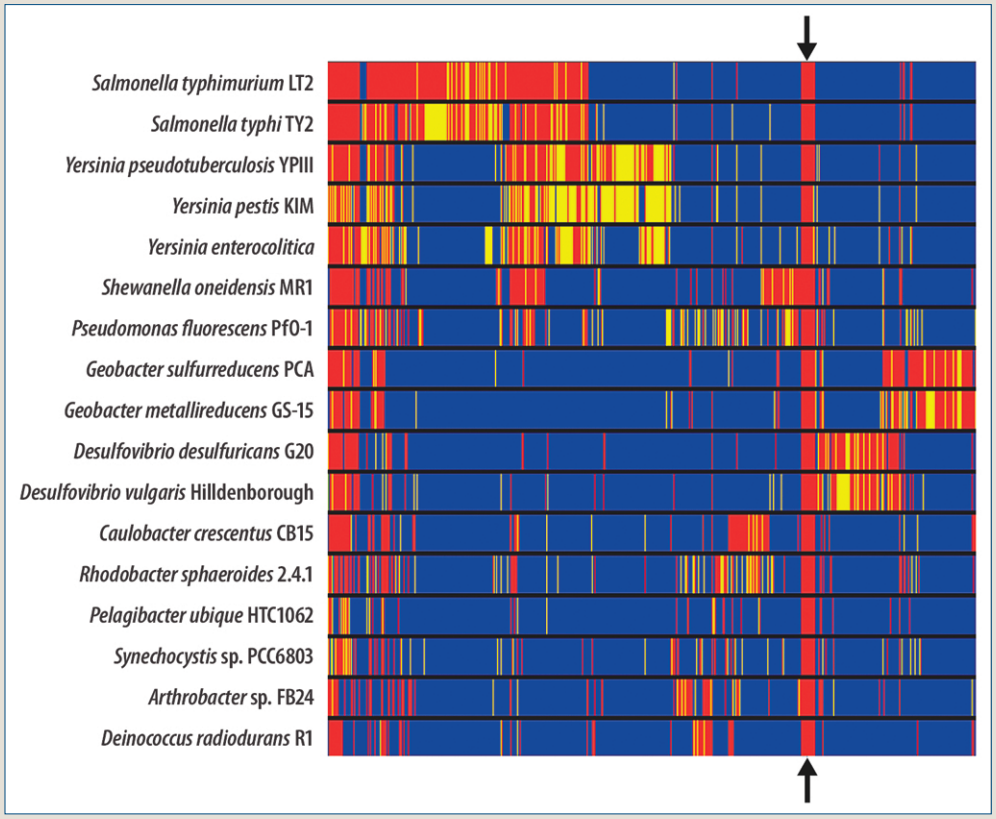


Fig. 3.3. Phases in Knowledgebase Development and Functionality.

Integration of Proteomics and Genomics Reveals a Core Proteome

Investigation of a universal set of proteins expressed among 17 environmental and pathogenic bacteria was enabled by a proteome database encompassing about 967,000 experimentally determined unique peptides linked to specific protein information and genome sequences from the DOE Joint Genome Institute and other data sources (Callister et al. 2008). Bacteria selected for this investigation included the metabolically diverse organisms *Rhodobacter sphaeroides*, *Shewanella oneidensis*, and *Synechocystis* sp. PCC6803, as well as pathogens such as *Yersinia pestis* and *Salmonella typhimurium*. Facilitated by collaborations that have made samples available for proteomic analysis, this study represents the gathering and evaluation of proteomic measurements taken over the course of 6 years.

Genomic comparisons among the 17 bacteria predicted the existence of a core genome composed of 144 genes (see figure, this page). Proteins from 74% of these genes were observed within the database, with each protein identified by two unique peptides. A functional analysis revealed that a majority of core proteome proteins (~55%) have functions related to protein synthesis, not surprising considering that the ability of a bacterial cell to make proteins for cell maintenance and growth is a vital function. However, what did surprise researchers was the observation of proteins (~7%) having very little or no functional characterization (e.g., the iojap-like protein). This gene was predicted to be homologous across all 17 bacteria and also is found outside the bacterial domain. Yet, little is known about this protein's function in the bacterial cell (Galperin and Koonin 2004). The universal expression of these relatively uncharacterized proteins suggests they are critical for life. This study also demonstrates the novel insights that can arise from large-scale data integration studies.



The Core Proteome for 17 Bacteria. *Genomic comparisons identified genes common to two or more bacteria (orange), resulting in a core genome of 144 genes. Proteomic measurements were used to verify the existence of these genes as proteins (red), resulting in the identification of a core proteome (arrow).*

The ultimate aim of systems biology is creating realistic and predictive models, but representing inherent complexities is a major challenge. Though such models will greatly facilitate the useful manipulation and modification of biological systems, a realistic simulation of even the simplest cell at the molecular level is well beyond the capabilities of even the most sophisticated computer in the foreseeable future. Thus, some level of abstraction and simplification is necessary. Less clear is the optimal detail or "granularity" for modeling biological systems at different scales, although defining core pathways in cells has proven to be a powerful concept. For multicell systems, an effective strategy could be to first obtain information at the single-cell level and use network approaches to extrapolate single-cell responses to multiple cells. However, this approach would require understanding the sources of response heterogeneity in cell populations as well as the feedback systems that operate among the different cells. Defining these and other types of information needed for multiscale modeling is an essential prerequisite for determining the types of data that analytical technologies should gather.

There is unlikely to be a single "best" approach for modeling complex biological systems. Instead, the most effective methodologies probably will link analytical measurements and data analysis to models of a particular biological response. Building a predictive model requires the ability to measure effects and compare them to simulated results. This linkage between theory and experiment is technically difficult to achieve and has been most successful with relatively simple biological systems. As models increase in scale and complexity, however, their output will become correspondingly more complex. This will necessitate gathering and integrating even more types of data to test and validate these models. In the future, therefore, the types of data gathered on biological systems will be driven increasingly by those required for predictive modeling.

3.4.1 Data Integration and Predictive Modeling

One of the greatest challenges over the next decade will be integrating disparate data types (e.g., molecular and image) to extract quantitative predictive models of ecosystem function (McMahon et al. 2007). Classical modeling of metabolic reactions typically includes reaction kinetics that become untenable for global characterization of the thousands of reactions occurring in a single species, let alone in a community of organisms. One approach for bypassing this bottleneck is to ignore individual rate kinetics and conduct a so-called metabolic flux analysis (MFA), in which the overall flux of a system is modeled (Tang et al. 2008; Price et al. 2003; Llaneras and Picó 2008). MFA does not require kinetic parameters for the reactions involved and can be scaled to deal with complete genomes and, potentially, with metagenomes. A particular kind of MFA is flux balance analysis (FBA). It has been shown to provide accurate predictions of single phenotypic properties—such as growth and substrate-uptake rates—for pure cultures of, for example, *Escherichia coli* (Fong and Palsson 2004; Edwards, Ibarra, and Palsson 2001) and *Helicobacter pylori* (Schilling et al. 2002). Such global modeling applications still are in their infancy, especially for ecosystem modeling (Stolyar et al. 2007), and will require concerted research efforts and resources to mature.

Despite enormous advances in high-throughput methods, measuring all characteristics of all biological macromolecules in all cell types under all conditions at all time scales will not be possible. Thus, there is a need for further development and applications of machine-learning methods, broadly termed "active learning," in which cycles of data-driven predictive modeling are followed by estimation of which subsequent experiments would best improve the model.

