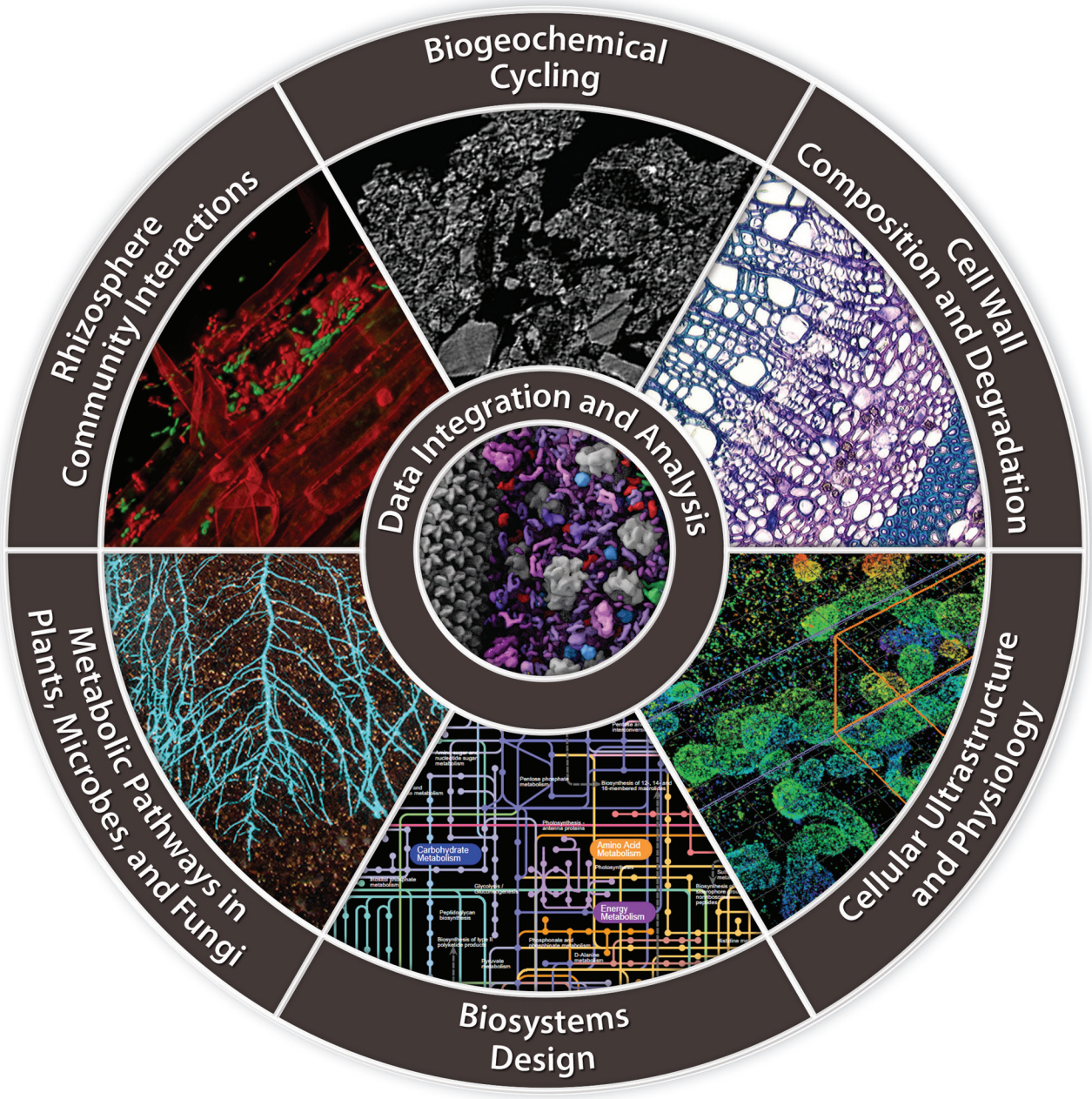
DOE/SC-0189

Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment

Workshop Report



U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Biological and Environmental Research

Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment Workshop

September 21–23, 2016

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

Co-Chairs

Paul Adams, Ph.D.
Lawrence Berkeley National Laboratory

Elizabeth Wright, Ph.D.
Emory University

Organizers

Biological Systems Science Division

Todd Anderson, Ph.D.

Amy Swain, Ph.D.
Amy.Swain@science.doe.gov
301.903.1828

This report is available at science.energy.gov/ber/community-resources/ and genomicscience.energy.gov.

Mission

The Office of Biological and Environmental Research (BER) advances world-class fundamental research programs and scientific user facilities to support the Department of Energy's energy, environment, and basic research missions. Addressing diverse and critical global challenges, the BER program seeks to understand how genomic information is translated to functional capabilities, enabling more confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, or contaminant bioremediation. BER research advances understanding of the roles of Earth's biogeochemical systems (the atmosphere, land, oceans, sea ice, and subsurface) in determining climate so that it can be predicted decades or centuries into the future, information needed to plan for future energy and resource needs. Solutions to these challenges are driven by a foundation of scientific knowledge and inquiry in atmospheric chemistry and physics, ecology, biology, and biogeochemisty.

Suggested citation for this report: U.S. DOE. 2017. *Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment*, DOE/SC-0189, U.S. Department of Energy Office of Science. science.energy.gov/ber/community-resources/.

DOE/SC-0189

Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment

Workshop Report

September 2017

U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Biological and Environmental Research

Contents

Executive Summary ..... **v**

Research Themes v

Overarching Challenges and Opportunities vii

1. Introduction **1**

1.1 Historical Perspective of BSSD Genomics Characterization 1

1.2 BSSD Structural Biology and Bioimaging Science: Atoms to Mesoscale 2

2. Cell Wall Composition and Degradation **5**

2.1 Current Science and Technologies 5

2.2 Major Challenges in Monitoring Plant Cell Walls 6

2.3 Advancing Biofuel and Bioproduct Design and Production 14

3. Rhizosphere Community Interactions **15**

3.1 Current Science and Technologies 16

3.2 Major Challenges in Detecting Rhizosphere Interactions 17

3.3 Major Needs in State-of-the-Art Technologies for Understanding Rhizospheres 19

3.4 Optimizing Plant Health, Soil Fertility, and Carbon Storage 27

4. BiogeochemicalCycling **29**

4.1 Current Science and Technologies 29

4.2 Major Challenges in Learning How Microbes Control Biogeochemistry Cycling and Mobility 30

4.3 Improving Model Predictions of Biological and Environmental Systems 35

5. Metabolic Pathways in Plants, Microbes, and Fungi **37**

5.1 Current Science and Technologies 38

5.2 Major Challenges in Detecting Metabolic Pathways 40

5.3 Predictively Understanding the Metabolome Across Scales, from Organisms to Ecosystems 43

6. Biosystems Design **45**

6.1 Current Science and Technologies 46

6.2 Major Challenges in Design of Engineered Biosystems 51

6.3 Leveraging Synthetic Biology for Future Energy and Resource Needs 56

7. Cellular Ultrastructureand Physiology **59**

7.1 Science and Technologies 59

7.2 Understanding Genotype-Phenotype Interactions to Improve Biofuel Feedstocks 69

8. Integration and Analysis **71**

8.1 Current Science and Technologies 71

8.2 Major Challenges in Data and Computation 74

8.3 Advancing Computational Platforms for Large-Scale Data Processing and Analysis 78

9. Summary and Conclusions **81**

Appendices **87**

Appendix A: Workshop Participants 87

Appendix B: Workshop Agenda 88

Appendix C: Figure Credits 91

Appendix D: Acronyms and Abbreviations 93

Appendix E: References 95

Executive Summary

The Biological Systems Science Division (BSSD) within the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) funds basic research on plants and microbes relevant to several DOE bioenergy and environmental mission areas. BSSD's long history of developing and supporting genomic characterization of biological systems has led, in part, to the high-throughput DNA sequencing technology available to researchers worldwide. In recent years, genome sequencing has vastly outpaced the ability to interpret genome function. Fully maximizing this wealth of data will require new technology advancements, along with improvements and an increase in the throughput of existing methods for characterizing molecular- to cellular-level processes important for inferring biological function.

BSSD research seeks to understand the fundamental genome-encoded properties of plants and microbes that can be harnessed or redesigned for beneficial purposes. Current emphases are leading to the discovery, development, and understanding of numerous plant and microbial species with traits suitable for the production of fuels and chemical products from renewable biomass that could be grown synergistically with food or animal feed crops while not competing with other societal needs. Additionally, BSSD further supports research leading to an understanding of the complex and essential interactions among plants, microbial communities, and the environment to find new ways to sustainably produce biomass for a range of bioenergy and bioproduct applications. This research also is relevant for incorporation into larger-scale environmental models such as those developed through the research supported by BER's Earth and Environmental Systems Sciences programs.

To engage the relevant scientific communities in discussions of these research areas, BER convened the Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment workshop on September 21–23, 2016 (see Appendix A, p. 87, and Appendix B, p. 88). Seeking to enable more comprehensive systems biology–based approaches, which typically require measurements of many samples, workshop participants highlighted the need for the development of highly sensitive methods to provide accurate measurements from small sample volumes and that are operable in high-throughput or highly parallel modes. Achieving these goals is critical to enabling predictive engineering of biological systems, as is further development of manipulation technologies. Biosystems design has the potential to revolutionize the way biology is exploited to produce economically valuable molecules; however, this will be possible only if high-throughput measurement technologies are combined with tools for precise genetic manipulation, and computational algorithms are devised for accurate prediction of phenotype resulting from genome manipulation. Participants at the workshop, organized by BSSD, concluded that multimodal methods will be required for many of the research needs discussed. As structural biology and imaging methods are converging, multiscale, multidisciplinary approaches to plant and microbial cell biology are increasingly emerging.

Research Themes

These fundamental research efforts require new and innovative methods and technologies to elucidate the foundational principles that drive biological systems of interest to DOE's energy and environmental missions. Characterizing biological systems involves analytical approaches that illuminate cellular components and their form, structure, size, function, spatial location, dynamics, and interactions with the environment. Workshop discussions identified new technologies and combinations of existing capabilities to address the challenges associated with

characterizing molecular and cellular systems relevant to bioenergy and environmental research. Participants included technology developers and biology researchers with expertise in cellular ultrastructure and physiology, bioenergy and bioproducts, and environmental microbiology. Attendees developed a series of research and technology development needs across six thematic areas spanning the range of BSSD-supported research. The challenges of studying these systems are many and broad in scope, covering time scales from femtoseconds to weeks and length scales from Angstroms to centimeters. This report addresses this very broad measurement range—from cells and their metabolism and mineralogy (Angstroms to micrometers), to rhizosphere ecosystem processes and community biochemical activity (millimeters to a meter). In this context, the range from micrometers to a meter is referred to as "mesoscale." Despite the breadth of the challenges, participants identified key needed technologies and improvements in current techniques that could advance BER science. These six major research themes are discussed below.

Cell Wall Composition and Degradation. The benefits from gaining a molecular-level understanding of plant cell wall composition and degradation were discussed in the context of using plants in the production of biofuels and bioproducts. As a renewable resource for biofuels and biomaterials, lignocellulosic biomass can partially replace the use of diminishing petroleum-based fuels and products and help meet increasing consumer demand for green chemicals. However, the varying structure and chemical composition of the cell walls of different plants and tissue types may hinder industrial-scale processes for converting biomass to bioproducts. Needed to address this challenge are better atomic- and molecular-level understandings of the structure and dynamics of naturally occurring cell wall processes, as well as the processes involved in the production of biofuels and other chemicals. Also needed are new characterization techniques with nanometer-scale resolution that require minimal sample preparation and keep the sample in close-to-natural conditions.

Several technologies and techniques identified will aid the understanding of plant cell wall properties at the anatomical, cellular, molecular, and genetic levels. Other new approaches suggested will provide as-yet-undiscovered molecular details about structural and temporal rearrangement of cell wall components during biomass deconstruction prior to conversion to biofuels and bioproducts.

Rhizosphere Community Interactions. A better understanding of ecosystems is yielding deeper insights into plant-microbe-mineral interactions important for bioenergy production. Knowing the complex interdependencies of these three systems is critical to understanding and developing sustainable biofuel production practices. Root system architecture has a dramatic bearing on plant viability and crop productivity in given soil conditions. Namely, the rhizosphere, the area immediately surrounding plant roots, is a nexus of biological activity and the foundational ecosystem for any plant-microbe system. Thus, studies are needed of all essential communal elements necessary for plant growth and yield across a range of geographic regions. Understanding these ecosystems can enable the design of optimally mutualistic plant-microbe interactions to improve biofuel crop sustainability.

For these studies, the development of penetrating imaging tools is needed to study entire, complex soil environments and root system architectures.

Biogeochemical Cycling. Environmental system function is intimately tied to the biogeochemical cycling of the major elements, particularly their reduction-oxidation (redox) transformations. Spatial and temporal imaging and measurements of biogeochemical systems are necessary for a mechanistic understanding of how different biogeochemical systems function. There also is a need for development of, and improvements to, technologies and approaches that will enable researchers (1) to understand and predict the dynamic interplay

between environmental biotic and abiotic factors that often are opaque to imaging tools, from the molecular to the mesoscale, and (2) to use this new understanding to predict larger-scale phenomena.

A combined and holistic use of a variety of dynamic imaging and characterization probes, coupled with multiomic and modeling approaches, is necessary to span spatial and temporal scales in biogeochemical systems to better understand their role in key environmental processes.

Metabolic Pathways in Plants, Microbes, and Fungi. Plants and microbes exchange metabolites in a community economy that ultimately determines the rates at which nutrients and water are extracted from soil and soil carbon is cycled (i.e., the biogeochemistry). A deeper understanding of the mechanisms by which organisms interact with each other in the environment, and the metabolic pathways and specific molecules involved in these interactions, will enable the modification of these pathways to improve nutrient-use efficiency and soil-carbon performance.

New tools are required for predicting and measuring metabolites from organisms key to BER bioenergy and environmental missions. Combined with new higher-resolution approaches, these technologies will need capabilities for determining spatiotemporal localization and mechanisms responsible for metabolite synthesis, transport, degradation, and perception. The ultimate goal is a "balanced record" of metabolite economy among plant-microbe-fungi interactions and the environment that fully accounts for all carbon and nutrient cycling in the system.

Biosystems Design. Synthetic biology provides a valuable approach to probe, study, and engineer new functions into biological systems through the introduction or modification of metabolic pathways, specifically generating biologically derived chemicals, fuels, and materials to ensure environmental sustainability. Challenges include (1) applying synthetic biology to intractable eukaryotic and multicellular organisms, (2) engineering communities of microorganisms and microbe-plant interfaces, (3) exploring genotype-phenotype landscapes resulting from genome engineering, (4) isolating engineered organisms with desired functions, and (5) safeguarding engineered biosystems.

Efficient tools for the precise manipulation of genomes in diverse target organisms will need to be combined with improved computational modeling methods to support predictive biology. These coupled approaches will require assistance from new methods for rapidly assaying function and fitness. They also must be applicable to technologies for controlling the containment of engineered systems and the products of engineered pathways.