3.4.2 Informatics for Image Data

Much of the current effort in bioinformatics is directed toward using gene sequence information as a common framework for integrating molecular-level data. Integrating imaging data with other data types presents an entirely different kind of challenge, especially when attempting to link cellular images with compositional data. Challenges stem both from the complexity of the images themselves and from the need to relate these images to a common reference framework.

While extensive progress has been made on automated analysis of some types of biological images (Glory and Murphy 2007; Peng 2008), there is an urgent need for improved methods for data mining of images, with

subsequent feature-detection capabilities for obtaining image statistics and measurements of properties such as cell size, cell-wall thickness, cell morphology, and label concentration. Unfortunately, the intrinsic variability of biological specimens complicates efforts to map spatial distributions observed in a single sample to a standard set of reference coordinates. New approaches are needed for linking images and their associated data at multiple scales using a type of co-registration scheme similar to a geographic information system (e.g., the Cell Centered Database, ccdb.ucsd.edu). A standard co-registration scheme could greatly facilitate the mapping of different cell types or cellular structures to data such as gene sequences, chemical compositions, environmental responses, and metabolic capabilities.

As in all data management efforts, the minimum metadata that should be associated with an image must be defined. Image data and associated metadata then need to be linked to experimental or systems-level data so that all data can be integrated with molecular-level characterization and image analysis and query. A framework allowing this type of data linkage across different scales and data types would greatly facilitate the ability to productively use imaging information in DOE's Systems Biology Knowledgebase (U.S. DOE 2009).

3.4.3 Modeling Scale

Since current computational resources do not capture the vast complexity of biology, defining an appropriate level of resolution and scale is critical for practical modeling of any biological process. Akin to quantum mechanical description at small scales and a mechanical description at larger scales is the importance of judging the level of granularity needed for understanding the phenomena being investigated. This determination directly impacts which computational resources or methods need to be employed: workstation, cluster computers, supercomputers, distributed computing, grid computing, lattice computations, or simple simulation tools such as Mathematica and Matlab.

Many biological systems will be associated with steep gradients in various important physical and chemical parameters (e.g., the oxygen gradient in a biofilm). This factor poses a significant challenge for solving continuum-based simulations but is less of an issue for stochastic simulations. However, except for the simplest systems, current stochastic simulations are very inefficient because they require an enormous number of simulation runs to adequately explore parameter space. For mesoscopic simulations, too many particle parameters preclude the use of stochastic simulations, but a continuum approach cannot be used either because of granularity and discontinuities (e.g., steep gradients). Nevertheless, connecting the microscale to the macroscale requires dealing with the mesoscale. New approaches, therefore, need to be developed, such as rule-based simulations or transfer function models. However, the adequacy of any new modeling approach must be defined in terms of its ability to capture and predict biological phenomena.

3.4.4 Heterogeneity

Heterogeneity—a fundamental aspect of living organisms and natural ecosystems—is critical to both the stability and evoluability of systems. One of at least three types of heterogeneity relates to physical aspects of the environment, such as the patchiness of mineralization or structure in substrates used as nutrient sources. At the systems level, another type of heterogeneity can arise from genetic variation in a population of organisms or from stochastic aspects of cellular functions (e.g., gene expression). This heterogeneity can be reflected by changes in the average behavior of cell populations or complex organisms. At the chemical level, variations in the structure of chemically identical molecules represent another kind of heterogeneity. This is the basis of a phenomenon in which multiple products can be synthesized from chemically identical substrates by enzymes that catalyze reactions at different rates. Accounting for heterogeneity introduces substantial complexity into models, but mechanistic models must include the basis of their behavior.

3.4.5 Validation of Multiscale Models

Building multiscale models requires linking and integrating models that operate at different temporal and spatial levels. Not only must existing methods for linking simulation schemes be improved to represent complex biological processes, but the sources of error contained in individual models must be identified as well. Representing rare events in complex models also is quite difficult because of the simplifications

used to make them computationally tractable. Thus, erroneous model results can arise from both incorrect structures and parameters, underscoring the need for methods to distinguish between the problems caused by each. Also needed is the ability to assess the sensitivity of the overall model to changes in characteristics of individual submodels.

To improve multiscale modeling of data, new and existing strategies for coupling micro- and macro-level models must be developed and tested. The information integration in these coupled, multiscale models possesses an inherent uncertainty that needs to be addressed by setting standards for data analysis. Uncertainty can be controlled by using numerical simulation techniques to approximate with quantitative error indicators on different scales—from isolated single-scale approaches to multiscale numerical simulation environments. This approach should take into account the inevitable and inherent errors in parameter estimates resulting from the limitations of instruments and analytical methodologies. Current model validation and verification efforts are constrained in part because of inadequate multiscale and upscaling methods. However, an even greater obstacle is the lack of realistic multiscale experimental models on which to test, calibrate, and validate the proposed methods. To calibrate any multiscale method, a model's measurable quantities must be assessed before a method can be established to define its datasets' hidden quantities. Once the models have been produced, specific assessment schemes will have to be developed to guarantee consistency of the resulting estimates.

Summary of Key Technological and Capability Challenges

3.1 Expanding Global Characterization Capabilities

- Advance microbial community genomic sequencing by increasing the coverage of more complex communities, analyzing metagenomic profiles over multiple time points, and developing single-cell sequencing and transcriptomics to investigate low-abundance organisms.
- Develop amplification methods for *de novo* cDNA sequencing and capabilities for detecting small regulatory RNAs.
- Provide more complete proteome characterizations by improving measurements of membrane and low-abundance proteins and by decreasing bias in protein coverage.
- Combine multiple characterization techniques (e.g., NMR and MS) to expand metabolome coverage, increase throughput for metabolic flux analysis, develop global measurements for lipids and carbohydrates, and address dynamic range and measurement sensitivity issues—especially for complex plant metabolomes.
- Obtain data on kinetics and changes in spatial patterns by processing smaller samples taken at multiple time points throughout the duration of particular biological activities.
- Concurrently measure kinetic data and chemical changes *in situ.*
- Enhance fine-scale resolution of omics methods to enable the ultimate goal of relating transcriptomic, proteomic, and metabolomic flux measurements for selected cells from natural environments.

3.2 Identifying and Measuring Rare Events, Molecules, and Cells Within Complex, Heterogeneous Biological Systems

- Improve the speed, sensitivity, and automation of cell imaging to enable the detection of rare cell types.
- Develop methods for detecting extremely small molecular and biochemical signals that regulate cellular responses *in situ*.

3.3 Seeing It All: Simultaneous Measurements Across Multiple Dimensions

- Design new multiplexed measurement platforms that combine multiple techniques for tracking, visualizing, and interpreting real-time changes in the structure, organization, and activities of biological systems spanning multiple scales in space (nanometer to kilometer) and time (milliseconds to years).
- Simultaneously conduct molecular-scale studies that focus on small areas within a larger scale system to obtain detailed molecular information spanning a larger spatial area (e.g., the cellular interface).

3.4 Integrating Information

- Understand the sources of heterogeneity in cell populations so that network approaches can be used to extrapolate single-cell responses to multiple cells.
- To overcome the bottleneck of dealing with individual rate kinetics for thousands of reactions, scale up metabolic flux analysis approaches to model overall biochemical flux for genomes and potentially metagenomes.
- Define the minimum metadata needed for each image type and develop a standard co-registration scheme for mapping images of molecular and cellular structures and cells to other data (e.g., DNA sequences, chemical composition, and environmental conditions).
- Automate image analysis and feature-detection capabilities that describe and measure the characteristics (e.g., cell size and cell-wall thickness) of objects and processes captured in each image.
- Improve the ability of continuum-based simulations to address steep gradients in biological systems (e.g., oxygen gradient in a biofilm).
- Increase the efficiency of stochastic simulations by establishing new methods that decrease the number of simulation runs needed to adequately explore parameter space.
- Develop and test new strategies for coupling micro- and macro-level models, understand the sources of error in individual models, and assess the sensitivity of the overall model to changes in individual submodels.

4. Path Forward

We now have an unprecedented opportunity to greatly expand the dimensions of biological measurements. Technologies must be challenged and boundaries pushed to enable characterization of multiple species within complex, heterogeneous cellular systems with increasing temporal and spatial resolution. Biology will be significantly advanced by integrating multiple measurements and placing them in the true physical and temporal context of the system. The major challenge, however, is measuring processes with minimal perturbation and in as natural a setting as possible (i.e., *in terra*, *in planta*, and *in vivo*). Meeting this challenge represents an exciting prospect to obtain a qualitatively different level of insight about complex biological processes.

Although still in its infancy relative to more-established sciences such as chemistry and physics, quantitative biology has advanced tremendously during the past decade, largely because of technological developments in DNA sequencing, proteomics, and other high-throughput genomics-driven technologies. Biologists, however, still lack a predictive understanding of cells, organisms, and communities—knowledge essential for their rational design and manipulation. These prediction and manipulation capabilities are challenged by the inherently complex, dynamic, and heterogeneous nature of living organisms. By making strategic investments in a new generation of technologies, our ability to understand complex biological systems will advance significantly (see box below). Such investments would

Examples of Needed Technology Advancements

Adding Dimensions to Biological Measurements To Capture Function

- Mass spectrometry approaches that provide both chemical *and* spatiotemporal information
- Single-cell metabolomic and proteomic measurements
- Microfluidic approaches for cell isolation and analysis
- Imaging approaches that provide both spatial *and* chemical information
- New nongenetic fluorescence probes and sets of multiplex probes
- Rapid, multiresolution, multidimensional imaging technologies

Expanding the Ability To Measure and Manipulate Biological Components

- Super-resolution optical spectroscopy at the nanometer scale
- Electrochemical imaging at the cellular level
- Novel isotope technologies including subcellular tracer studies
- Improved isolation and identification of gene products from natural samples
- Nuclear magnetic resonance to distinguish subtle biological molecular variation (e.g., among carbohydrates)
- Atomic force microscopy at the molecular scale for soft biology
- Higher-resolution electron microscopy to use with hydrated samples

Identifying Important Events in Heterogeneous Environments

- Secondary ion mass spectrometry at the nanoscale
- Nondestructive imaging for dynamics across long time scales
- Synchrotron-based approaches including infrared, X-ray fluorescence, and tomography for *in situ* analysis

Modeling Complex Systems

- Improved mechanistic models of intercellular and intracellular networks
- New approaches to modularize complex networks
- Multicellular models
- Hybrid models to link multiple temporal and spatial scales
- Models representing long-term outcomes of rare events

continue DOE's long history of facilitating important technological breakthroughs in the biological sciences and would improve the scientific foundation for solving the nation's significant energy and environmental challenges.

As detailed in Chapter 2, beginning on p. 5, workshop participants identified several critical biological challenges that would benefit greatly from investments in technologies to enable their timely achievement. With predictive biology as the ultimate goal, these technology investments should be guided by the need to answer specific scientific questions, not simply to stimulate the development of technology for technology's sake. Answering these questions should be of highest priority and will require integrated developments on a number of fronts.

Existing techniques can begin to address some of these issues, but almost all have limitations such as being static, able to measure only a single component, or restricted in spatial resolution. The challenge of simultaneously making global, dynamic, and multiscale measurements might be solved by improving these techniques or creating new ones. However, a single technology or related suite of technologies unlikely will be able to answer all of the complex, systems-level questions that need to be addressed. Thus, many different technical approaches should be explored simultaneously.

Technologies cannot be developed or validated in the absence of an appropriate experimental model, and it is unclear whether suitable experimental model systems currently exist for addressing all the questions discussed in this report. A reasonable path forward thus involves not only identifying technologies that will have the greatest impact on answering important biological questions, but also identifying or developing the most appropriate experimental models. Clearly needed are well-characterized and tractable model systems—on the organismal and multiorganismal scale—that more closely approximate natural settings and are relevant to DOE scientific mission areas. Development of such model systems is best done concurrently with technology development and application.

Collecting and integrating information that results from new technologies also will be critical for eventually predicting and manipulating biological function. If information can be effectively integrated, capability advancements in one technical area can synergize with those in others, potentially resulting in revolutionary progress in understanding complex biological systems.

4.1 Technology: Meeting the Measurement Challenge

The technologies needed to characterize a system ultimately will depend on the particulars of the system being examined and the research objectives being carried out. The broad technological challenges outlined in Chapter 3, beginning on p. 19, should be addressed in the context of specific model systems, with the aim of solving particular biological problems. As multicellular model systems are developed and natural systems more thoroughly characterized, technologies will need to be subsequently adapted and changed. Moreover, there are a number of important capabilities whose development is likely to be foundational to understanding complex biological systems and thus probably will require continuous improvement.

In the short term, development efforts should focus on adapting current technologies for tackling more-complex, heterogeneous systems and on adding dynamic measurements to static techniques. Such improvements can be accomplished most effectively by increasing the speed and resolution of current technologies. For example, the slow speed of switchable fluorescent probes and image acquisition currently restricts the use of subdiffraction-limited optical imaging to fixed samples. Similarly, the slow speed of proteomics measurements limits their use in kinetic studies. In the medium term (5 to 10 years), current technologies should be combined or enhanced to enable multimodal, multidimensional analyses. In the long term (10 years or more), analyses will need to be expanded to allow measurement of complex living systems and their surrounding environment at appropriate length and time scales. Making comprehensive biological and environmental measurements will not always be practical. Therefore, it is critical to define the key measurements necessary for understanding and predicting system behavior and to determine the required scale and density of information. In addition, technology developments should be closely tied

to improvements in data analysis and computational modeling to anticipate and deliver the needed measurement tools.

4.1.1 Adding Dimensions to Biological Measurements To Capture Function

A key finding from the workshop was the clear need for simultaneously assessing multiple molecular species with appropriate temporal and spatial resolution. In the near term, this will entail developing new labeling or probing technologies to spatially map the activity of multiple cellular processes *in situ*. Fluorescence-based labeling technologies and related nanoprobes were identified as critical areas in which immediate expansion is needed. In addition, microfluidic approaches for single-cell sorting and analysis should be pursued simultaneously with new genome amplification technologies to enable full genome sequencing of single cells. Highly sensitive mass spectrometry techniques also need to be developed for single-cell metabolite and proteomic measurements.

Capabilities for making multimodal measurements with spatial and temporal registration should be addressed in the near term, but significant technical challenges must be overcome to speed their development. For example, the current tradeoff between spatial and temporal measurements will hinder new developments until measurement rates can be increased appreciably. Emphasis thus is needed on developing faster detectors and brighter or more portable light sources that will enable new approaches to multiplexed measurements. Also needed are advanced nanoSIMS and other technologies for single-cell isotope imaging. Such advancements will increase the number of measurable parameters in complex systems and should be especially applicable to investigating multicellular systems.

Linking cellular, multicellular, and environmental scales is an ultimate goal requiring long-term investments in the development of approaches to simultaneously and noninvasively measure cellular states (e.g., the proteome, metabolome, and transcriptome). This linking also must be conducted for more-complex natural systems, necessitating the development of advanced imaging systems capable of interrogating large areas in three dimensions at high resolution. The extent of resolution needed for these technologies should be established by advanced modeling approaches addressing the particular biological question being pursued. Because of their widespread use and enabling nature, multiresolution and multidimensional imaging technologies should be priorities for continuous development and enhancement.