Cellular Ultrastructure and Physiology. BER research examines a range of plant and microbial cell structures and organization, from the atomic level to complex molecular machines, cellular compartments, scaffolds, and whole cells. Workshop participants identified several measurement challenges, including how to detect and visualize cellular dynamic processes such as metabolic cycles, signaling and trafficking in plants, and interactions among microbial and fungal communities. They described needs for improved (1) structural imaging at the atomic and molecular level, (2) methods for illuminating whole organisms to understand the internal organization of cells, and (3) imaging chemical events that underlie biology. These needs include methods to determine the locations and dynamic parameters of enzyme reactions within cells, as well as the flow of chemicals and macromolecules within and between cells. The structural and dynamical insights from such studies will inform and enable more accurate modeling of biogeochemical cycling and metabolic pathways important in rhizospheric communities and biofuel or bioproduct processes.

Overarching Challenges and Opportunities

Several challenges common to all the research themes emerged throughout workshop discussions. Translating information from genomic

studies to the molecular and cellular realm for characterization will require increased throughput for existing technologies and the development of new high-throughput approaches. Achieving these goals will necessarily involve more automation and computational algorithms to manage the high data volumes that will be produced. Improved machine-learning approaches and large data-handling capacity will be essential. Integration of disparate data types from multiple and heterogeneous sources remains a challenge, so continued development of integrative and interpretive computational approaches is needed. Similar needs also were discussed at a workshop hosted by DOE's Office of Advanced Scientific Computing Research (ASCR), the DOE Exascale Requirements Review, held March 28–31, 2016, in Rockville, Md., which generated the meeting report, *ASCR Exascale Requirements Review* (science.energy.gov/ascr/community-resources/program-documents/).

The tools and methods described in this BER report are critical for advancing the deep understanding of complex, multicomponent systems that are central to bioenergy and the environment. While new technologies are needed for advancing leading-edge biological insights, they are of limited value if they are not readily accessible by the scientists who need them to conduct their research. As new instruments, platforms, and approaches are created, it is important that they be developed in ways that ultimately enable biology researchers to use them, either by adopting them in their own laboratories or by having access to the tools, appropriate expertise, and support at national user facilities. Elements will include robust hardware, physiologically relevant sample preparation and measurement conditions, automation, sophisticated analytical algorithms, and user-friendly interfaces. For facility-based technologies, long-term and productive community access requires recognition of the need for ongoing operational support.

Described herein are some of the workshop's identified challenges to studying the biological systems of interest to BSSD, which has a history of developing and supporting highly sophisticated research tools and techniques and ensuring that researchers can access them to advance science in support of the division's goals. Workshop discussions reflected in this document will help guide the next generation of imaging and analytical instrumentation needed to gain a predictive understanding of biological systems supporting DOE's energy and environmental missions.

BER appreciates the tireless efforts of the workshop co-chairs and the contributions of all the workshop participants. BER also extends special thanks to Kris Christen, Holly Haun, Brett Hopwood, Sheryl Martin, Stacey McCray, Marissa Mills, Judy Wyrick, and Betty Mansfield of Oak Ridge National Laboratory's Biological and Environmental Research Information System for editing and preparing this report for publication.

Chapter 1

Introduction

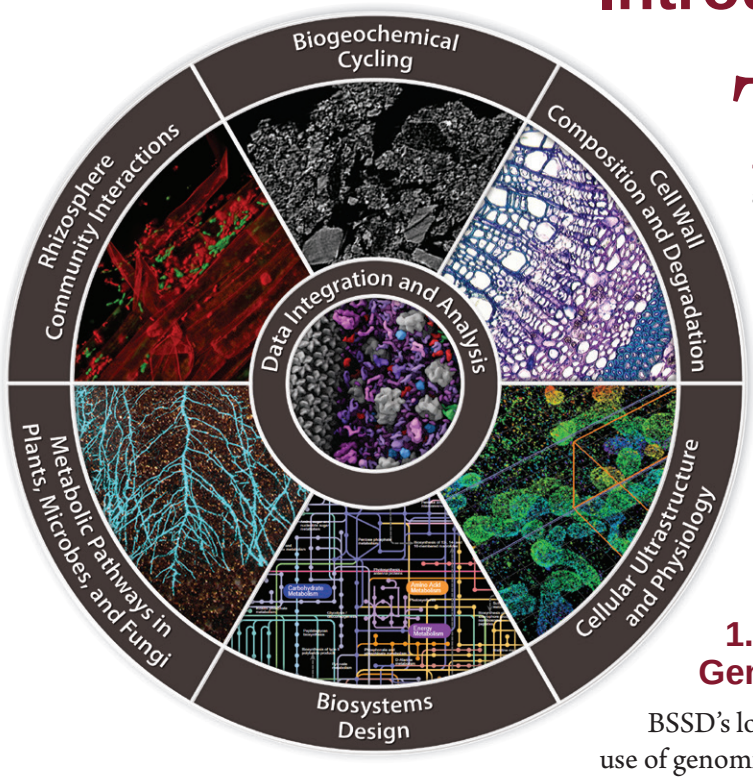


Fig. 1.1. Research Targets Supported by the Office of Biological and Environmental Research (BER) Biological Systems Science Division (BSSD). BER is managed within the U.S. Department of Energy's Office of Science. **Clockwise from top: (1)** Biogeochemical cycling; **(2)** cell wall composition and degradation; **(3)** cellular ultrastructure and physiology; **(4)** biosystems design; **(5)** metabolic pathways in plants, microbes, and fungi; **(6)** rhizosphere community; **(center)** mesoscale molecular model. BSSD research is linked via data integration and analysis. [See Appendix C, p. 91, to view image attributions and permissions.]

The Biological Systems Science Division (BSSD) of the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) supports systems biology research that targets the foundational principles driving biological systems of microbes, plants, and multispecies communities relevant to BSSD's bioenergy and environmental missions (see Fig. 1.1, this page). Managed within DOE's Office of Science, this research includes pursuit of a systems-level understanding of the spatiotemporal expression of biomolecules and structures within microbial and plant cells.

1.1 Historical Perspective of BSSD Genomics Characterization

BSSD's long history of developing and supporting the use of genomic characterization of biological systems has led, in part, to the high-throughput DNA sequencing technology available to researchers worldwide. In recent years, genome sequencing has vastly outpaced the ability to interpret genome function. There is a great opportunity for the further development of methods for characterizing biological processes at the molecular and cellular level to enable the inference of function. BSSD research seeks to understand the fundamental genome-encoded properties of plants and microbes that can be harnessed or redesigned for beneficial purposes. Current emphases are leading to the discovery, development, and understanding of numerous plant and microbial species with traits suitable for the production of fuels and chemical products from renewable biomass that could be grown synergistically with food or animal feed crops while not competing with other societal needs. Additionally,

BSSD supports research leading to an understanding of the complex and essential interactions among plants, microbial communities, and the environment. Also aiming to create a more mechanistic understanding of the dynamic nature of cellular metabolism, much of the BSSD research portfolio spans the following three broad areas.

Bioenergy and Bioproduct Production. Central to DOE's mission is support of fundamental research to provide the knowledge underpinning development of renewable energy sources. Plant biomass is a long-recognized source of renewable sugars and other compounds for the biological production of fuels and other useful chemicals. Successful development of these biological approaches ultimately will require a detailed understanding of how to optimize plants and microbes for biomass production and sustainable growth, how biomass is constructed and how to deconstruct it, how to optimize enzymatic pathways that produce desired molecules and biopolymers (bioproducts), and how these pathways are regulated in the context of cell metabolism.

Environmental Microbiology. Microbes and their communities significantly affect biogeochemical transformations in a wide range of diverse ecosystems. BSSD-funded research efforts focus on how microbial consortia communicate; evolve; share resources; interact with other organisms in the rhizosphere and aquatic environments; are affected by changes in the environment; and, ultimately, play a role in defining the Earth's landscape. The improvement and use of novel instrumentation and methodologies will further inform BER-sponsored research directions associated with microbial ecology and environmental and climatic changes.

Cellular Ultrastructure and Physiology. BSSD supports research that focuses on examining the structure and function of whole microbial organisms and plant cells to assess how cellular and subcellular structures of individual organisms correlate with specific biochemical, molecular, genetic, and behavioral pathways. Successful investigations may lead to the development of synthetic systems that can replicate functions of natural systems or carry out novel functions not observed in nature.

1.2 BSSD Structural Biology and Bioimaging Science: Atoms to Mesoscale

Analysis of biological systems generally extends across many orders of magnitude in length and time (see Fig. 1.2, p. 3). Environmental phenomena at the meter-length scale are intrinsically linked with the activities of biological systems at the cellular level, which in turn are the result of the activities of individual enzymes at the atomic level. This report addresses the very broad measurement ranges required—from cells and their metabolism and mineralogy (Angstroms to micrometers), to rhizosphere ecosystem processes and community biochemical activity (millimeters to a meter). In this context, the range from micrometers to a meter is referred to as "mesoscale." Combined experiments that traverse all these length scales are currently rare, highlighting the important need for capabilities to measure these systems at very different length scales and resolutions and then rigorously correlate the results. The time domain in biology is equally as important as length. Emergent environmental phenomena may evolve over months or years but, ultimately, they are influenced by cellular activities on the millisecond time scale. In turn, cellular activities are a result of enzymatic activities whose key steps may be atomic rearrangements at the subpicosecond time scale. The ability to predict the time evolution of complex biological systems demonstrates a profound level of understanding and opens up the possibility of their control and manipulation for defined outcomes. Over the last century, multiple techniques have been developed to measure biological systems across these length and time scales.

Participants in BER's Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment workshop in September 2016 considered gaps in technologies for

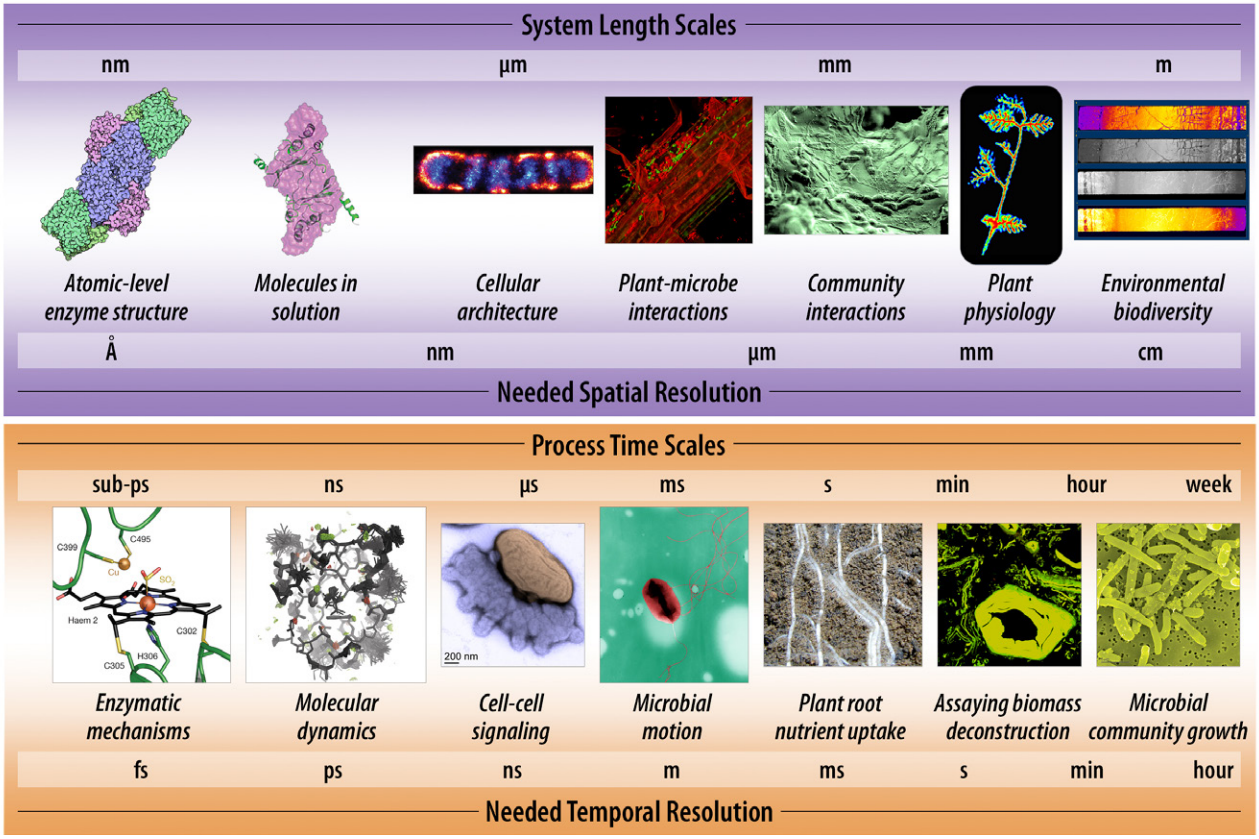


Fig.1.2. Overview of Needed Length and Time Scales. Biological and Environmental Research program biological interests include complex processes that span a size range from Angstroms to a meter and a dynamic range from femtoseconds to a week. Representative processes and their elements are illustrated. [See Appendix C, p. 91, to view image attributions and permissions.]

Key: Å, Angstrom; **cm,** centimeter; **fs,** femtosecond; **m**, meter; **mm**, millimeter; **min**, minute; **µm**, micrometer; **µs**, microsecond; **ms**, millisecond; **nm**, nanometer; **ns**, nanosecond; **ps**, picosecond; **s**, second; **sub-ps**, subpicosecond.

characterizing biological systems. They cited the importance of knowing what technologies exist, but also, on a more practical level, what is currently available to BSSD researchers. Furthermore, participants recognized that BER supports a number of resources that should be kept at the leading edge of science. Development or implementation of new technologies, where appropriate and needed, may strategically leverage or add to existing resources. Following are examples of the BER-supported advanced technology resources available to and used by the BSSD research community.

- The DOE Joint Genome Institute (JGI; jgi.doe.gov) provides advanced sequencing and synthesis of plant and microbial genomes.
- The DOE Systems Biology Knowledgebase (KBase; kbase.us) is an openly available informatics resource for collaborative, computational modeling of plant, microbial, and community systems.
- BER-supported structural biology beamline resources (berstructuralbioportal.org) are available at synchrotron and neutron facilities

supported by the DOE Office of Basic Energy Sciences (BES). These beamlines elucidate structures ranging from the atomic to the tissue scale and provide dynamic information from femtosecond to seconds.

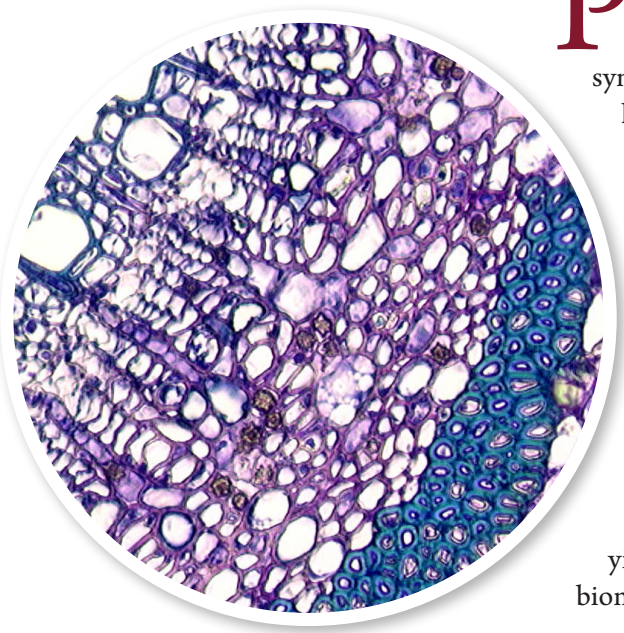
- The BER Environmental Molecular Sciences Laboratory (EMSL; emsl.pnl.gov/emslweb/) offers a suite of over 50 advanced and varied capabilities, many of which are immensely useful for BSSD researchers, including computing resources.
- BER researchers also have access to advanced computing resources (science.energy.gov/ascr/facilities/); in particular, the National Energy Research Scientific Computing Center (NERSC; nersc.gov).

BSSD's Bioimaging Technology program (science.energy.gov/ber/bioimaging-technology/) targets creation of novel multifunctional technologies to image, measure, and model key metabolic processes within and among microbial cells and multicellular plant tissues. This ongoing program supports the development of stand-alone, *in situ*, nondestructive imaging platforms. As they are made robust, these instrument platforms are made available to BSSD researchers.

The following six chapters summarize the driving research needs in areas key to BSSD, the current technologies available, and the needed new and improved technologies and their potential impact. The penultimate chapter discusses the need for new computational and modeling approaches to enable these areas of BSSD-supported biological research.

Chapter 2

Cell Wall Composition and Degradation



Plant cell walls consist primarily of cellulose, hemicelluloses, and lignin that, in addition to starch, are the major carbon-containing products of photosynthesis. This lignocellulosic fraction of plant biomass has great potential as a renewable feedstock for biofuels and biomaterials. Achieving this potential can offset the diminishing availability of fossil fuels and meet increasing consumer demand for green chemicals. Emerging biorefinery methods use thermochemical pretreatment and enzymatic hydrolysis to deconstruct plant cell walls to monomeric sugars that microbes then ferment into biofuels. Advancing the viability of biorefineries requires deeper understanding of the biosynthesis of plant cell walls and their physiochemical properties as well as the rate and yield of chemical and enzyme processes used for biomass deconstruction.

Understanding the complex organization of plant tissues and the cell wall polymers that comprise them requires the use of many techniques to characterize differences in cellular ultrastructure and chemical composition across spatial scales of millimeters to nanometers. Scientists need improved spatial resolution correlated with chemical bonding information to advance knowledge of how chemical treatments and biological catalysts work synergistically to convert plant feedstocks into useful sugars that are the basis for biofuels.

2.1 Current Science and Technologies

Cellulose. Of the three major structural polymers constituting the plant cell wall, the most abundant is cellulose,

composed of linear β-(1,4)-glucan chains. Together, these chains form a cellulose elementary fibril (CEF), which contains glucan chains packed in parallel and associated through extensive hydrogen bond networks. The exact number of CEF glucan chains and their geometric arrangement are still subjects of debate (Ding and Himmel 2006; Evert 2006; Ding et al. 2012.) Studies also show that CEFs aggregate into large bundles called macrofibrils and that the number of CEFs in a macrofibril varies among cell types and cell wall layers (Ding et al. 2012). Tightly packed and highly hydrogen bonded, cellulose polymers are not easily accessible to hydrolytic enzymes, presenting a significant challenge for biomass deconstruction.

Hemicelluloses. This class of branched or unbranched polysaccharides comprises a β-(1,4)-linked sugar backbone with short side chains consisting of a wide variety of sugar residues linked with different glycosidic bonds. Hemicelluloses sometimes also include sugar acids and noncarbohydrate subunits and contain both C5 (e.g., xylose) and C6 (e.g., glucose) sugars. For example, the hemicellulose xyloglucan has the same β-(1,4)-glucan backbone as cellulose along with side chains composed of xylose and other sugars. The β-(1,4)-glucan backbone may facilitate interactions with cellulose and serve as a bridge between CEFs. Hemicellulose branching and side groups also form covalent bonds with other cell wall polymers, such as pectin and lignins in lignified walls. These covalent bonds, and the interactions with cellulose, make hemicelluloses barriers to enzymatic access to cellulose, although they are much more amenable to enzymatic breakdown because of their noncrystalline state.

Lignin. The second most abundant polymer in plant biomass, lignin is a covalently linked heterogeneous composition of aromatic phenols. Despite its abundance, the structure of native lignin in a plant cell wall is poorly understood. In addition to providing mechanical support to plants, lignin is believed to be the major factor in biomass recalcitrance—the resistance of plant cell walls to microbial and enzymatic deconstruction—because it impedes enzymatic accessibility to the polysaccharide substrates (Zeng et al. 2014).

2.2 Major Challenges in Monitoring Plant Cell Walls

The changes that occur in lignocellulosic biomass structure and associations during and after pretreatments are not well understood, but this knowledge is essential for efficient deconstruction. Pretreatments that use different chemical and thermal conditions appear to change specific cell wall components in different ways—all increasing enzymatic digestibility, but the mechanisms are poorly understood. Identified changes include transitions in cellulose crystallinity and the rearrangement or removal of matrix copolymers with subsequent improvements in hydrolysis. Most current information pertaining to cellulose involves unpretreated samples, so characterizations of the transitions that occur following removal of disruptive agents, washing, drying, and rehydration are incomplete. Improved understanding requires technologies for atomic- and molecular-level investigations of the structure and dynamics of both naturally occurring and pretreated fibrous cellulose. Also poorly defined are structural transformations in lignin and interactions among lignin, hemicellulose, and cellulose during and after different pretreatments.

Understanding the Genetic and Molecular Basis of Plant Cell Wall Properties

Over the past 2 decades, field studies of plant cell wall biosynthesis have yielded valuable insights into the molecular mechanisms of synthesis and deposition of cellulose, lignin, hemicellulose, and pectin (Kalluri et al. 2014). However, critical questions in cell wall biosynthesis and biomass formation remain:

- What does the transition zone between primary and secondary walls look like?
- Which cellular processes lead to unique cell wall properties of a given type of biomass?

Cell Wall Composition and Degradation Needs

Effectively Modeling and Predicting Plant Cell Wall Systems

Technological improvements and innovative developments for monitoring molecular-scale structural changes in the cell wall (e.g., chemical-physical changes accompanying biopolymer deposition).

- *In situ* and *in vivo* approaches at atomic- and molecular-scale resolution to use with minimal sample preparation and in conditions closely resembling the natural environment.
- Effective modeling and prediction of spatiotemporally emergent plant system properties to improve the precision and pace of biomass research.

Characterizing Cell Wall Structure and Chemistry

Further development of imaging techniques and correlative approaches for conducting real-time visualization of plant cell wall biosynthesis (including secondary cell wall deposition, where most lignification occurs) and biomass deconstruction processes *in vivo* and *in planta*.

- Real-time, nanometer-scale imaging techniques to visualize the trafficking of cell wall synthases and modification enzymes and the dynamics of cytoskeletal networks.
- Techniques to characterize *in situ* the bioassembly of plant cell wall polymers, such as cellulose microfibril networks and interactions among cellulose, pectin, hemicellulose, and lignification.
- Improved methods to label proteins and increase throughput, combined with new omics technologies to improve the pace and precision of structure-function prediction.
- Synthetic biology approaches to incorporate deuterium into cell wall polymers in a controlled manner.
- Systems biology approaches to study plant cell wall biosynthesis, specifically relevant genes and their regulatory networks.

- What determines cellulose variations (commonly found in all plant cell wall types) and the accompanying cellulose composition?
- What controls the arrangement and turnover of protein complexes that synthesize wall polymers (i.e., cellulose synthase complexes) and the directionality of cellulose deposition?
- Do the spatial distribution patterns of hemicellulose and lignin polymers in the cell wall relate to patterns of transport and delivery of precursor-containing secretory vesicles?

Further elaborating the identity of intra- and interpolymer cross-linkages and the role of wall proteins and metals will be important, as will clarifying the extent of uniformity in cell wall architecture and wall polymer distribution along the cell boundary. Also of great interest is resolving the system dynamics that underlie and control achievement of defined cell wall phenotypes. These dynamics include changes in (1) membrane lipid dynamics, (2) metabolites and the local environment (e.g., shifts in H^+, Ca^{2+}, and reactive oxygen species in the apoplast, membrane, and cytosolic space), and (3) cellular signaling and regulatory events.

The Golgi apparatus and vesicle trafficking system are integral players in cell wall biosynthesis, which in turn determines biomass properties (see Fig. 2.1, p. 8). Golgi-derived vesicles carry both precursors of wall polymers (i.e., lignin and hemicellulose) and cellulose synthase complexes (CSCs) for cellulose synthesis in plasma membranes (Bashline et al. 2014). Research also shows that clathrin-mediated endocytosis (CME) underlies cellulose synthase (CesA) recycling and that CSC turnover affects the degree of cellulose

polymerization and patterning (McFarlane et al. 2014). Despite these findings, there is a critical lack in the understanding of cell wall formation occurring in these organelles and vesicles. Molecular and cellular (mesoscale) technologies are needed to fill these knowledge gaps. On the molecular front, the ability to monitor trafficking of vesicles and identify their contents *in situ* and in real time, while simultaneously detecting changes in cell wall properties, will open up the next frontier of cell wall research. Extending the resolution of live-cell imaging at single-molecule resolution—such as the recently demonstrated two-color nanoscale imaging of intracellular targets (Bottanelli et al. 2016)—will help researchers examine the range of bioenergy-relevant biomolecules in plant cells.

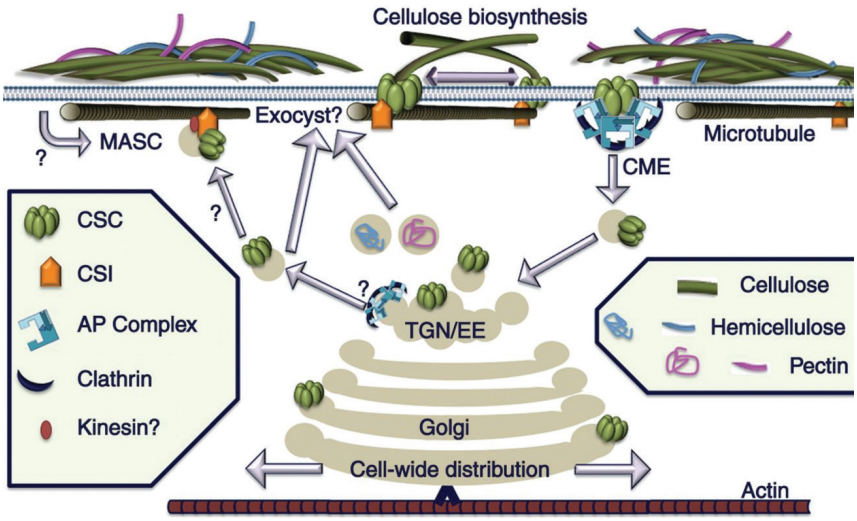


Fig. 2.1. Overview of the Localization and Proposed Trafficking Pathways, Compartments, and Mechanisms of Cellulose Synthase Complexes (CSCs). Although pectin and hemicellulose are synthesized in the Golgi apparatus and secreted to the cell wall, cellulose is exclusively synthesized by CSCs located at the plasma membrane. CSCs are thought to be assembled in the Golgi apparatus, which is responsible for the actin-dependent, cell-wide distribution of these complexes. Researchers believe CSCs are secreted through the trans-Golgi network/early endosome (TGN/EE) and may be partitioned into a specific domain within it. [Image reprinted from Bashline L., S. Li, and Y. Gu. 2014. "The Trafficking of the Cellulose Synthase Complex in Higher Plants," *Annals of Botany* **114**(6), 1059–67, by permission of Oxford University Press.]

Key: AP, adaptor protein; **CME**, clathrin-mediated endocytosis; **CSI**, compound structure identification; **MASC,** macrotubule-associated cellulose synthase compartments.

Multimodal Biomass Imaging Methods

Traditionally, biomass composition is analyzed by wet chemistry (Sluiter et al. 2013). The research community has used many other analytical methods to characterize plant cell wall structure, such as (1) electron microscopy (Ohad et al. 1962; Ohad and Danon 1964; Ha et al. 1998), (2) ^{13}C solid-state nuclear magnetic resonance (NMR; Ha et al. 1998; Sturcova et al. 2004), (3) X-ray diffraction, (4) small-angle neutron scattering (SANS; Sugiyama et al. 1998), and (5) Fourier transform infrared (FTIR) spectroscopy (Sene et al. 1994).

Advances in some technologies particularly applicable to plant cell walls include:

- Atomic force microscopy (AFM). A scanning probe microscopy imaging technique offering sub-nanometer resolution of surface topography under aqueous conditions (see Fig. 2.2, p. 9).
- Advanced transmission electron microscopy (TEM) techniques. Scanning TEM for generating nanometer-resolution tomographic reconstructions of plant cell wall sections (~1 micrometer thick), both before and after pretreatment are under development.
- Soft X-ray microscopy and Fourier transform infrared (FTIR) microscopy. Two techniques for

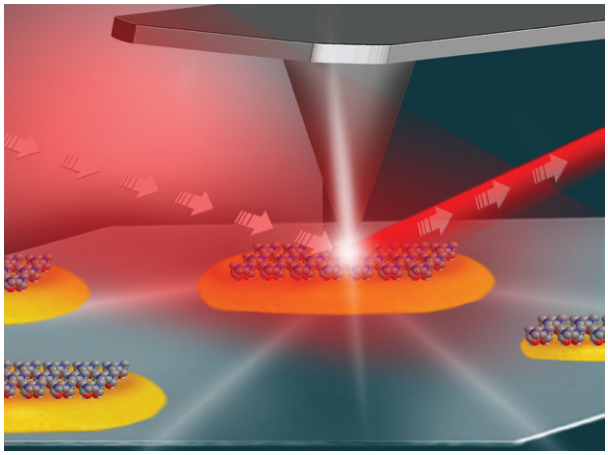


Fig. 2.2. Infrared Light Scattered off a Metallic Atomic Force Microscope Tip. The combination of synchrotron infrared radiation with scattering, scanning near-field optical microscopy (s-SNOM) enables infrared spectroscopic investigations with ~20 nanometer spatial resolution. This synchrotron infrared nanospectroscopy (SINS) technique can be applied to hard and soft matter, including biominerals, proteins, bacteria, fungi, and other biomaterials to identify and measure local surface properties (instead of chemistry). The illustration shows the setup for an experiment at Lawrence Berkeley National Laboratory's Advanced Light Source, in which SINS measurements identified site-dependent reactivity of chemically active molecules that were anchored to the surface of metallic nanoparticles. [Reproduced with permission from Elad Gross, The Hebrew University of Jerusalem, Israel; from *Faraday Discussions* **188**, 345–53, with permission from The Royal Society of Chemistry.]

providing needed spatial information about the distribution of lignocellulosic polymers.

- Nano-Secondary Ion Mass Spectrometry (Nano-SIMS). Imaging technique enabling elemental compositional analysis of samples with a resolution of ~50 nanometers.
- Coherent Raman scattering (CRS) and synchrotron infrared nanospectroscopy (SINS). Two techniques providing chemical information at the level of functional groups with ~20 to 30 nm resolution (Bechtel et al. 2014; Amenbar et al. 2017).