4.1.2 Expanding the Ability To Measure and Manipulate Biological Components

The types of studies that can be done with cells are restricted by limitations in what we can actually measure. Our ability to detect and quantify fundamental components such as proteins and nucleic acids is relatively advanced, but many varieties of different molecules are currently difficult to detect. Technologies need to be developed for completing the "parts list" of cellular components (e.g., metabolites and carbohydrates) that currently are invisible or poorly characterized. In the near term, the focus should be on expanding the number of different metabolites, ions, and well-defined products of enzymatic pathways that can be monitored. Especially important is detecting species that indicate the activity of cellular and intercellular pathways. Again, these technologies will be most informative if they can be applied at the level of the individual cell in a high-throughput manner.

In the midterm, measuring dynamics and, in particular, the flux of specific components within multicellular biological and natural systems will be critical. Approaches to image the spatial distribution and activity of less-characterized components also will be needed. A longer-term goal is developing general approaches for identifying complex extracellular molecules produced by cells or detecting the modification of the extracellular space by organisms. Such insight will be necessary to understand how cells dynamically alter their microenvironment and how this subsequently changes their physiological responses.

Simply cataloging components will be insufficient for understanding their function. Progress in this area will require developing approaches to specifically manipulate and measure the activity of these poorly characterized components. For example, chelators have been very useful for understanding the role of specific ions in cellular function, as have targeted inhibitors to

enzymes. Activity-based protein profiling techniques are being developed to characterize enzyme function, and such approaches should be extended to measure other components. Also needed are general techniques to inhibit or increase the availability or activity of poorly understood components, such as modified lipids. Manipulating the spatial distribution and activity of components would be a particularly powerful approach for understanding their physiological function in the normal cellular context. Experimental manipulation of complex system variables, at the molecular level, will be necessary for understanding the functional processes from the organismal to the environmental scale. A long-term goal is making such manipulations routine, while simultaneously measuring their effects on system processes.

4.1.3 Identifying Important Events in Heterogeneous Environments

Measuring rare events or minority components and relating them to functional outcomes are critical long-term research goals that first require an improved ability to detect such events and cell types. As described in Section 3.2, p. 23, a necessary foundation for achieving these objectives is advancements in single-molecule and single-cell measurement technologies. These technologies then need to be enhanced to identify and detect single or small populations of molecules or cells within complex heterogeneous backgrounds. Better measurements of single molecules or cells subsequently should improve our ability to detect small differences in a population's composition or activity.

While sensitive detection technologies will be critical, equally important is the need for these technologies to be high throughput and parallel. Measurements on a single cell can be insightful, but much greater value could be obtained from the comparative measurements of hundreds, thousands, or tens of thousands of cells. This comparative capability could allow rare events to be reliably identified. To understand the outcome of these rare events, however, we need to be able to follow the same population over long time periods using nondestructive methods. Simultaneous improvements in the development of model systems and analytical technologies should be directed toward achieving this aim.

4.1.4 Modeling Complex Systems

Understanding which measurements are most informative for characterizing a complex biological system will require significant advances in computational modeling. Deterministic, ordinary differential equation–based models are effective when applied to relatively simple biological systems, but they involve too many constraints, specified parameters, and computational resources to be generally useful for large-scale networks. Nonmechanistic approaches, such as Bayesian networks, lack the specificity needed to constrain experimental measurements. The disparity between the current scale of biological modeling and analytical technologies is one of the most severe bottlenecks in systems biology. However, applying systems biology to problems relevant to DOE will require developing scalable approaches that are applicable to multicellular systems.

In the near term, computational models should focus on describing physical relationships that can be experimentally measured (e.g., a model of a metabolic pathway should represent a physically measurable relationship). Likewise, parameters needed to constrain the model also should be measurable, such as protein abundance and stoichiometries. This will require modifying current models to align them to the experimental systems and biological challenges outlined in Chapter 2, beginning on p. 5. These models should focus on fundamental pathways and regulatory networks in cells. This level of representation is likely to reveal important and universal principles of cellular design.

In the midterm, new modeling approaches should be developed for representing characteristics of multicellular systems, such as the relationships between different cell types and the flux of energy and materials through the system. These capabilities likely will require improved mathematical representations, such as rule-based modeling and new computational approaches. Including heterogeneity in the models likely will be very challenging but nevertheless essential for accurately re-creating the behavior of cellular communities from knowledge of their individual members.

In the long term, we should strive to develop predictive models with an appropriate level of granularity. This will require coordinating analytical and modeling efforts and developing practical approaches for building

scalable models. Creating such models involves developing hybrid models or more-effective mathematical approaches for representing multidimensional models. The accuracy of the models should be improved to enable evaluation of the effects of rare events. Such advancement is likely to require far more powerful computers than currently available or novel approaches for distributed computing.

4.2 Research Strategy: Linking Observations from Cells to Natural Communities

Most traditional biological models were chosen for their convenience and simplicity rather than relevance for solving particular biological problems. Addressing many of the problems relevant to DOE's energy and environmental missions, however, will require understanding how biological systems function *in situ*. Such a goal is beyond our current technical capabilities in many cases, and thus appropriate experimental systems need to be established in the laboratory to serve as a bridge to reach this goal. This is particularly the case for microbial communities whose inherent complexity challenges efforts to understand the mechanisms of their interactions. Several of the technologies described in Chapter 3, beginning on p. 19, have the potential to address the challenges associated with natural biological systems, but unless investigators work toward defining and validating workable experimental models, the development of these technologies will lag.

Because they are relatively simple and well characterized, current experimental model systems are likely to be the starting points for any new technology development efforts. In some cases, relatively simple plant and microbial species relevant to various DOE missions already have been identified and initially characterized. For example, microalgal systems proposed as a platform for biofuel development also would be excellent systems for understanding basic cellular processes. The initial systems must be relatively simple so that they are tractable using current technologies, but they also should be suitable for investigating the questions posed in this report. It would be best if the scientific community identified a limited set of model organisms to facilitate collaborative efforts, but it is recognized that sometimes a particular biological problem is best approached by the use of a unique model system. For cases in which the choice of a particular strain or species of organism is not critical, every effort should be made to use DOE-relevant model systems and modify current technologies to work with them.

Tools for manipulating and probing current model organisms must be improved, but new classes of methodologies likely are needed to perform manipulative experiments on more-advanced model systems. There is an urgent need to develop these advanced systems, particularly those that involve the co-cultivation of multiple organisms. Ideally, these multiorganism systems should be capable of being spatially defined and should support the development of multidimensional analytical technologies. These technologies should be developed in conjunction with the experimental systems, and the limitations of both should be defined.

In the medium term, model biological systems need to transition toward those with increasing complexity representative of natural systems. This transition will require an improved understanding of natural biological systems and the factors controlling their stability. Needed are better approaches to identify and characterize the species in natural systems, as well as improvements made in defining their spatial and temporal organization and functional capabilities. The data obtained from natural systems then should be used as a basis for creating simplified models in the lab. Creating mimics of natural systems likely will require a new generation of growth chambers compatible with multiple analytical devices and probes. In the longer term, investigators should be able to conduct complex analytical measurements in the field.

Eventually, multicellular model systems will be used to explore a variety of different processes, such as energy flow through communities in a controlled environment. These model systems also should support the development of more-advanced technologies, especially multimodal approaches for simultaneously measuring intercellular and extracellular environments. A critical focus should be not only on measuring relevant characteristics, but also on precisely manipulating these characteristics to learn how information, materials, and energy flow through the system. Necessary capabilities will include selectively analyzing individual cells within

complex mixtures of organisms and within heterogeneous chemical and physical environments. Attaining these goals and realizing a predictive understanding of flux through these systems will require the iterative testing of models that, in turn, must be supported by effective collection and integration of measurement data.