The ability to overlay results of these emerging techniques is promising, yielding high spatial, temporal, and chemical information on plant cell walls at the molecular (nano-) through cellular (micro-) scales. Further development of these correlative imaging techniques will allow real-time visualization of plant cell wall biosynthesis and biomass deconstruction processes *in vivo* and *in planta.* Techniques that can co-register and overlay chemical, structural, and biomechanical characteristics will be essential in closing existing knowledge gaps in biomass structure and composition and in supporting the development of more effective cell wall utilization methods (see Fig. 2.3, p. 10, and Fig. 2.4, p. 11). Recently demonstrated multimodal chemical imaging technologies include the Hybrid Photonic Mode-Synthesizing Atomic Force Microscopy (HPFM), which combines the disciplines of nanospectroscopy, nanomechanical microscopy (10 to 50 nm), and AFM with Raman, fluorescence, and infrared imaging. Another example is multimodal imaging with mass spectrometry (MS) that provides spatially resolved molecular and elemental mass–based chemical imaging. Newly available technologies include matrix-assisted laser desorption ionization (MALDI) time-of-flight mass spectrometry (TOF-MS), secondary ion mass spectrometry (SIMS), helium ion microscopy with SIMS (HIM-SIMS), and SINS. A number of the methods proposed for measuring metabolites at the cellular level also will be important for multimodal biomass imaging (see Chapter 5. Metabolic Pathways in Plants, Microbes, and Fungi, p. 37).

Multiscale Computational and Modeling Approaches

As a complement to imaging and spectroscopic analyses of plant cell wall polymers, multiscale computational approaches, such as those below, can play a critical role in linking cell wall structure and dynamics to natural and engineered changes in composition.

- At the atomic level, quantum chemical calculations yield assessments of the strength of

association between biomass components such as cellulose and lignin.

- At the molecular level, molecular dynamics simulations, both atomistic and coarse grained, can provide information on plant polymer structure at nanometer to micrometer resolution and dynamics at nanosecond- to microsecond-length scales.

Scientists also can use such simulations to predict the effects of thermochemical pretreatment. Molecular simulation is highly complementary with neutron and X-ray scattering experiments that probe similar time and length scales. High-performance computing extends the time and length scales accessible to high-resolution physical simulations. Furthermore, lower-resolution, finite-element methods can help characterize the elastic and viscoelastic responses of plant cell walls, important for cell expansion during plant growth.

Generating a basic model of a plant cell is the unifying goal in developing and applying technologies to probe molecular and phenotypic responses in the same or similar plant cells. Achieving this goal will entail defined monitoring, visualizations, models of plant cell wall biosynthesis and deposition, and remodeling within a given plant cell type in optimal growth conditions and in a time-resolved manner. Developing this basic model, where none exists today, will open doors for independent model iterations. The design of these models will need to accommodate cell-type heterogeneity within a given plant, species-level distinctions, and dynamics in response to both internal cues (e.g., developmental and physiological) and external ones (e.g., biotic and abiotic stressors).

Capabilities are needed to monitor (i.e., image) molecular changes (e.g., in nucleic acids, proteins, protein complexes, and metabolites) in the context of fine-scale phenotypic changes such as chemical-physical shifts accompanying biopolymer deposition. These developments will be powerful in accurately co-registering information for effective use in modeling efforts. Also needed are advancements enabling these technologies to probe cells *in situ* and *in vivo*. Finally, addressing discovery science–driven goals will require expanding

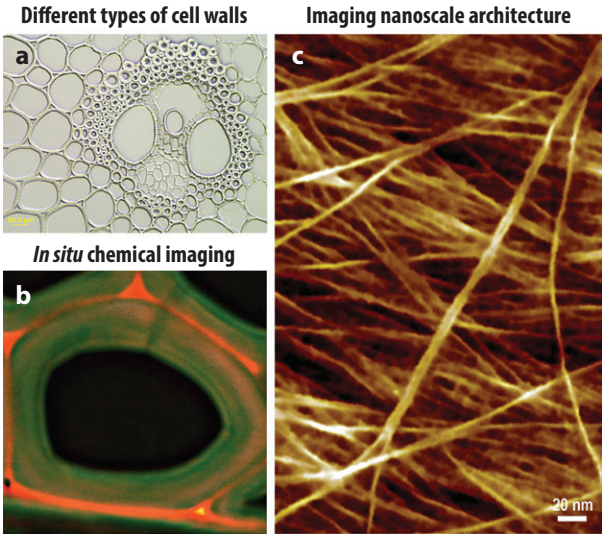


Fig. 2.3. Exploring Plant Cell Wall Architecture and Chemistry. Plant cell walls are structurally and chemically complex at the mesoscale and nanoscale. This complexity can be measured by spatial and chemical imaging techniques. At the tissue level, these approaches include optical microscopy, which can provide diffraction-limited spatial resolution, and micro-spectroscopy for *in situ* chemical resolution. Other analytical techniques such as mass spectrometry, nuclear magnetic resonance, and X-ray scattering can reveal specific physical-chemical information of bulk biomass. At the subnanometer scale, atomic force microscopy and electron microscopy can image the cell wall. Beyond these technologies, correlative and nondestructive imaging techniques are needed to improve the understanding of the structure and chemistry of plant cell walls, as well as their biosynthesis and bioconversion processes. [**(a)** Image courtesy Shi-You Ding, Department of Plant Biology, Michigan State University from Ding, S., et al. 2014. "Size, Shape, and Arrangement of Native Cellulose Fibrils in Maize Cell Walls," *Cellulose* **21**(2), 863–71, with permission of Springer. **(b)** Reprinted from Zeng, Y., et al. 2014. "Lignin Plays a Negative Role in the Biochemical Process for Producing Lignocellulosic Biofuels," *Current Opinion in Biotechnology* **27**, 38–45, with permission from Elsevier. **(c)** From Ding, S., et al. 2012. "How Does Plant Cell Wall Nanoscale Architecture Correlate with Enzymatic Digestibility?" *Science* **338**(6110), 1055–60. Reprinted with permission from AAAS.]

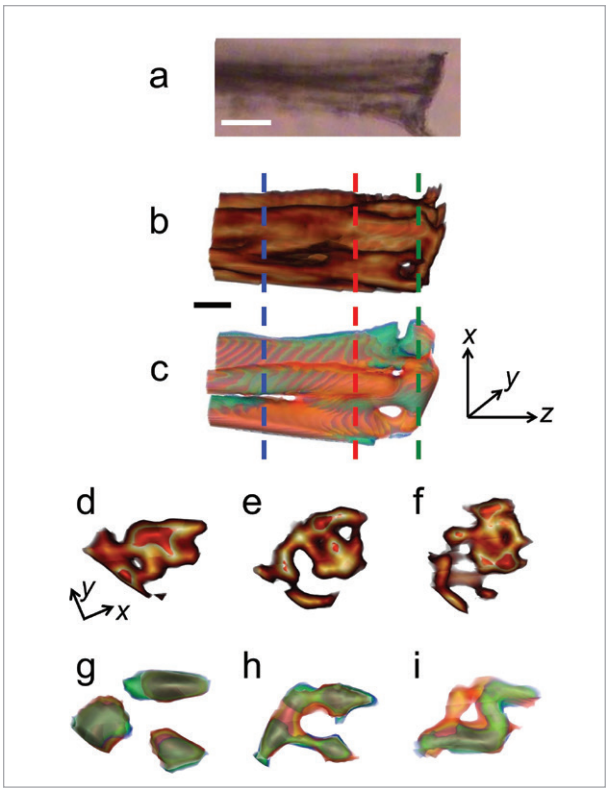


Fig. 2.4 Three-Dimensional (3D) Tomographic Reconstructions of *Populus* Wood Using Synchrotron Fourier Transform Infrared Spectro-Microtomography. Elucidating the 3D molecular architecture of plant cell walls is one of the most challenging problems in plant biology, and the deconstruction of lignified cell walls is a critical step in converting biomass to liquid biofuels and other value-added products. **(a)** Bright-field image of plant biomass. **(b, d–f)** The brown in these tomograms represents reconstructed intensities of hydrocarbon stretching absorption modes. **(c, g–i)** The red in these images is spectrally associated with lignin superimposed on the blue-green colors associated with holocellulose. Panels d–i are virtual slices 10 micrometers (µm) thick across the three longitudinal vessels of this specimen at locations indicated by the dashed lines. Holocellulose is more prominent in the middle of the wall, whereas lignin dominates around the exterior of the wall and middle lamellae between vessels. (Scale bars, 20 µm). [Reprinted by permission from Macmillan Publishers Ltd: Martin, M. C., et al. 2013. "3D Spectral Imaging with Synchrotron Fourier Transform Infrared Spectro-Microtomography," *Nature Methods* **10**, 861–64.]

phenotypic databases with genotype-to-phenotype correlative analyses to enable early identification of the molecular and phenotypic properties that can effectively predict emergent properties.

Characterizing Cell Wall Structure and Chemistry

Development of Correlative Nondestructive Methods. The ability to track single-molecule behavior has changed the fundamental approach to studying biological processes in cell walls. Further development of correlative imaging techniques will allow real-time visualization of biosynthesis and biomass deconstruction processes *in vivo* and *in planta*. Early approaches to single-molecule imaging in biology primarily have focused on fluorescence-based microscopy, in which a fluorophore—such as a dye, quantum dot, or fluorescent protein—is chemically or genetically tagged to a molecule of interest and deterministic or stochastic super-resolution techniques track the molecule in two or three dimensions (2D or 3D). However, labeling potentially can interrupt the biomolecule's functions, especially in complex systems such as biomass conversion reactions. The desire to analyze at nanometer-scale resolution, with minimal sample preparation, and under natural conditions excludes most high-resolution electron microscopy techniques. These techniques, however, are completely amenable to AFM and the recently developed stimulated Raman scattering (SRS) microscopy for real-time visualization of the reaction interface and *in situ* mapping of cell wall chemistry, respectively.

AFM Investigations of the Cell Wall Surface. Although limited in application to images of a substrate's surface, AFM is especially advantageous for studying cell wall accessibility and digestibility and thus is a powerful tool for mapping surface

properties. In addition, the cantilever tip (<5 nm) is smaller than most enzymes. As demonstrated in the literature (Ding et al. 2012), surface properties measured by AFM possibly are correlated to enzyme accessibility. Notably, researchers can apply AFM imaging under nearly the same physiological conditions as enzymatic digestion, meaning that cell wall surface interactions are essentially the same with both the AFM tip and the enzymes.

NMR and Polymer Structure and Bonds. NMR can provide detailed molecular information about lignin and polysaccharide structures and detect covalent linkages between these two biopolymers (del Río et al. 2016). Both solution- and solid-state NMR frequently are used to monitor changes resulting from biomass pretreatment (Trajano et al. 2013; Petersen et al. 2014). Carbon-13 labeling of whole plants allows sensitive acquisition of 2D solid-state carbon-carbon correlation spectra of cell walls, enabling detailed study of cell wall architecture (Wang and Hong 2016). These studies have shown that cellulose microfibrils are cross-linked more through pectin than through xyloglucan (Dick-Perez et al. 2011) and have helped determine the number of cellulose chains that make up a microfibril (Wang and Hong 2016). Comprehensive multiphase NMR—a combination of liquid, gel, and solid-phase NMR—can investigate intact plants, avoiding the need for any type of extraction (Wheeler et al. 2015). This approach uses filtering techniques to observe components present separately in the three phases.

MS Techniques for Cell Wall Characterization. Several MS-based approaches provide chemical and structural information on cell wall polymers. Techniques such as negative and positive ion electrospray (ESI) tandem mass spectrometry (MS/MS) can analyze the sequencing and linkages of underivatized oligosaccharides obtained from partial depolymerization of several α- and β-glucans (Palma et al. 2015). ESI-MS/MS also can structurally characterize branched hemicellulose oligosaccharides (Quemener et al. 2015). Ion mobility–MS, in combination with ESI-MS, is able to separate and distinguish closely related hemicellulose oligosaccharide isomers (Plancot et al. 2014). Researchers can use MALDI-MS imaging to identify and localize cell wall polysaccharides on biomass samples by performing partial enzymatic digestions directly on the biomass, without translocation of the released oligosaccharides. They then can detect and characterize these polymers by MALDI-MS and localize them at micrometer resolution (Veličković et al. 2014; Veličković et al. 2016). Finally, scientists can determine lignin structure and its changes, including ratios of constituent monolignols and of C5 and C6 sugars, in a top-down approach and in high-throughput fashion by several means. They include pyrolysis–molecular beam MS, ESI-MS, atmospheric pressure photoionization (APPI), MALDI-MS, and MS/MS (Kelley et al. 2002; Banoub et al. 2015).

Nondestructive Imaging Using Neutrons. Many of the techniques described above can be used to provide information about structural changes in lignocellulose components before and after a pretreatment regime. However, to best understand the morphological changes that occur during pretreatment, structural analysis must be performed *in situ*. Neutron scattering techniques are ideally suited to this task because they can probe the length and time scales relevant to lignocellulose characterization (see Fig. 2.5, p. 13). Moreover, because neutrons do not destroy delicate biological samples, scientists can study structural changes to biomass *in situ* and in real time using a specialized reaction cell. A unique property of neutron scattering is that it enables contrast variation techniques that allow separation of scattering contributions from different components within intact lignocellulose. This process occurs through the controlled replacement of hydrogen with its isotope deuterium. Developing synthetic biology approaches to incorporate deuterium into cell wall polymers in a controlled manner would provide new details about the structural and temporal rearrangement of lignocellulose during pretreatment. For example, feeding deuterated variants of specific metabolic intermediates such as GDP-fucose or GDP-rhamannose may produce

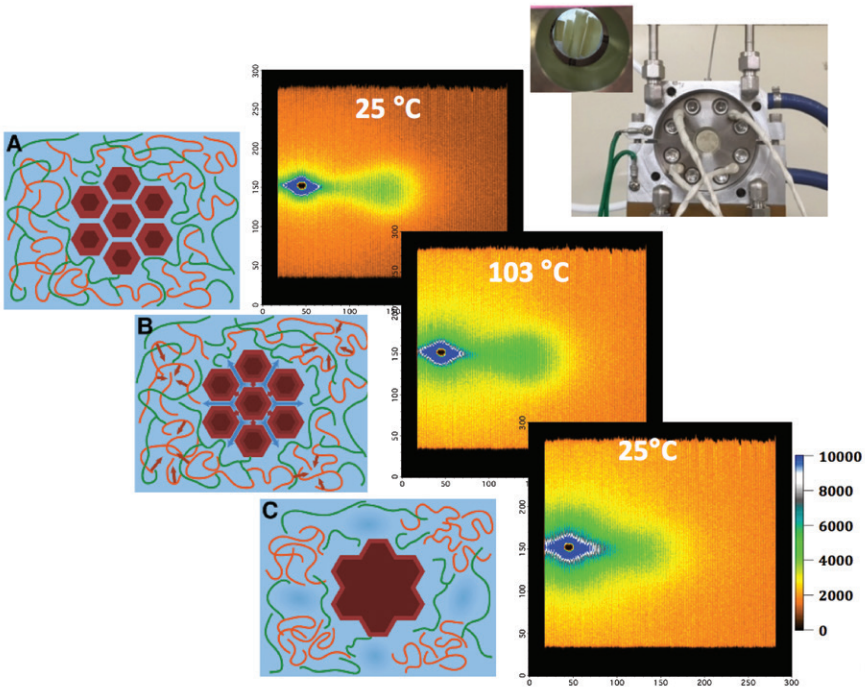


Fig. 2.5. Real-Time Small-Angle Neutron Scattering (SANS) Supported by Molecular Dynamics Simulation for Identifying Processes Occurring During Biomass Thermochemical Pretreatment. (Left panels) Illustration of changes to lignocellulose. Changes in cellulose morphology (brown hexagons), lignin (red chains) aggregates, and hemicellulose (green chains) occur during the pretreatment process. **(Middle)** Two-dimensional SANS images. **(Right)** *In situ* SANS reaction cell. [**(Left)** Pingali, S. V., et al. 2014. "Morphological Changes in the Cellulose and Lignin Components of Biomass Occur at Different Stages During Steam Pretreatment," *Cellulose* **21**(2), 873–78, with permission of Springer. **(Left)** Reproduced from Langan, P., et al. 2014. "Common processes drive the thermochemical pretreatment of lignocellulosic biomass," *Green Chemistry* **16**(1), 63–68, with permission from The Royal Society of Chemistry.]

specific labeling of Fuc and Rha polymers, which then can be probed with neutrons.

Techniques for Characterizing Holoproteins and Their Interactions. Cellulytic enzymes are typically multidomain proteins with well-folded domains separated by flexible linkers. Although X-ray diffraction and NMR have provided atomic-resolution structural information about the ordered domains of these proteins, only low-resolution information, obtained using small-angle X-ray and neutron scattering, is available for holoproteins. Recent advancements in cryo-electron microscopy (or cryoEM) techniques (see Chapter 6. Biosystems Design, p. 45) with new direct electron detector technology, have enabled (1) the determination of increasingly complex systems at near-atomic resolution and (2) the characterization of different conformational states of biomolecules resulting from conformational heterogeneity in samples. This technique has the potential to overcome limitations of other techniques and provide new insights into the structure of complex cellulolytic complexes such as the bacterial cellulosome.

New approaches to Investigate How Cellulases Interact with Insoluble Substrates. Typically, studies are limited to using soluble oligosaccharides or indirect approaches such as binding assays. Methods to prepare more realistic polysaccharide substrates with minimal structural heterogeneity will enhance structural and functional studies. SANS with contrast variation and biomolecule deuterium labeling has the potential to provide structural insight into cellulase-cellulose interactions. Researchers also could use this approach to understand the mechanism of cellulase inhibition by nonproductive interactions with lignin and other compounds. Such studies could yield more-specific

structural information on enzyme-lignin-cellulose interactions needed to elucidate these mechanisms.

Improved Methods for Elucidating Interaction Mechanisms. Molecular interactions, such as enzyme with substrate or enzyme with donor can be probed by saturation transfer difference NMR spectroscopy (STD) or transferred nuclear Overhauser effect spectroscopy (tr-NOESY) NMR (Marchetti et al. 2016). In STD, selected protons throughout the protein are magnetically saturated. This saturation then is transferred to the carbohydrate ligand, leading to reduced signal intensity of protons that interact with the protein and thus enabling elucidation of the interaction mechanism between binding partners. The use of tr-NOESY allows detection of NOE contacts present only in the bound state, thus characterizing the conformation of the interacting oligosaccharide. Methods to prepare the necessary labeled proteins are known but could be improved along with greater spectral sensitivity to increase the throughput of these approaches.

Sequence-to-Function Continuum. To accelerate the pace of biomass improvement efforts, new genomic technologies are needed to manipulate plant systems at the molecular, cellular, tissue, and organismal levels. Genome and transcriptome sequencing technologies are yielding a wealth of information, yet it is only partially annotated or interpreted for function. Greater pace and precision would enable prediction of (1) the functional consequence of gene sequence variation on cell wall and biomass properties and (2) sustainability traits based on a comprehensive protein sequence-structure-function knowledgebase. Key to maximizing the functional interpretation of genome sequences and sequence variation is to develop and make available molecular and mesoscale technologies that can assist in understanding protein structure, activity, and function. Progress in this area could advance significantly from the iteration of computational prediction and simulation methods with experimental characterization and verification. New technologies would transform the current research landscape, enabling characterization of plant cell wall biosynthesis and remodeling pathways based on structural biology and multiplexed protein assay platforms that include multiplexed protein-ligand binding, enzyme-substrate conversion, and single-molecule protein pull-down (SiMPULL; Jain et al. 2012).

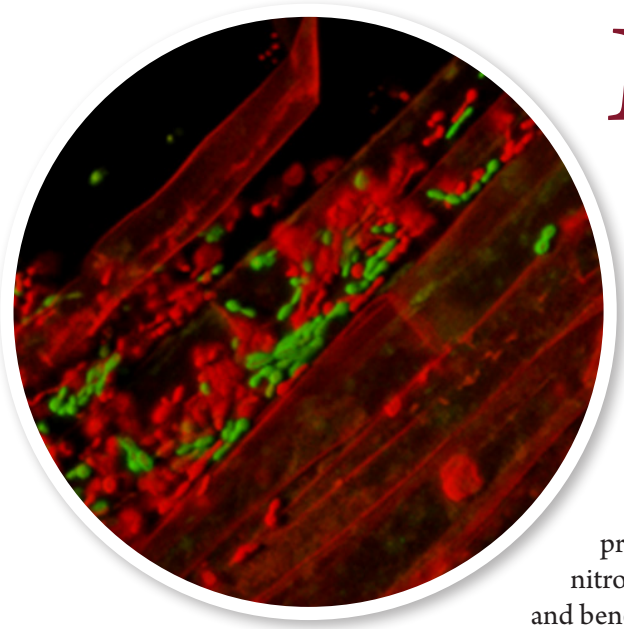
2.3 Advancing Biofuel and Bioproduct Design and Production

A deeper understanding of plant cell wall architecture at the molecular to mesoscale levels will provide knowledge of the structure and dynamics of native biomass and the response of its components to pretreatment regimes and subsequent enzymatic processing. To achieve this understanding, scientists must develop approaches that combine disparate datasets from analytical chemistry, direct and indirect imaging techniques, and spectroscopies. Incorporating experimental datasets in computational cell wall models will overcome the knowledge gap that exists in understanding the structure of the plant cell wall and the critical structural rearrangements that increase biomass digestibility for biofuel and bioproduct production. These advances will transform the current pace and depth of understanding of plant biomass improvement efforts.

Genomic studies have identified many genes involved in plant cell wall biosynthesis and modification. However, the regulation of these genes and how they communicate (i.e., interact) with environmental factors are still poorly understood. Research in understanding biosynthesis and trafficking of cell wall components is critical to enable the rational design and production of improved biomass. Achieving this understanding requires new tool development for visualizing *in vivo* and real-time biosynthesis events, such as deposition of cellulose microfibrils, secretion of hemicelluloses, and transformation of lignin monomers and lignification. Also needed is cell imaging with single-molecule resolution, which will enable visualization of a range of cellular biomolecules. These capabilities will open up the next frontier of cell wall research and lead to technological improvements in bioenergy feedstock plants.

Chapter 3

Rhizosphere Community Interactions



Microbes play critically important roles in the environment, shaping plant health and productivity, the terrestrial carbon cycle, and environmental remediation. Although the ability to identify the organisms involved in these processes has significantly improved with modern sequencing technologies, characterizing microbial activities and their interactions with plants, viruses, soil fauna, and the abiotic environment remains a significant challenge. These interactions are particularly impactful in the *rhizosphere*, the biologically active area immediately surrounding plant roots. Rhizosphere microbiomes (consisting of bacteria, fungi, archaea, protists, and phages) include diverse taxa such as nitrogen fixers and crop pathogens, mycorrhizal fungi, and beneficial bacteria that enhance plant nutrient acquisition and drought tolerance. Stimulated by root exudates and root decay, rhizosphere organisms interact to move carbon from root tissues to the surrounding soil, a process that ultimately regulates both soil carbon stabilization and ecosystem processes such as trace gas production. Plant-microbe interactions occurring within plant tissues and on leaf surfaces also play important roles in plant health. However, knowledge of plant-microbe interactions is constrained to only a few model systems. Little is known about interactions among root-associated microbes and even less about their interactions with other members of the soil food web (e.g., fungi, fauna, and phages; see Fig. 3.1, p. 16).

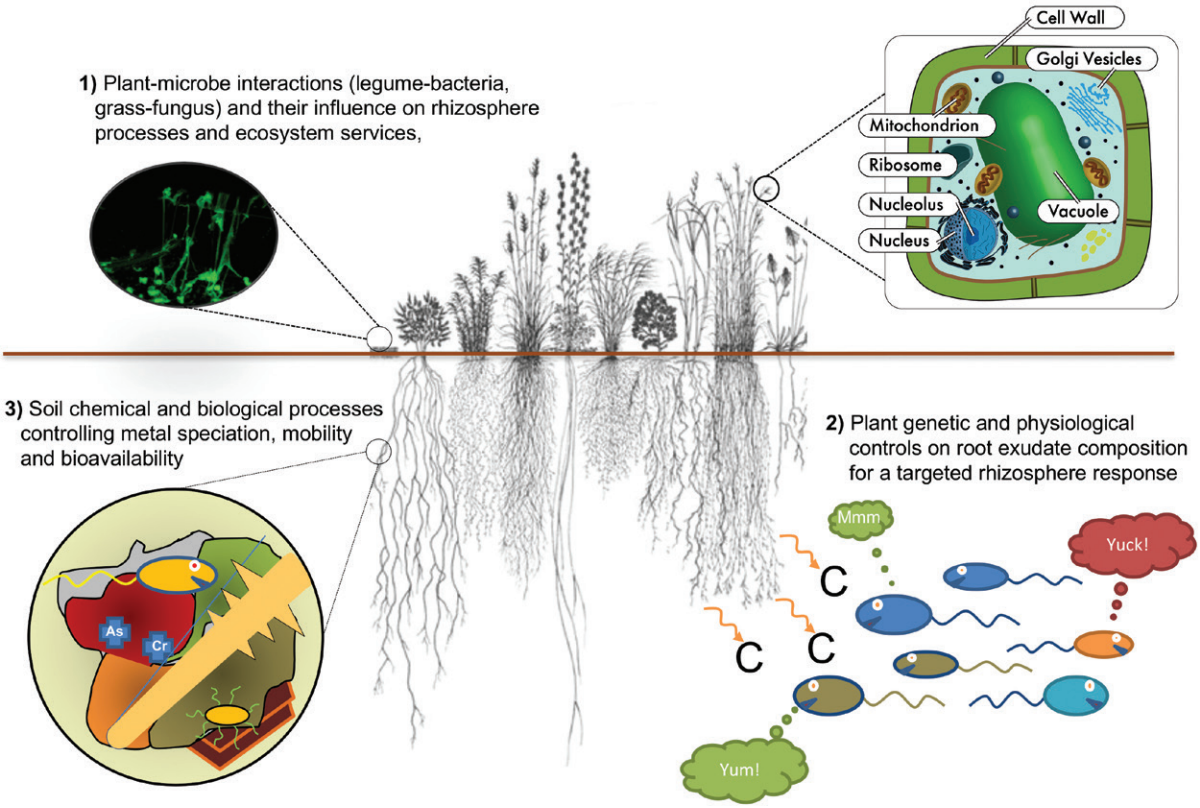


Fig. 3.1. Significant Effects of Plants, Microbes, and Their Communities on Biogeochemical Transformations at Multiple Scales. New methodologies and technologies are needed to inform a wide range of mission-relevant questions for the U.S. Department of Energy Office of Biological and Environmental Research. [Image adapted from David McNear, University of Kentucky Rhizosphere Science Laboratory.]

3.1 Current Science and Technologies

Rhizospheric microbial communities drive fundamental processes in the global carbon cycle and regulate levels of atmospheric carbon dioxide (CO_2) and soil carbon storage. Microbiome research seeks to define community membership, ecological relationships among organisms, and the roles specific taxa play in systems-level chemical and biological processes. Regardless of their habitat, microbiomes comprise many different taxa, exploiting an energy source, yet these microbial assemblages often are inherently interdependent and dynamic in both space and time. The comprehensive understanding of *in situ* microbiome ecology has become tantalizingly possible with the advent of high-throughput sequencing, advanced microscopy, and stable-isotope tracing techniques. However, current microbiome studies often are highly descriptive, focused on correlation patterns or simple one-on-one interactions between culturable organisms.

This chapter describes current and needed technologies that address (1) microbe-microbe and plant-microbe interactions and their influence on rhizosphere processes and ecosystem services that benefit humankind, (2) plant genetic and physiological controls on root exudate composition and beneficial interactions with microbial symbionts, (3) soil chemical and biological processes, and (4) interkingdom interactions (e.g., algae-bacteria,

Rhizosphere Community Interactions Needs

Revealing and Monitoring Plant and Microbe Interactions

Improved biofuel and bioproduct production with sustainable agriculture practices and better comprehension of both beneficial and harmful plant-microbe interactions for making quantitative, *in situ*, and three-dimensional measurements of dynamic molecular phenomena with nanometer to centimeter resolution.

- New ways to collect plant exudates and metabolites under realistic conditions.
- Reference databases to help identify the detected transcripts, proteins, and metabolites.
- Methods to noninvasively monitor root growth in a field setting over a growing season and root impacts on soil carbon or water stocks.
- Efforts to determine minimal necessary biological information (e.g., key microbial processes and critical environmental drivers) to parameterize soil models.
- Biotechnologies to enable persistence of relationships between plants and growth-promoting microbes in field settings.

Improving Image or Sensor-Based Systems to Monitor Root-Microbe Interactions

Improved image- or sensor-based systems for monitoring root exudates and rhizospheric microorganisms *in situ* at biologically relevant length and time scales for revealing deeper insights into interdependencies within the rhizosphere.

- Methods to monitor root-microbe interactions in real time in the soil.
- Novel frameworks to perform ^{13}C metabolic flux analysis for consortia systems having more than three members and harnessed for nonmodel rhizospheric systems.
- Reference metabolism databases to use for nonmodel microbes in the rhizosphere.
- Improvements in detector technologies and optics to increase spatial resolution to less than 10 micrometers for cellular-level imaging.
- Novel contrast agents to enable new experiments for following dynamic processes in plants and their associated rhizosphere.
- Genetic tools to study systems with no available draft genome and with polyploid conditions that limit genetic-modification effectiveness.

fungi-bacteria, and bacteria-phage). Many of these approaches seek to link the identity of uncultivated microbes with their potential to metabolize compounds in the environment—a topic that remains a "grand challenge" area for the field of microbial ecology (Neufeld et al. 2007).

3.2 Major Challenges in Detecting Rhizosphere Interactions

Capturing Molecular Exchanges Underpinning Plant-Microbe Interactions

Critical research in environmental microbiology and phytobiome studies hinges on questions such as, "*Which organisms live where, and what do they do*?" However, obtaining simultaneous spatial, temporal, chemical, and phylogenetic information for native microbial communities is extremely difficult, particularly in opaque systems such as the soil or faunal guts. To better comprehend both beneficial and harmful plant-microbe interactions, researchers need capabilities to make quantitative, *in situ*, and three-dimensional (3D) measurements of dynamic molecular phenomena with nanometer to centimeter resolution. These technologies are needed to discover the ephemeral molecular signals serving as communication pathways between plants and their microbial partners. While measurement of

intra- and intercellular fluxes is currently possible in simple mixtures, such as between microalgae and bacteria (Hom et al. 2015), novel approaches are needed to extend this capability into the rhizosphere and capture the molecular exchanges that underpin multitrophic interactions (e.g., soil food webs).