The long-term goal is integrating information from model systems with data from the direct interrogation of natural communities *in situ*. Ideally, this should be possible at a field site, but some technologies are likely to be restricted to the laboratory for the foreseeable future. Thus, transplantation of natural systems into the laboratory setting likely will be necessary. This will require improvements in the controlled environmental chambers that can be used with various imaging techniques. These should move beyond state-of-the-art environmental chambers, such as those that are beginning to allow electron microscopy to be conducted with hydrated samples, toward those that include microfluidics to allow direct experimental manipulation during analysis.

4.3 Knowledge Visualization: Overcoming Challenges of Data Integration and Systems Analysis

Development of characterization technologies for multicellular systems should take into consideration the idea that such technologies must enable experimentation at a much larger scale than that of traditional biological research while retaining the necessary detail and resolution. As the focus shifts from characterizing the average behavior of populations to enumerating the specific behavior of their constituent members, the magnitude of data that must be gathered, processed, and understood will expand greatly. This information explosion must be anticipated and managed up front. The experience of the Internet has shown that data accessible from creation is far easier to manage and manipulate than data that reside in legacy repositories. Early integration between experimental biologists, technology developers, and computational biologists is essential.

Data volume also can be managed by restricting the scale of measurements to that which is needed to adequately reflect the biological system. This first will require advancements in computational modeling technologies to understand the level of granularity needed to represent a given biological process. Nevertheless, there should be some early effort to define "recommended scales of measurements for biological systems" that could be used as guidelines or references for the types and amounts of measurements needed to predict, model, and quantify specific systems.

Even if efforts are made to restrict data volume, the amount that will be gathered and managed is still likely to be enormous. Multimodal data that will be generated over the next few years will present an entirely new challenge. Thus, necessary in the medium term will be development of tools for integrating and interpreting complex datasets. The ultimate aim of systems biology is predictive modeling of biological systems, which will require an iterative cycle of experiment-simulation-experiment. Because of the necessity of comparing the results of the models with those of the experiments, measurement technologies need to be developed in parallel with the models so that outputs of both approaches are comparable. This will ensure the generation of useful databases and computational approaches for integrating measurements and models at multiple scales.

Another important medium-term goal should be defining the modularity, or systems-level rules that could be inferred using novel approaches for analyzing complex datasets. This first will require the storage and accessibility of the data, as described above. Making sense of complex datasets also will require new approaches for visualizing multidimensional data. Attaining these visualization capabilities will require investing in the development of computational tools that enable the integrated graphic display of multidimensional cellular parameters (e.g., genome, proteome, metabolome, transcriptome, and bibliome). Such advances need to be coordinated with the development of new analytical tools because of the current paucity of ways to productively query and comparatively display such data. Automated or semiautomated data mining likely will play an important role as well.

In the long term, sophisticated computational tools should be able to take data acquired from high-throughput instruments and automatically build a predictive model of the system. These "virtualization" technologies should allow for rapid cycling of

experimentation and modeling, thus greatly assisting in the rational exploration of complex biological systems.

4.4 Integration: Teaming at Disciplinary Interfaces To Advance Frontiers in Biological Research

Clearly, there is a strong need to integrate the biological sciences with analytical and computational technologies. Simply stimulating development in the area of measurement technology is insufficient. Moving the science forward requires integration, assimilation, and understanding. This will not be accomplished by technology, but by people. The exponential increase in the speed and scale of biological technologies has created the need for teaming in biology as never before. DOE BER has a tradition of driving multidisciplinary collaborations in the physical and biological sciences in concert with technology development and is ideally positioned to continue such efforts. In support of advancing the frontiers in biological research, BER should foster integration across biology, physics, mathematics, chemistry, computation, engineering, and materials science. In particular, it should facilitate the integrating of "omics" technologies with more traditional measurement modalities.

In the near term, DOE could help achieve the goal of increasing interdisciplinary science by supporting

- Integrative training for graduates, post graduates, and principal investigators
- Collaborative incentives to individuals and institutions
- Facilitated access to high-end technologies through more accessible user facilities that also provide adequate technical support for data interpretation
- Development of high-capability instruments at affordable prices (spread technology and opportunities to do good science across laboratories).

DOE also should strive to enhance communication within and across fields by

- Defining research and measurement standards across disciplines that are necessary to allow data exchange
- Generating large public domains for data exchange
- Identifying common overarching multiscale problems between fields
- Integrating experimental tools and design across BER programs.

The complexity of the biological challenges described in this workshop report requires scientific collaboration, which should be the expectation of this type of research. An important goal of any program addressing the frontiers of biological research should be to stimulate greater interactions among experimentalists, technology developers, computer scientists, and biological modelers. One way this can be done is to support the creation and development of integrated teams of biologists, physicists, chemists, material scientists, computation scientists, and engineers who work under one roof to tackle biological and technical challenges.

Just as the genome era ushered in the current biological revolution, emerging technological advances are poised to reveal untold biological mysteries. It was not so long ago that we could not quite conceive of sequencing an entire microbial genome, and now sequencing a human genome is increasingly commonplace. DOE was the first to take up the challenge of genome sequencing and has opened up an entirely new area of biological research. It now has the opportunity once again to change the future of biology. We do not know all of the discoveries that will ensue from advancing the frontiers of biological knowledge, but history indicates that many opportunities await us.

Appendix 1

New Frontiers in Characterizing Biological Systems Workshop

Agenda

Bethesda Marriott Hotel
May 13–14, 2009

May 12

7:00 – 9:00 p.m.	Dinner Meeting with Working Session Chairs

May 13

7:30 – 8:30 a.m.	Registration Sign-In and Continental Breakfast
8:30 – 9:45 a.m.	Welcome and Opening Remarks – Anna Palmisano and Sharlene Weatherwax
8:45 – 9:00 a.m.	Objective of Workshop – Arthur Katz
9:00 – 10:30 a.m.	Key Issues, Challenges, and Technology Needs • Biofuels: Brian Davison • Carbon Cycle: Colin Murrell • Environmental Remediation: Alfred Spormann
10:30 – 11:00 a.m.	Break
11:00 –11:30 p.m.	Parallel Breakout Sessions Technology Overview Mark Ellisman (Technology Perspective)
11:30 – 1:00 p.m.	Buffet –Working Lunch; Eat in Breakout Sessions
1:00 – 3:00 p.m.	Breakout Session Continues
3:00 – 3:30 p.m.	Break
3:30 – 4:30 p.m.	Plenary Session
4:30 – 5:30 p.m.	Plenary Session Continues - Interim Report from Each Breakout Group
6:30 – 7:30 p.m.	Dinner Meeting with Working Session Chairs • Update on the Progress and Issues from Each Group • Draft Outline for Group Report

May 14

7:30 – 8:30 a.m.	Breakfast
8:30 – 10:00a.m.	Parallel Breakout Session Wrap-Up
10:00 – 10:30 a.m.	Break
10:30 – 12:00 p.m.	Plenary Session – Final Report from Breakout Sessions
12:00 – 1:00 p.m.	Working Lunch (Meeting with Session Chairs)
1:00 – 3:00 p.m.	Plenary Session Round-Table Discussion with Session Chairs
3:00 – 3:15 p.m.	Workshop Concludes – DOE Concluding Remarks
3:15 – 5:30 p.m.	Writing Team Leaders Meeting • Complete Breakout Session Reports • Plans and Timeline for the Workshop Report

Appendix 2

New Frontiers in Characterizing Biological Systems Workshop

Writing Team

Barcellos-Hoff, Mary Helen*
New York University School of Medicine

Brandizzi, Federica*
Michigan State University

Davison, Brian
Chief Scientist, Genomic Science program (beginning Oct. 1, 2009)
Oak Ridge National Laboratory

Doktycz, Mitchel**
Oak Ridge National Laboratory

Fredrickson, Jim
Chief Scientist, Genomic Science program (through Sept. 30, 2009)
Pacific Northwest National Laboratory

Hugenholtz, Phillip**
DOE Joint Genome Institute

Sobecky, Patricia†
University of Alabama

Sweedler, Jonathan†
University of Illinois

Wiley, H. Steven
Pacific Northwest National Laboratory

*Cochair, Multicellular Session. **Cochair, Interface Session. †Cochair, Cellular Session.