Particularly in the rhizosphere, these interactions often are mediated by competition, cooperation, or interconnected metabolisms involving mixed populations of microbes, fungi, protists, microfauna, and phages. The use of metagenomic sequencing provides the genomic profile of all members of such a communal system. Recent advances in metagenomics have enabled researchers to identify and "count" members within these communities and suggest a system's genetic potential, but without quantitative and temporal resolution. Approaches such as metatranscriptomic sequencing and metaproteomic or metabolite profiling may reflect current community activity, but they continue to be limited by the quality of reference databases, which are needed to help identify the detected transcripts, proteins, and metabolites through comparison with known samples. Another key challenge in this area is the identification of a community's eukaryotic components (e.g., protists, fungi, and microfauna), whose genomes often are large, difficult to assemble, and highly underrepresented in reference databases.

Genomic, transcriptomic, metabolomic, and proteomic methods have advanced considerably, but they still are most often applied to single-microbe, cultured systems. Results from these data streams need to be more effectively integrated and their predictions verified through manipulative experiments, metabolic modeling, and quantitative observations of metabolic fluxes within and between organisms (via imaging or biochemical characterization). Natural microbiomes are teeming with a mixture of organisms that interact not only with each other, but at surface interfaces, in environmentally sensitive zones, and at points of biocatalysis (i.e., enzymatic transformation of organic compounds). Although conducting "bulk" measurements is sometimes desirable for mixed microbial systems, determining interaction mechanisms among the system members often requires measuring reactions at key biological interfaces in real time, with high spatial resolution [1 micrometer (μm) to 10 mm]. For example, an ideal scenario would involve discovering the repertoire of enzymes and the composition of chemical breakout products generated by saprotrophic fungi and bacteria during lignocellulosic breakdown of a decaying root. Similarly, scientists can derive knowledge of plant metabolites primarily from highly simplified hydroponic systems, but plants grow naturally in a soil matrix where metabolites of interest likely are intercepted by the surrounding microbiome before they can be collected. Thus, new ways are needed to collect plant exudates and metabolites under realistic conditions.

Understanding the Soil Carbon Cycle to Aid Sustainable Biomass Agriculture for Bioenergy Production

Microbial communities drive fundamental processes in the global carbon cycle and regulate levels of atmospheric CO_2 and soil carbon storage. Underlying these activities is the ability of microbiota to aggregate newly formed minerals and oxidized residues of microbial biomass to create a structured, porous matrix that serves as a reactor for cycling matter and as a reservoir of nutrients and water. However, a mechanistic, genomically informed molecular understanding of the fate and turnover of soil organic matter (SOM) is lacking. Myriad decomposition processes shape SOM—originating from root exudates, root turnover, and microbial and invertebrate necromass. These processes determine nutrient availability and the *tilth* of a soil, a critical yet difficult-to-measure trait reflecting soil structure, aggregation, and overall suitability for agriculture. Efforts to make agricultural practices more sustainable for biofeedstock production, and thus bioenergy endeavors, require *in situ* measurements of soil tilth attributes and new technologies that can track SOM from its source to its fate with high molecular, spatial, and temporal resolution.

KBase Narratives

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase
kbase.us

Within the DOE Systems Biology Knowledgebase (KBase), computational experiments are captured in dynamic, interactive documents called Narratives that promote collaboration and reproducibility of scientific results. In addition to all the data, parameters, analysis steps, and output reports associated with an experiment, Narratives can include user images, notes, and commentary. The KBase system also maintains Narrative provenance and versioning, tracking a user's edits and app re-executions, along with resulting updates in the underlying data. Narratives can be kept private, shared with colleagues and collaborators, or made public for the benefit of the research community. Because the Narrative Interface is built on the Jupyter Notebook, users can write custom scripts in their Narratives. A selection of KBase Narratives is showcased in the Narrative library (kbase.us/narrative-library/).

Optimizing System Effects of Plant Genotypes

Optimizing plant feedstocks for bioenergy and sustainable agriculture also requires a better comprehension of the genetic underpinnings of plant gene–environment interactions, particularly for root architecture and function. Although the architecture of the root system is widely known to have a strong genetic component, its topology, cell morphology, and function can be very responsive to environmental drivers and the surrounding microbial holobiome. Currently, there are no means of noninvasively monitoring root growth over a growing season in a field setting or monitoring root impacts on soil carbon or water stocks. Other needs are biotechnologies that enable persistence of relationships between plants and growth-promoting microbes in field settings (see Chapter 6. Biosystems Biodesign, p. 45). Well-designed mutualistic plant-microbe interactions could improve plant nutrient and water acquisition and stress tolerance while simultaneously promoting carbon stabilization, soil fertility, and a diverse soil biota.

Scaling Data from Genomics to Ecosystem Processes

Researchers currently cannot model microbial activities at a scale useful for predicting responses to environmental changes such as temperature variation, drought, and land-use shifts. Some have proposed using trait-based modeling approaches to represent microbial genetic capacities in ecosystem-scale carbon cycle models, and early results suggest significant positive effects on the predicted responses of soil carbon (Wieder et al. 2013; Hagerty et al. 2014). Additional efforts are needed to determine the minimal biological information (e.g., key microbial processes and critical environmental drivers) necessary to parameterize soil models. An example of current capabilities in modeling a community's metabolic networks and improving the understanding of them is what the DOE Systems Biology Knowledgebase (KBase; see box, this page) calls a Narrative (narrative.kbase.us/narrative/ws.13807.obj.1). Presented at the meeting was an example of a KBase computational workflow associated with a publication (Henry et al. 2016) that tests the use of community-level data for network reconstruction for interspecies interactions when members are not well characterized and cannot be experimentally isolated.

3.3 Major Needs in State-of-the-Art Technologies for Understanding Rhizospheres

Acquiring New Genetic Tools for Recalcitrant Organisms

Unfortunately, the vast majority of microbes within environmental consortia cannot be cultured, leading to a lack of understanding of how microbial communities are structured and evolve (Solomon et al. 2014) and limiting the research community's ability to harness these microbes for bioenergy applications through genetic manipulation (see Fig. 3.2, p. 20). With the advent of approaches enabled by clustered regularly interspaced short palindromic repeats

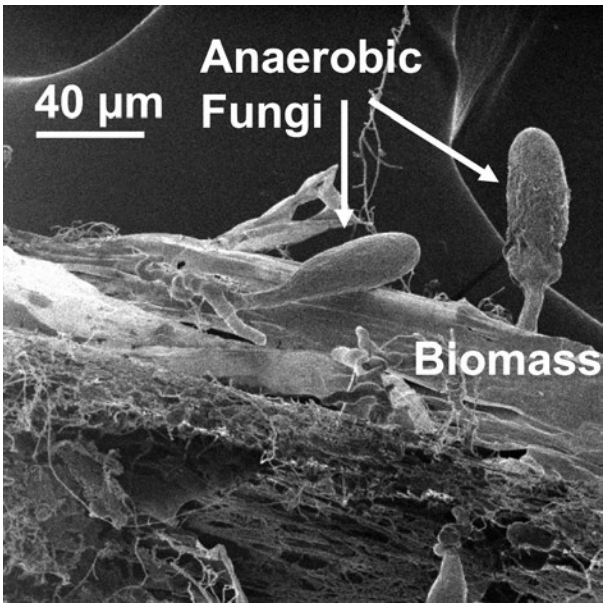


Fig. 3.2. Anaerobic Fungi Isolated from the Digestive Tract of Large Herbivores. These fungi thrive on the hydrolyzed sugars from crude biomass, potentially an aid in bioenergy production. As yet, there are no genetic tools that can modify the function of anaerobic fungi because of a lack of genomic information coupled with the formidable barrier of the fungal cell wall. In the image, helium ion microscopy illustrates growth of the novel isolate, *Anaeromyces robustus*, a depolymerizing crude reed canary grass. [Image courtesy Chuck Smallwood, Pacific Northwest National Laboratory.]

(CRISPRs; Cong et al. 2013; Qi et al. 2013) targeted to both prokaryotic and eukaryotic systems, researchers can now genetically modify an expanded set of organisms. However, additional barriers mitigate the effectiveness even of CRISPR systems, including but not limited to cell wall thickness, recombination, and transformation efficiency. Strategies need to be developed for systems with no available draft genome and with polyploid conditions that limit the effectiveness of genetic modification.

Characterizing Metagenomes for Proteomic and Metabolomic Analyses

Whole-community proteomics and metabolomics are promising approaches for integrated systems biology research addressing questions about microbial activities and interactions of uncultured, *in situ* taxa. However, many components of environmental and root microbiomes have no associated genomic sequence information, necessitating full metagenomic characterization prior to proteomic and metabolomic analyses. Quantitative mass spectrometry and shotgun proteomic approaches, the current standards for investigating these systems, are expensive, extremely limited in throughput, and inaccessible to many laboratories. Numerous research groups could benefit from increased access to shared informatics tools and data storage. Another key challenge requires technologies with sensitivity for low-abundance proteins and metabolites in natural samples such as soils and sediments, especially when considering metaproteomic characterization of an interacting system.

Quantitative proteomic approaches are excellent for assessing overall changes in the "physiology" of microbial communities and for helping identify biomarkers of the co-culture community that may indicate its stability (or instability), the presence or absence of particular organisms, and other important features of interest. Monitoring of such biomarkers is best performed using selected reaction monitoring (SRM) assays on a mass spectrometer that also can be multiplexed in multiple reaction monitoring (MRM). MRMs are fast, sensitive, and quantitative approaches to simultaneously monitor multiple proteins or peptides of interest. These technologies are rapidly replacing immunology-based tests (e.g., enzyme-linked immunosorbent assay, or ELISA) developed mostly for human health applications (Choi et al. 2013) but also applicable to environmental biological systems.

Monitoring Microbial Activity with Isotope Tracing

Currently, a large emphasis in microbial ecology is on sequencing DNA and RNA of natural microbial populations. Ideally, sequencing should be paired with direct monitoring of microbial activity to determine the relationship between genetic potential and actual process rates. Isotope-tracing

experiments are useful for monitoring microbial activity and have been used to study specific interactions in plant systems as well (Hernández et al. 2015; Moyes et al. 2016), including mycorrhizal fungi (Nuccio et al. 2013; Kaiser et al. 2015) and rhizosphere communities (Hernández et al. 2015; Shi et al. 2015; Pett-Ridge and Firestone 2017). To analyze more complex communities, researchers have applied combinations of (1) stable-isotope probing (SIP; Radajewski et al. 2000; Mayali et al. 2012; Hungate et al. 2015); (2) nanoscale secondary ion mass spectrometry (NanoSIMS) imaging (Pett-Ridge and Weber 2012); see Fig. 3.3, this page); and, for microbe identification, (3) fluorescence *in situ* hybridization (FISH; Behrens et al. 2008; Dekas and Orphan 2011; Musat et al. 2016) or Raman single-cell sorting (Eichorst et al. 2015). All these techniques have provided valuable links between identity and function in mixed microbial communities but need further adaptation for applications in soil and on root surfaces. New approaches also are needed for molecular-scale imaging of root exudates and microbial metabolites to capture the extremely rapid specific exchanges between plants and microbes. In these cases, the targeting of metabolically active microorganisms in detail may be best suited to molecular imaging techniques such as nanostructure-initiator mass spectrometry (NIMS; Northen et al. 2008) and SIP-based approaches including SIP-metagenomics, mRNA-SIP (Haichar et al. 2012), protein-SIP (Jehmlich et al. 2010), and metabolomics-SIP (Freund and Hegeman 2017).

Measuring Taxa-Specific Secondary Metabolites, Metabolic Fluxes, and Modeling

Natural ecosystems consist of thousands of different microbial species that share the same habitat. Although some organisms may compete for the same substrate, the waste product of one microbe could inhibit a neighbor, or two different microbes might simbiotically feed off each

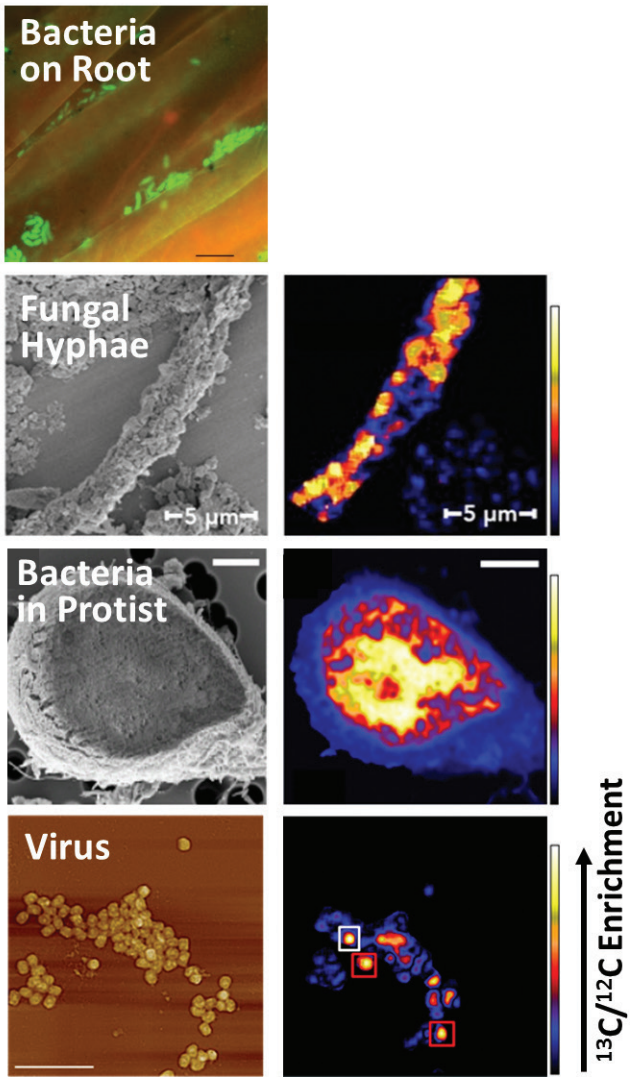


Fig. 3.3. Multimode Imaging of Organismal Interactions Relevant to the Rhizosphere. (From top) Fluorescence *in situ* hybridization (FISH)–tagged bacteria on an *Avena fatua* root. Scanning electron microscopy (SEM) and nanoscale secondary ion mass spectrometry (nanoSIMS) images of kaolinite-coated fungal hypha from a ^{13}C-plant incubation; a sectioned protist cell containing ^{13}C-cellulose–degrading bacteria; viral particles with ^{15}N-thymidine–enriched DNA. [Bacteria in Protist images from Carpenter, K. J., et al. 2013. "Correlated SEM, TEM, and NanoSIMS Imaging of Microbes from the Hindgut of a Lower Termite: Methods for *In Situ* Functional and Ecological Studies of Uncultivable Microbes," *Microscopy and Microanalysis* **19**(6), 1490–501, reproduced with permission. All other images courtesy Jennifer Pett-Ridge, Rhona Stuart, Rachel Neurath, Kevin Carpenter, Peter Weber, Alex Malkin, and Sean Gates.]

other's byproducts (Morris et al. 2013). All these processes happen simultaneously across countless microbes, controlling community structure and function as well as interactions with associated root networks. Of key interest are the secondary metabolites within microbial communities, which often are produced by cryptic environmental triggers that are difficult to reproduce in a laboratory (Cane et al. 1998). Current technologies are limited to genome-mining methods (e.g., antiSMASH) to posit putative (Blin et al. 2013) structures of these metabolites. New technologies are needed to advance genomic, transcriptomic, and metabolomic capabilities to (1) provide accurate predictions of novel metabolite structures, (2) quantify their production spatially in native environments, and (3) link metabolite generation to specific constituents within microbial communities.