Additional contributors: Wolfgang Fink, Bruce Fouke, Hoi-Ying Holman, Jan Liphardt, Tijana Rajh, Alfred Spormann, and Sunney Xie.

Participants

Barcellos-Hoff, Mary Helen
New York University School of Medicine

Bard, Allen
The University of Texas at Austin

Batt, Carl
Cornell University

Bhat, Talapady
National Institute of Standards and Technology

Brandizzi, Federica
Michigan State University

Chen, Liaohai
Argonne National Laboratory

Cooks, Graham
Purdue University

Cramer, Stephen
University of California, Davis

Davison, Brian
Oak Ridge National Laboratory

Doktycz, Mitchel
Oak Ridge National Laboratory

Ellisman, Mark
University of California, San Diego

Fink, Wolfgang
California Institute of Technology

Fouke, Bruce
University of Illinois

Fredrickson, Jim
Pacific Northwest National Laboratory

Gilna, Paul
University of California, San Diego

Haswell, Elizabeth
Washington University, St. Louis

Hlatky, Lynn
Tufts University School of Medicine

Holman, Hoi-Ying
Lawrence Berkeley National Laboratory

Hugenholtz, Phillip
DOE Joint Genome Institute

Knight, Rob
University of Colorado, Boulder

Lane, Stephen
University of California, Davis

Liphardt, Jan
Lawrence Berkeley National Laboratory

Markley, John
University of Wisconsin-Madison

Miller, Lisa
Brookhaven National Laboratory

Murphy, Robert
Carnegie Mellon University

Murrell, Colin
University of Warwick

Myles, Dean
Oak Ridge National Laboratory

Patterson, George
National Institutes of Health

Pianetta, Piero
Stanford Linear Accelerator Center

Rajh, Tijana
Argonne National Laboratory

Sobecky, Patricia
University of Alabama

Spormann, Alfred
Stanford University

Sweedler, Jonathan
University of Illinois

Taylor, Kenneth
Florida State University

van der Lelie, Daniel
Brookhaven National Laboratory

Vo-Dinh, Tuan
Duke University

Wallenstein, Matthew
Colorado State University

Wiley, H. Steven
Pacific Northwest National Laboratory

Xie, Sunney
Harvard University

Observers

Anderson, Todd
U.S. Department of Energy

Cole, Dean
U.S. Department of Energy

Graber, Joseph
U.S. Department of Energy

Gregurick, Susan
U.S. Department of Energy

Hirsch, Roland
U.S. Department of Energy

Katz, Arthur
U.S. Department of Energy

Mansfield, Betty
Oak Ridge National Laboratory

Marron, Michael
National Center for Research Resources

McLean, Gail
U.S. Department of Energy

Metting, Noelle
U.S. Department of Energy

Palmisano, Anna
U.S. Department of Energy

Sayler, Gary
(Biological and Environmental Research Advisory Committee)
University of Tennessee

Srivastava, Prem
U.S. Department of Energy

Stacey, Gary
(Biological and Environmental Research Advisory Committee)
University of Missouri

Stack, Robert
U.S. Department of Energy

Swain, Amy
National Center for Research Resources

Weatherwax, Sharlene
U.S. Department of Energy

Appendix 3

Bibliography

American Academy of Microbiology. 2006. "Microbial Energy Conversion: A Report from the American Academy of Microbiology," eds. Merry Buckley and Judy Wall (http://academy.asm.org/index.php?option=com_content&task=view&id=51&Itemid=54).

Bahatyrova, S., et al. 2004. "The Native Architecture of a Photosynthetic Membrane," *Nature* **430**(7003), 1058–62.

Bayer, E. A., Y. Shoham, and R. Lamed. 2006. "Cellulose-Decomposing Bacteria and Their Enzyme Systems." In *The Prokaryotes*, 3rd ed., Vol. 2: Ecophysiology and Biochemistry, edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, 578–617. Springer, New York.

Behrens, S., et al. 2008. "Linking Microbial Phylogeny to Metabolic Activity at the Single-Cell Level by Using Enhanced Element Labeling-Catalyzed Reporter Deposition Fluorescence In Situ Hybridization (EL-FISH) and NanoSIMS," *Applied and Environmental Microbiology* **74**(10), 3143–50. doi:10.1128/AEM.00191-08.

Bergeron, A., et al. 2007. "Divide and Conquer: Enriching Environmental Sequencing Data," PLoS ONE **2**(9), e830. doi:10.1371/journal.pone.0000830.

Betzig, E., et al. 2006. "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution," *Science* **313**, 1642–45.

Bisseling, T., J. L. Dangl, and P. Schulze-Lefert. 2009. "Next-Generation Communication," Science **324**(5928), 741. Special issue, *Plant-Microbe Interactions*.

Blakeslee, J. J., et al. 2007. "Interactions among PIN-FORMED and P-Glycoprotein Auxin Transporters in *Arabidopsis*," *The Plant Cell* **19**(1), 131–47.

Blow, N. 2009. "Systems Biology: Untangling the Protein Web," *Nature* **460**, 415–18. doi:10.1038/460415a.

Callister S. J., et al. 2008. "Comparative Bacterial Proteomics: Analytics of the Core Genome Concept," *PLoS ONE* **3**(2), e1542.

Cell Centered Database, ccdb.ucsd.edu.

Chan, J. W., et al. 2007. "Monitoring Dynamic Protein Expression in Living *E. coli* Bacterial Cells by Laser Tweezers Raman Spectroscopy," *Cytometry* **71A**(7), 468–74.

Choi, P., et al. 2008. "A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell," *Science* **322**(5900), 442–46.

Dong, M., et al. 2008. "A 'Tagless' Strategy for Identification of Stable Protein Complexes Genome-Wide by Multidimensional Orthogonal Chromatographic Separation and iTRAQ Reagent Tracking," *Journal of Proteome Research* **7**(5), 1836–49.

Dumont, C. G., and J. C. Murrell. 2005. "Stable Isotope Probing—Linking Microbial Identity to Function," *Nature Reviews Microbiology* **3**, 499–504.

Edwards, J. S., R. U. Ibarra, and B. O. Palsson. 2001. "*In silico* Predictions of *Escherichia coli* Metabolic Capabilities Are Consistent with Experimental Data," *Nature Biotechnology* **19**(2), 125–30. doi:10.1038/84379.

Elf, J., G.-W. Li, and X. S. Xie. 2007. "Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell," *Science* **316**, 1191–94.

Elowitz, M. B., et al. 2002. "Stochastic Gene Expression in a Single Cell," *Science* **297**, 1183–86.

Faley, S., et al. 2008. "Microfluidic Platform for Real-Time Signaling Analysis of Multiple Single T Cells in Parallel," *Lab on a Chip* **8**, 1700–12.

Feng, X., et al. 2009. "Characterization of the Central Metabolic Pathways in *Thermoanaerobacter* sp. Strain X514 via Isotopomer-Assisted Metabolite Analysis," *Applied and Environmental Microbiology* **75**(15), 5001–08.

Fink, W. 2008. "Stochastic Optimization Framework (SOF) for Computer-Optimized Design, Engineering, and Performance of Multi-Dimensional Systems and Processes," *Proceedings of the SPIE Conference on Space Exploration Technologies*, **6960**(69600N), doi:10.1117/12.784440.

Fong, S. S., and B. O. Palsson. 2004. "Metabolic Gene-Deletion Strains of *Escherichia coli* Evolve to Computationally Predicted Growth Phenotypes," *Nature Genetics* **36**(10), 1056–58. doi:10.1038/ng1432.

Fouke, B. W., et al. 2000. "Depositional Facies and Aqueous-Solid Geochemistry of Travertine-Depositing Hot Springs (Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, U.S.A.)," *Journal of Sedimentary Research* **70**(3), 565–85.

Fouke, B. W., et al. 2003. "Partitioning of Bacterial Communities Between Travertine Depositional Facies at Mammoth Hot Springs, Yellowstone National Park, U.S.A.," *Canadian Journal of Earth Sciences* **40**(11), 1531–48.

Frias-Lopez, J., et al. 2008. "Microbial Community Gene Expression in Ocean Surface Waters," *Proceedings of the National Academy of Sciences of the United States of America* **105**(10), 3805–10. doi: 10.1073/pnas.0708897105.

Galperin, M. Y., and E. V. Koonin. 2004. "'Conserved Hypothetical' Proteins: Prioritization of Targets for Experimental Study," *Nucleic Acids Research* **32**(18), 5452–63.

Glory, E., and R. F. Murphy. 2007. "Automated Subcellular Location Determination and High-Throughput Microscopy," *Developmental Cell* **12**(1), 7–16. doi:10.1016/j.devcel.2006.12.007.

Greenfield, D., et al. 2009. "Self-Organization of the *E. coli* Chemotaxis System Imaged with Super-Resolution Light Microscopy," *PLoS Biology* **7**(6), e1000137. doi: 10.1371/journal.pbio.1000137.

Gustafsson, M. G. L. 2005. "Nonlinear Structured-Illumination Microscopy: Wide-Field Fluorescence Imaging with Theoretically Unlimited Resolution," *Proceedings of the National Academy of Sciences of the United States of America* **102**(37), 13081–86.

Hell, S. 2003. "Toward Fluorescence Nanoscopy," *Nature Biotechnology* **21**(11), 1347–55.

Hess, S. T., T. P. Girirajan, and M. D. Mason. 2006. "Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy," *Biophysical Journal* **91**(11), 4258–72.

Huang, B., et al. 2008. "Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy," *Science* **319**, 810–13.

Huang, W. E., et al. 2007. "Raman-FISH: Combining Stable-Isotope Raman Spectroscopy and Fluorescence In Situ Hybridization for the Single Cell Analysis of Identity and Function," *Environmental Microbiology* **9**, 1878–89.

Hugenholtz, P., and G. W. Tyson. 2008. "Microbiology: Metagenomics," *Nature* **455**, 481–83. doi:10.1038/455481a.

Hura, G. L., et al. 2009. "Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS)," *Nature Methods* **6**(8), 606–12.

Kandianis, M., et al. 2008. "Microbial Biomass: A Catalyst for $CaCO_3$ Precipitation in Advection-Dominated Transport Regimes," *Geological Society of America Bulletin*, **120**(3–4), 442–50.

Kirchhoff, H., U. Mukherjee, and H. J. Galla. 2002. "Molecular Architecture of the Thylakoid Membrane: Lipid Diffusion Space for Plastoquinone," *Biochemistry* **41**(15), 4872–82.

Kunin, V., et al. 2008. "Millimeter-Scale Genetic Gradients and Community-Level Molecular Convergence in a Hypersaline Microbial Mat," *Molecular Systems Biology* **4**, 198. doi:10.1038/msb.2008.35.

Kruger, N. J., and R. G. Ratcliffe. 2009. "Insights into Plant Metabolic Networks from Steady-State Metabolic Flux Analysis," *Biochimie* **91**(6), 697–702.

Llaneras, F., and J. Picó. 2008. "Stoichiometric Modelling of Cell Metabolism," *Journal of Bioscience and Bioengineering* **105**(1), 1–11. doi:10.1263/jbb.105.1.

McMahon, K., H. García Martín, and P. Hugenholtz. 2007. "Integrating Ecology into Biotechnology," *Current Opinion in Biotechnology* **18**(3), 287–92. doi:10.1016/j.copbio.2007.04.007.

National Research Council, Committee on Review of the Department of Energy's Genomics: GTL Program. 2006. *Review of the Department of Energy's Genomics:GTL Program*, National Academies Press, Washington, D.C.

Nealson, K., and W. A. Ghiorse. 2001. *Geobiology: Exploring the Interface Between the Biosphere and the Geosphere*, American Academy of Microbiology, Washington, D.C.

Neufeld, J. D., M. Wagner, and J. C. Murrell. 2007. "Who Eats What, Where and When? Isotope-Labelling Experiments Are Coming of Age," *ISME Journal* **1**, 103–10.

Pelletier, D. A., et al. 2008. "A General System for Studying Protein-Protein Interactions in Gram-Negative Bacteria," *Journal of Proteome Research* **7**(8), 3319–28.

Peng, H. 2008. "Bioimage Informatics: A New Area of Engineering Biology," *Bioinformatics* **24**(17), 1827–36.

Pinchuk, G. E., D. A. Rodionov, C. Yang, et al. 2009. "Genomic Reconstruction of *Shewanella oneidensis* MR-1 Metabolism Reveals a Previously Uncharacterized Machinery for Lactate Utilization," *Proceedings of the National Academy of Sciences of the United States of America* **106**(8), 2874–79.

Price, N., et al. 2003. "Genome-Scale Microbial *In Silico* Models: The Constraints-Based Approach," *Trends in Biotechnology* **21**(4), 162–69.

Pushkarev, D., N. F. Neff, and S. R. Quake. 2009. "Single-Molecule Sequencing of an Individual Human Genome," *Nature Biotechnology* **27**, 847–50. doi:10.1038/nbt.1561.

Ratcliffe, R. G., and Y. Shachar-Hill. 2006. "Measuring Multiple Fluxes Through Plant Metabolic Networks," *Plant* Journal **45**(4), 490–511.

Rozhkova, E. A., et al. 2009. "A High-Performance Nanobio Photocatalyst for Targeted Brain Cancer Therapy," *Nano Letters*, Article ASAP. doi: 10.1021/nl901610f.

Rust, M. J., M. Bates, and X. Zhuang. 2006. "Sub-Diffraction–Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)," *Nature Methods* **3**(10), 793–95.

Schilling, C. H., et al. 2002. "Genome-Scale Metabolic Model of *Helicobacter pylori* 26695," *Journal of Bacteriology* **184**(16), 4582–93. doi:10.1128/JB.184.16.4582-4593.2002.

Sener, M. K., et al. 2007. "Atomic-Level Structural and Functional Model of a Bacterial Photosynthetic Membrane Vesicle," *Proceedings of the National Academy of Sciences of the United States of America* **104**(40), 15723–28.

Shaner, N. C., G. H. Patterson, and M. W. Davidson. 2007. "Advances in Fluorescent Protein Technology," *Journal of Cell Science* **120**(24), 4247–60.