Specific chemical signaling (both nonvolatile and volatile) between plants and microorganisms is understudied, with the exception of a few examples such as rhizobia and arbuscular-mycorrhizal fungi (AMF). Less is known about signaling between plants and nonendophytes or directly between microorganisms in soil. Fungi-bacteria interactions are of special interest, particularly in cases where nutrient foraging partnerships are expected (e.g., with AMF). Determining available metabolic networks within a suite of interacting organisms is critical to monitoring the community's key integration points. The development of radioactive tracer experiments has enabled verification of metabolic pathways predicted from genomic information (Stewart et al. 2010), but additional work using high-sensitivity approaches [e.g., accelerator mass spectrometry (AMS)] is needed to quantitatively parameterize multitaxa metabolic models. Novel frameworks for performing ^{13}C metabolic flux analysis (^{13}C-MFA) are now a reality for co-culture systems (Antoniewicz 2014; Gebreselassie and Antoniewicz 2014) but have yet to prove accurate for consortia systems with more than three members. Although promising, these frameworks are limited to model systems such as *Escherichia coli* and have not yet been harnessed for the nonmodel systems of interacting microbes more typically found in the rhizosphere. In addition, approaches using optimal isotopic tracers for ^{13}C-flux analysis often provide unclear results but still are needed to amplify differences for detection in cellular metabolism for the different community members. Typical outputs of these experiments are measured using gas chromatography mass spectrometry, or GC-MS and then are computationally deconvoluted using in-house software tools to produce species-specific metabolic fluxes. These computational tools also lack metabolic models for nonmodel microbes within the rhizosphere.

Imaging Plant-Microbe Interfaces

The rhizosphere is highly dynamic in time and space, and its spatial relationships difficult to study in a high-throughput manner because of the soil's opacity and large surface area. A suite of well-validated imaging methods could play a greater role in imaging plant-microbe interactions in both the rhizosphere and within plant tissues (Oburger and Schmidt 2016; see Fig. 3.4, p. 23). These methods include neutron radiography (in combination with deuterated water tracing; see Fig. 3.5, p. 24), ^{14}C phosphor-imaging, optode–diffusive gradient in thin films (DGT) sensor imaging, zymography, confocal scanning laser microscopy along with FISH combined with catalyzed reporter deposition (CARD-FISH), and labeling of target microorganisms with green fluorescent protein (GFP).

Information on the temporal behavior of localized chemical events is fundamental to understanding the highly dynamic interactions at the plant-microbe interface. As such, the ability to image chemical gradients and processes at the nano- to micrometer scale in real time will provide the highly informative data needed to unravel the complex networks of interface processes. Existing synchrotron FTIR (sFTIR) imaging methods have enabled insights into (1) micrometer-scale microbial community activities (Hazen et al. 2010;

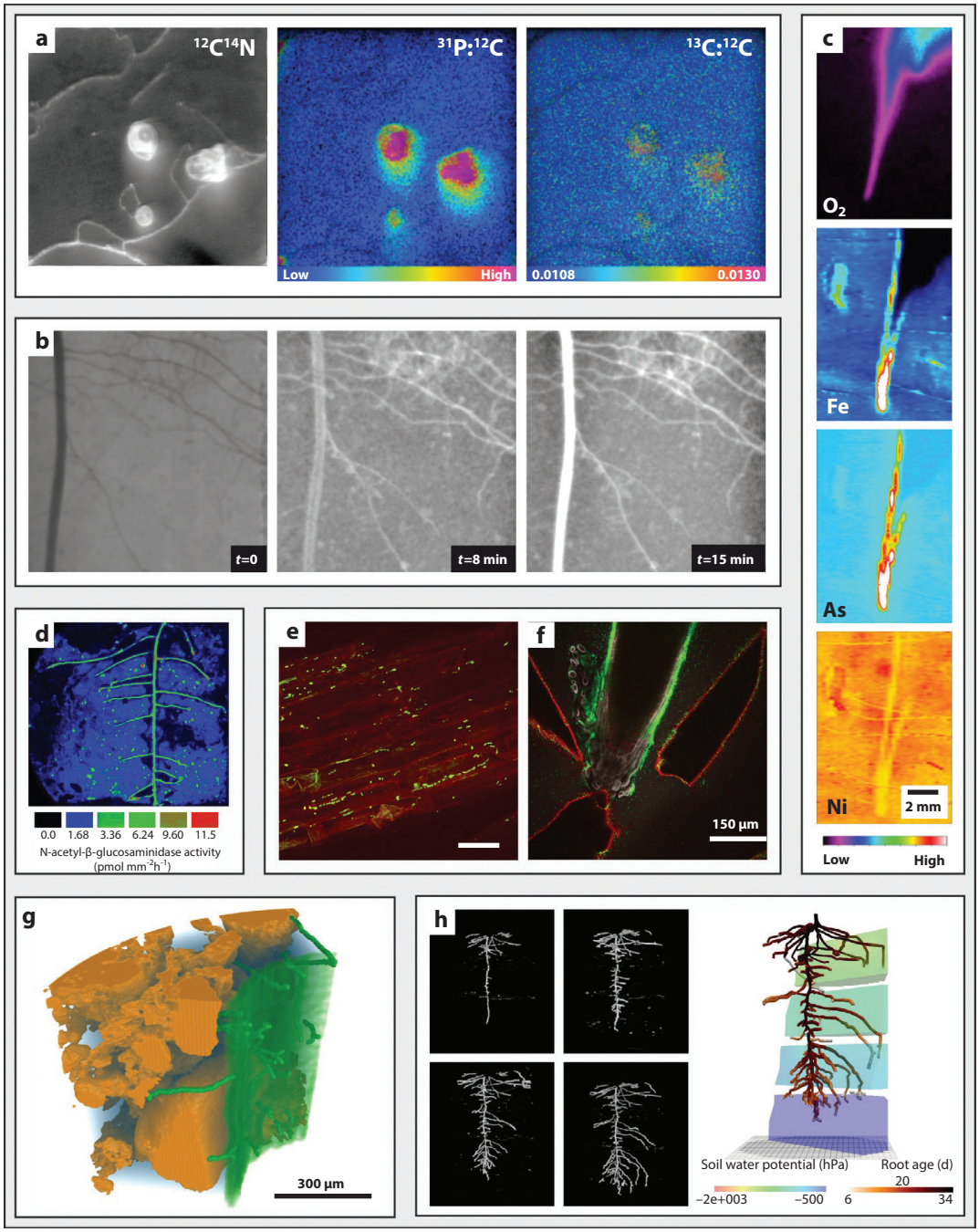


Fig. 3.4. Imaging of Processes at Root-Microbe-Soil Interfaces on Various Scales. (a) Nanoscale secondary ion mass spectrometry (nanoSIMS) of arbuscular-mycorrhizal fungal hyphae in a wheat root; **(b)** neutron radiographs showing flow of deuterated water into roots; **(c)** optode–diffusive gradients in thin film (DGT) sensor images of oxygen, iron, arsenic, and nickel in the rhizosphere of rice; **(d)** zymograph showing nitrogen-acetyl-glucosaminidase (chitinase) activity around lupin roots; **(e, f)** confocal scanning laser microscope images of microbes labeled with green fluorescent protein (GFP) and fluorescence *in situ* hybridization (FISH) in the rhizoplane, **(g)** synchrotron tomography illustrating rhizoplane root hairs (green), soil particles (brown) and water (blue), **(h)** X-ray microtomography reconstruction of bean root system architecture. [Reprinted from Oburger, E., and H. Schmidt. 2016. "New Methods to Unravel Rhizosphere Processes," *Trends in Plant Science* **21**(3), 243–55. © 2016, with permission from Elsevier.]

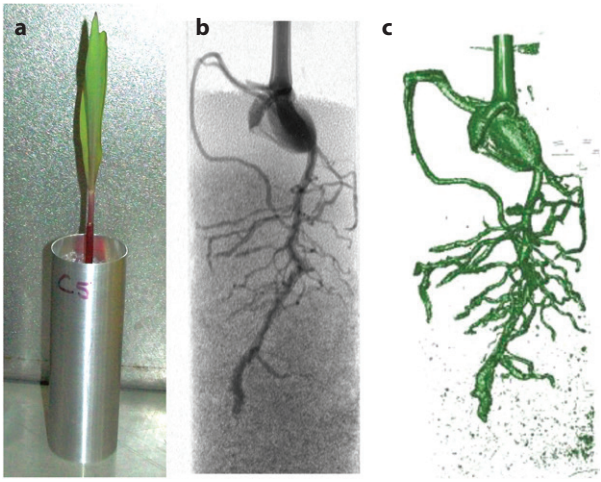


Fig. 3.5. Neutron Imaging Reveals Root Architecture and Internal Plant Water Dynamics *In Situ*. (a) Maize seedlings grown in sand imaged using radiography and tomography. **(b)** Pulses of water or deuterium oxide tracked through the root systems. Water flux within individual roots responded differentially to foliar illumination based on supply and demand of water in the roots. **(c)** Tomographic reconstruction of root architecture. [Reprinted from Warren, J. M., et al. 2013. "Neutron Imaging Reveals Internal Plant Water Dynamics," *Plant and Soil* **366**(1–2), 683–93, with permission of Springer.]

Baelum et al. 2012; Mason et al. 2012; Probst et al. 2013; Bouskill et al. 2016) and (2) plant root processes at micrometer spatial resolution (Cohen et al. 2015). Researchers need access to methods to characterize various plant-microbe interfaces, as well as additional sFTIR developments to image interface processes at nanometer length scales *in situ* and in real time.

Using X-ray and Neutron Imaging

Using X-rays to penetrate soil would seem like an obvious solution to the problem of soil opacity. However, such analysis has three primary impediments: (1) roots have the same attenuation level as water and organic matter, (2) the resolution of X-ray systems has precluded the detection of fine roots, and (3) X-ray system use has been restricted to the laboratory and with relatively small containers. Despite these challenges, the research community has made recent improvements on all fronts. For example, a group in the United Kingdom has developed robust software to identify roots in imaging data from complex soil environments (Mairhofer et al. 2013; Pound et al. 2013; Mairhofer et al. 2015). They also have built a high-powered X-ray system that can image plants grown in larger containers (Daly et al. 2015). Others have demonstrated the use of X-ray computed tomography for detecting rice roots, which are relatively thin, and have used this technology to investigate differences in root system architecture for the same genotypes in different soil substrates (Rogers et al. 2016).

Neutrons are uniquely sensitive to light elements such as hydrogen and its isotope deuterium; they are highly penetrating and, unlike X-rays, do not cause radiation damage to sensitive biological samples. Neutron radiography and tomography have made possible the study of (1) root system architecture and (2) water dynamics in the roots and the surrounding rhizosphere *in situ* using deuterium oxide as a contrast agent (Warren et al. 2013; Kroener et al. 2014; Totzke et al. 2017). The spatial resolution of neutron radiography is currently ~50 μm, and temporal resolution is 1 second. Improvements in detector technologies and optics to increase spatial resolution to less than 10 μm would enable imaging at the cellular level. In addition, the development of new contrast agents would allow new types of experiments to follow dynamic processes in plants and the associated rhizosphere.

Modeling Internal Root Anatomy to Optimize Function

Researchers have used mathematical models to simulate root growth under various environmental conditions (Walk et al. 2004; Dathe et al. 2016). These models make assumptions about the rates and direction of growth for different types of roots in response to nutrients, toxins, and water. One assumption is that growth rates for any particular root type in a homogeneous medium should be relatively constant. Challenging this assumption are recent observations of root growth rates over time that have detected

high levels of stochasticity in growth rates for roots from the same plant (Symonova et al. 2015). Modeling and experimentation indicate that internal root anatomy can play a role in optimizing root function. In particular, the formation of air spaces called "aerenchyma" appears to reduce carbon allocation to the root (Postma and Lynch 2011), while large cortical cell size may affect drought tolerance (Chimungu et al. 2014). Efforts have begun to integrate available phenotypic and molecular information into a platform called, "Plants *In Silico*" (Zhu et al. 2015). However, improved and validated modeling resources are needed to analyze this compiled information to determine which physiological and genetic traits determine differences in function.

Creating Nondestructive Methods to Detect Root Biomass, Architecture, and Function

In addition to interacting with soil microbes, roots anchor plants and are a plant's primary sites of nutrient and water acquisition. An improved understanding of root biology could help reduce fertilizer and irrigation use and improve carbon cycling models. As noted previously, soil opacity is the primary impediment to understanding root biology. Ideally, researchers want to be able to noninvasively determine the location of roots in soil and how they grow over time. Needed are new approaches for characterizing root system architecture in the field to target both shallow and deep roots. For shallow roots, "shovelomics" is the most common approach, involving cutting the root system close to the stem and characterizing the type and angle of visible roots (York and Lynch 2015). In addition, a recently developed image analysis platform can digitally characterize images of washed roots (Bucksch et al. 2014; Das et al. 2015), but this approach also has a major drawback—its destructive nature that precludes measurement of the same root system over time. Moreover, analysis of deeper roots relies primarily on various types of excavation, including total roots, soil cores, trenches, access shafts, caves, and mines (Maeght et al. 2013). Though an alternative approach employs cameras buried in tubes called mini-rhizotrons, their burial also disturbs the soil, creating preferred paths for root growth, among other problems (Maeght et al. 2013). Clearly, improved capabilities are needed for nondestructive characterization of shallow and deep roots *in situ*.

Development of noninvasive technologies such as ground-penetrating radar (GPR) also could enable *in situ* root detection. Primarily, GPR is useful for tree root detection (Barton and Montagu 2004; Jayawickreme et al. 2014), with a resolution limit around 5 cm. Lowering this resolution limit to detect crop roots and increasing the penetration depth to enable deeper root detection would make this technology more broadly useful. However, challenges would remain, including the need to move the GPR equipment within the narrow confines of normal breeding plots and the potential cost of its frequent use during the growing season.

The use of X-rays offers another possible means of measuring root biomass *in situ*. An example is the combination of X-ray, MS, and isotopic techniques that has enabled the tracking of organic molecules in rhizosphere environments in the laboratory (see Fig. 3.6, p. 26). Although potentially feasible for use in the field, the use of X-rays is limited there because of requirements for a very powerful generator and placement of a detector on the opposite side of the roots. One solution could be the creation of simplified, controlled environmental systems, called Ecotrons (Lawton et al. 1993), for developing new imaging approaches and conducting *in situ* experiments.

A final approach being investigated for *in situ* root detection involves the use of sensors that report contact with or close proximity to roots. New, miniaturized sensors attached to a cage-like device buried in the soil could provide a low-cost and noninvasive alternative to the direct imaging of plants.