Shi, Y., G. W. Tyson, and E. F. DeLong. 2009. "Metatranscriptomics Reveals Unique Microbial Small RNAs in the Ocean's Water Column," *Nature* **459**, 266–69. doi:10.1038/nature08055.

Shroff, H., et al. 2008. "Live-Cell Photoactivated Localization Microscopy of Nanoscale Adhesion Dynamics," *Nature Methods* **5**(5), 417–23.

Stolyar, S., et al. 2007. "Metabolic Modeling of a Mutualistic Microbial Community," *Molecular Systems Biology* **3**(92). doi:10.1038/msb4100131.

Tang, Y. J., et al. 2008. "Advances in Analysis of Microbial Metabolic Fluxes Via ^{13}C Isotopic Labeling," *Mass Spectrometry Reviews* **28**(2), 362–75.

Tringe, S. G., and P. Hugenholtz. 2008. "A Renaissance for the Pioneering 16S rRNA Gene," *Current Opinion in Microbiology* **11**(5), 442–46.

U.S. DOE. 2006. *Breaking the Barriers to Cellulosic Ethanol: A Joint Research Agenda*. DOE/SC-0095 (http://genomicscience.energy.gov/biofuels/b2bworkshop.shtm).

U.S. DOE. 2008. *Carbon Cycling and Biosequestration: Integrating Biology and Climate Through Systems Science.* DOE/SC-108 (http://genomicscience.energy.gov/carboncycle/).

U.S. DOE. 2009. *U.S. Department of Energy Office of Science Systems Biology Knowledgebase for a New Era in Biology: A Genomics:GTL Report from the May 2008 Workshop*, DOE/SC-113, U.S. Department of Energy Office of Science (http://genomicsgtl.energy.gov/compbio).

Vaziri, A., et al. 2008. "Multilayer Three-Dimensional Super Resolution Imaging of Thick Biological Samples," *Proceedings of the National Academy of Sciences of the United States of America* **105**(51), 20221–26.

VerBerkmoes, N. C., et al. 2009. "Systems Biology: Functional Analysis of Natural Microbial Consortia Using Community Proteomics," *Nature Reviews Microbiology* **7**, 196–205. doi:10.1038/nrmicro2080.

Veysey, J., et al. 2008. "Reconstruction of Water Temperature, pH, and Flux of Ancient Hot Springs from Travertine Depositional Facies," *Journal of Sedimentary Research*, **78**(2), 69–76.

Vinson, V., and G. Chin. 2007. "Lights, Camera, Action," *Science* **316**(5828), 1143–58. Special issue on single molecules. doi: 10.1126/science.316.5828.

Warnecke, F., et al. 2007. "Metagenomic and Functional Analysis of Hindgut Microbiota of a Wood-Feeding Higher Termite," *Nature* **450**, 560–65.

Woese, C. R. 2004. "A New Biology for a New Century," *Microbiology and Molecular Biology Reviews* **68**, 173–86.

Xie, X. S., et al. 2008. "Single-Molecule Approach to Molecular Biology in Living Bacterial Cells," *Annual Reviews of Biophysics* **37**, 417–44.

Yu, J., et al. 2006. "Probing Gene Expression in Live Cells One Protein Molecule at a Time," *Science* **311**, 1600–03.

Appendix 4

Figure Credits

Cover Images

Ecosystem image courtesy of U.S. Department of Agriculture Natural Resources Conservation Service. Microbial community image © Simko/Visuals Unlimited. Plant tissue image (switchgrass stem cross-section) courtesy of S. Ding, DOE BioEnergy Science Center. Cellular systems image of *E. coli* protein expression from Greenfield et al. 2009. Cellular systems image of switchgrass plant cell walls using atomic force microscope courtesy of S. Singh and B. Simmons, DOE Joint BioEnergy Institute. Molecular and cellular systems image showing protein complexes courtesy of M. Biggin, Lawrence Berkeley National Laboratory.

Figure 2.2

RuBisCO image from PDB ID: 9RUB, Lundqvist, T., and G. Schneider. 1991. "Crystal Structure of Activated Ribulose-1,5-bisphosphate Carboxylase Complexed with Its Substrate, Ribulose-1,5-bisphosphate," *Journal of Biological Chemistry* **266**(19), 12604–611. Bacterial ribosome image courtesy of H. Noller, Center for Molecular Biology of RNA, University of California, Santa Cruz. Bacterial cellulosome image from Y. Bomble, M. Crowley, and M. Himmel, DOE BioEnergy Science Center. Carboxysome image courtesy of D. Morris and G. Jensen and reprinted from *Journal of Molecular Biology* **372**, Iancu, C., et al., "The Structure of Isolated *Synechococcus* Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography," 764–73, © 2007, with permission from Elsevier. *Synechocystis* image from W. Vermaas, Arizona State University. Bacterial periplasm image and *Shewanella* bacteria image from A. Dohnalkova, Environmental Molecular Sciences Laboratory, DOE Pacific Northwest National Laboratory. Plant cell-wall thickness image courtesy of S. Jung and A. Ragauskas, Georgia Tech, DOE BioEnergy Science Center. Chloroplast image © G. Chapman/Visuals Unlimited. Mitochondrion image © H. Berg/Visuals Unlimited. Plant cell image © M. Powell/Visuals Unlimited.

Figure 3.2

Ecosystem image courtesy of U.S. Department of Agriculture Natural Resources Conservation Service. Microbial community image from micro*scope, http://microscope.mbl.edu. Plant cell image © M. Powell/Visuals Unlimited. Cyanobacterial cell image from W. Vermaas, Arizona State University. Carboxysome image courtesy of D. Morris and G. Jensen and reprinted from *Journal of Molecular Biology* **372**, Iancu, C., et al., "The Structure of Isolated *Synechococcus* Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography," 764–73, © 2007, with permission from Elsevier. Protein complex image from PDB ID: 9RUB, Lundqvist, T., and G. Schneider. 1991. "Crystal Structure of Activated Ribulose-1,5-bisphosphate Carboxylase Complexed with Its Substrate, Ribulose-1,5-bisphosphate," *Journal of Biological Chemistry* **266**(19), 12604–611.

Acronym List

AFM	atomic force microscopy
BER	DOE Office of Biological and Environmental Research
CARD-FISH	catalyzed reporter deposition–fluorescence in situ hybridization
DNA	deoxyribonucleic acid
DOE	Department of Energy
EL-FISH	elemental labeling–fluorescence in situ hybridization
EM	electron microscopy
FBA	flux balance analysis
FISH	fluorescence in situ hybridization
FRET	fluorescence resonance energy transfer
IR	infrared radiation
MAR-FISH	microautoradiography–fluorescence in situ hybridization
MFA	metabolic flux analysis
MRI	nuclear magnetic resonance imaging
MS	mass spectrometry
nanoSIMS	nano–secondary ion mass spectrometry
NMR	nuclear magnetic resonance
PALM	photoactivated localization microscopy
PET	positron emission tomography
RNA	ribonucleic acid
RuBisCO	ribulose-1,5-bisphosphate carboxylase oxygenase
STED	stimulated emission depletion
STORM	stochastic optical reconstruction microscopy
UV	ultraviolet

Reference Table of Numeric Prefixes

10^{24}	yotta
10^{21}	zetta
10^{18}	exa
10^{15}	peta
10^{12}	tera
10^{9}	giga
10^{6}	mega
10^{3}	kilo
10^{0}	
10^{-3}	milli
10^{-6}	micro
10^{-9}	nano
10^{-12}	pico
10^{-15}	femto
10^{-18}	atto
10^{-21}	zepto
10^{-24}	yocto

Printed with soy ink on recycled paper

U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE 2009