Another useful capability, in addition to determining root system architecture, would be the ability to discover the internal root anatomy. For example, research indicates that aerenchyma formation is

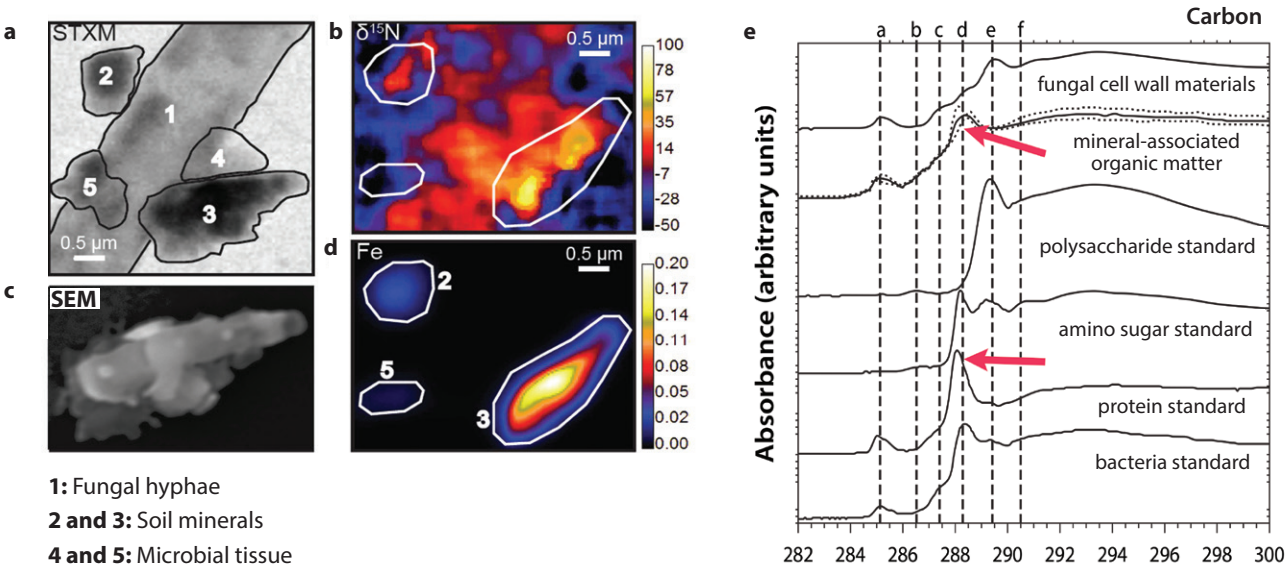


Fig. 3.6. Characterization of Organic Molecules from Soil Using Multimodal Imaging Techniques. Combined imaging using high-resolution technologies with both isotopic and molecular sensitivity enables researchers to track several activities of organic molecules: (1) transfer among system components, (2) biological modification, (3) assimilation into cells, or (4) association with inorganic particles (Keiluweit et al. 2012). In these figures, sequential imaging with scanning transmission X-ray microscopy (STXM), scanning electron microscopy (SEM), and nanoscale secondary ion mass spectrometry (NanoSIMS) illustrate how isotopically enriched (^{13}C) fungal cell material is transformed during soil decomposition and subsequently becomes preferentially sorbed to iron oxide mineral surfaces. **(a, e)** STXM and near-edge X-ray absorption fine structure carbon image and spectra. **(c)** SEM of iron oxide mineral. **(b, d)** NanoSIMS images showing nitrogen enrichment and iron distribution. [Images courtesy Jennifer Pett-Ridge. Reprinted from Kuweit, M., et al. 2012. "Nano-Scale Investigation of the Association of Microbial Nitrogen Residues with Iron (Hydr)oxides in a Forest Soil O-horizon," *Geochimica et Cosmochimica Acta* **95**, 213–26. © 2016, with permission from Elsevier.]

important under certain environmental conditions. To date, this type of analysis has been limited to destructive methods, but techniques enabling *in situ* anatomical analysis would be very valuable. Monitoring root functional attributes, such as nutrient and water dynamics, is another area of interest. Although researchers have examined these attributes in the laboratory with positron emission tomography imaging (PET) and NMR, in-field approaches capturing these physiological functions apparently do not yet exist. Finally, other needed capabilities involve nondestructive tracking of microbe-root interactions in soil. To date, researchers document these interactions through excavation and root washing. A transforming development in rhizosphere research would be a sensor-based system to detect microbes associated with roots in soil and perhaps even metabolites associated with specific microbial taxa.

Developing Transparent Model Soil System

Development of a transparent medium having the properties of soil potentially could make laboratory analysis of root growth more relevant to the field. Treating soil with the polymer Nafion renders it transparent with a refractive index close to that of water, and this treated soil has some of the

physical properties of native soils. Its primary use to date has been to observe root-microbe interactions (Downie et al. 2012).

Modeling Interactions Among Fauna and Microorganisms in the Rhizosphere

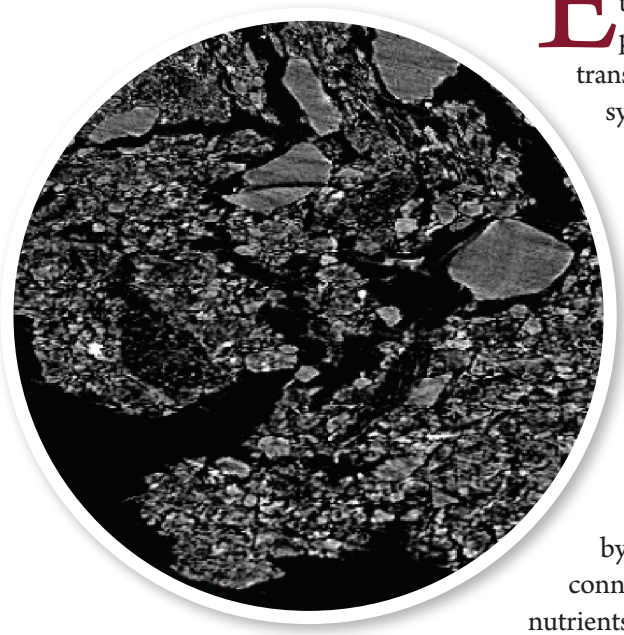
Carbon dynamics within soil ecosystems is dominated by a web of biotic transformations mediated by photosynthetic and heterotrophic microbes and metazoans. Recognition of these dynamics has enabled development of consumer-resource models of food webs to (1) predict the quantitative effects of species removal on the remaining taxa within intertidal food webs (Berlow et al. 2009) and (2) forecast seasonal changes in microbial and metazoan carbon dynamics in a complex aquatic ecosystem (Boit et al. 2012). The cryptic nature of soil ecosystems has prevented the understanding of soil food webs from developing as rapidly as that of more easily observed aquatic systems (Brose and Scheu 2014). The discovery of more powerful "omics" and isotopic techniques is rapidly increasing the ability to more accurately model soil ecosystem carbon dynamics (Jakobsen and Rosendahl 1990; Digel et al. 2014). For example, research has recently identified unique structural aspects of complex soil food webs (Riede et al. 2010) and explored the effects of temperature and moisture on the dynamics of simple soil food chains—theoretically (Binzer et al. 2012) and experimentally (Lang et al. 2012; Lang, B., et al. 2014). The next challenges are to (1) integrate simple components of soil food webs into more realistic and predictive networks by resolving resource flow and incorporating key additional processes and (2) add the ecological and biogeochemical roles of understudied groups such as viruses and microfauna.

3.4 Optimizing Plant Health, Soil Fertility, and Carbon Storage

The condition of the root system of a plant directly relates to its aboveground health and productivity, and thus the amount of biofeedstock produced. Microbial communities drive fundamental processes in the global carbon cycle, regulating both atmospheric CO_2 levels and soil carbon storage. These communities are particularly impactful in the rhizosphere, where they may assist the plant host with nutrient or water acquisition, provide pathogen defense, or mediate the transformation of root tissues into stabilized SOM. To better understand plant-microbe interactions, researchers need the ability to make quantitative, *in situ*, 3D measurements of dynamic molecular phenomena and to capture the interactions and exchanges between plants and their microbial partners at scales ranging from single cells to field plots. A related need is the ability to monitor root growth noninvasively over a growing season in a field setting, capturing root impacts on soil carbon and water stocks. Well-designed mutualistic plant-microbe interactions could improve plant nutrient and water acquisition and stress tolerance while simultaneously promoting carbon stabilization, soil fertility, and a diverse soil biota.

Chapter 4

Biogeochemical Cycling



Environmental system function is intimately tied to biogeochemical cycling of the major elements, particularly their reduction-oxidation (redox) transformations. The inherent complexity of natural systems creates a network of pathways for electron transfer that closely couples biogeochemical cycles. Microorganisms inhabiting these environments, such as rhizospheres or leaf surfaces, are known to play important roles in global biogeochemical processes by catalyzing the chemical transformations and movement of carbon, nutrients, and contaminants (Anderson et al. 2003; Karl et al. 2012). The populations within a community respond differently to environmental parameters that include biotic and abiotic variables within the system's boundaries. The cumulative effect caused by these multiple populations, directly or indirectly connected, facilitates the net movement of energy and nutrients through the system. However, the understanding of all these interactions from molecular to microscale to macroscale, even for simplified communities, is limited. To understand, model, and predict microbial behavior and function, conditions for microbial communities must be delineated for spatial and temporal relationships among phylogenetic diversity, functional potential (e.g., carbon fixation, phosphorous and nitrogen cycling, and metal binding), and habitat.

4.1 Current Science and Technologies

Some research efforts funded by the U.S. Department of Energy's Office of Biological and Environmental Research (BER) focus on microbial populations—how they communicate; evolve; share resources; interact with abiotic

constituents and other organisms in subsurface, rhizosphere, and aquatic environments; are affected by environmental changes; and, ultimately, define the Earth's landscape. Research techniques are needed to enable and transform mechanistic understanding of structure-function relationships, from biological cells at the nanoscale to complex microbial communities at the macroscale, in three dimensions and across time, to improve capabilities for modeling and predicting microbial behavior and function.

BER-relevant ecosystems and environments include the subsurface and rhizosphere, surface waters, groundwater, rivers, wetlands, peatland and bogs, permafrosts, grasslands, forests, agricultural soils, and engineered systems (e.g., bioreactors). Within the ecosystems of these terrestrial and aquatic environments, microorganisms, plants, and soil and sediment constituents biogeochemically interact with and affect each other, exerting great control on the partitioning of elements and chemical compounds among the air, water, and land. In turn, such biogeochemical interactions have a major impact on critical environmental phenomena such as nutrient availability to living organisms, contaminant fate and transport, and water quality. Interactions from the molecular scale (Angstrom to nanometers) to the mesoscale (millimeters to a meter) also can drive emergent biogeochemical processes at larger scales. Moreover, with unknown metabolic interactions distributed across heterogeneous matrices that have complex diffusion characteristics, there is difficulty determining the impact of biotic and abiotic structures across temporal and spatial scales. For example, metabolic crosstalk in specific and nonspecific consortia, redox zonation resulting from combined biotic and abiotic interactions, hydrology and water quality linked to microbial activity, and carbon reduction and oxidation across oxic and anoxic boundaries are all affected by these complex biogeochemical interactions.

The geocycling of elements and other more complex molecules plays an important role in shaping the environment. Of the elements, iron and sulfur play particularly important roles because of their abundance and chemistry. Iron is highly abundant and often a dominant redox-active, biogeochemically cycled element in Earth's outermost shell, the lithosphere. Iron-bearing clay minerals and iron oxides are common constituents of soils and sediments. Iron's biogeochemistry in most aquatic and terrestrial environments is driven largely by microbial activity, particularly in iron-rich soils and sediments, where iron redox cycling between Fe(II) and Fe(III) by microorganisms is a significant component of carbon and nutrient cycling and energy flux (Canfield et al. 1993; Nealson and Saffarini 1994; Roden and Wetzel 1996; Lovley 2000; Thamdrup 2000). Although sulfur at times is less abundant than iron in typical aquatic and terrestrial environments, it too is one of the main biogeochemically cycled elements within the lithosphere that is redox active. Sulfur's biogeochemical transformations are more complex because of the variety of sulfur oxidation states (VI, V, IV, III, II, I, 0, –I, and –II). Sulfur species with intermediate valence states are key intermediates in sulfur redox cycling, which, for iron as well, is driven largely by microbial activity in most aquatic and terrestrial environments. The presence of sulfide (as either dissolved sulfide or ferrous sulfide minerals) can profoundly affect the fate, transport, and bioavailability of a variety of nutrients, as well as control the evolving nature of microbial community structure and function within soils and sediments.

4.2 Major Challenges in Learning How Microbes Control Biogeochemistry Cycling and Mobility

Understanding Coupled Biogeochemical Cycles in Redox-Dynamic Environments

The sulfide generated by microbial sulfate reduction and its subsequent reaction with reduced iron, produced by iron-reducing organisms, can lead to the formation of insoluble iron sulfides. Such pathways are important links coupling iron and sulfur biogeochemistry. Similarly, the redox processes coupling the biogeochemical cycles of iron and sulfur

Biogeochemical Cycling Needs

Predicting Regional- and Global-Scale Processes

New insights for understanding and, ultimately, predicting the dynamic interplay among biotic and abiotic environmental components, from molecular to mesoscale and at multiscale spatial and temporal resolutions.

- Probes with a variety of spatial and temporal resolutions to study the relevant biogeochemical system.
- New and improved technologies to image components at resolutions of nanometers to millimeters and subseconds.
- Contrast agents and tags to probe molecular to mesoscale processes in biogeochemical systems.
- Advancement and standardization in imaging technologies to routinely image the soil pore network together with the life that inhabits it.
- Advanced approaches in the coupled development of X-ray contrast agents and X-ray microtomographic imaging to distinguish cells from pores, organic matter, and minerals within opaque environmental media.

with carbon, nitrogen, oxygen, and manganese in redox-dynamic environments play a critical role in major ecosystem processes, which include carbon assimilation, mineralization of organic matter and its accompanying release of carbon dioxide (CO_2), methane (CH_4) formation and oxidation, uptake and release of nutrients (e.g., nitrogen and phosphorus), and contaminant availability and transformation. Thus, improved understanding of the coupled iron and sulfur redox dynamics and their subsequent effects on nutrient cycling is key to advancing understanding of how biogeochemical processes help to control the cycling and mobility of materials in aquatic and terrestrial systems.

A variety of competing biogeochemical processes determine the flux of terrestrial greenhouse gases (e.g., CO_2, CH_4, and nitrogen) to the atmosphere. These processes are mediated by both microbes and plants. For example, the interaction between methanogens (methane producers), methanotrophs (methane consumers), and other anaerobic microorganisms (e.g., iron reducers, sulfate reducers, and denitrifiers) can create a complex, dynamic network of interacting redox processes that directly affects global biogeochemical cycling and CO_2 and CH_4 flux. Key genes in the microbial biochemical pathways that mediate CH_4 flux (e.g., *mcr, pmoA, and mmo*; see box, this page) display substantial diversity that varies significantly among sites, even within a given region, suggesting a much greater degree of metabolic diversity than is known from studies of isolated strains. The illustration in Fig. 4.1, p. 32, illustrates some of the potentially coupled biogeochemical cycles that can affect emissions of greenhouse gases such as CO_2 and CH_4 from a redox-active environment such as wetland sediments. The coupling of these dynamic biogeochemical processes is highly dependent on the spatial arrangement of microbial metabolisms along linear redox gradients. The gradients run perpendicular to the sediment-water interfaces and radial redox gradients emanating from wetland plant arenchyma. For use in simulation and prediction of such processes, these biogeochemical interactions need to be illuminated and quantified across multiple spatial scales—from nanometer- to micrometer-scale cellular metabolism

Key Methane Biochemical Pathway Genes

mcr – methyl coenzyme m reductase gene

pmoA – particulate methane monooxygenase gene

mmo – methane monooxygenase gene

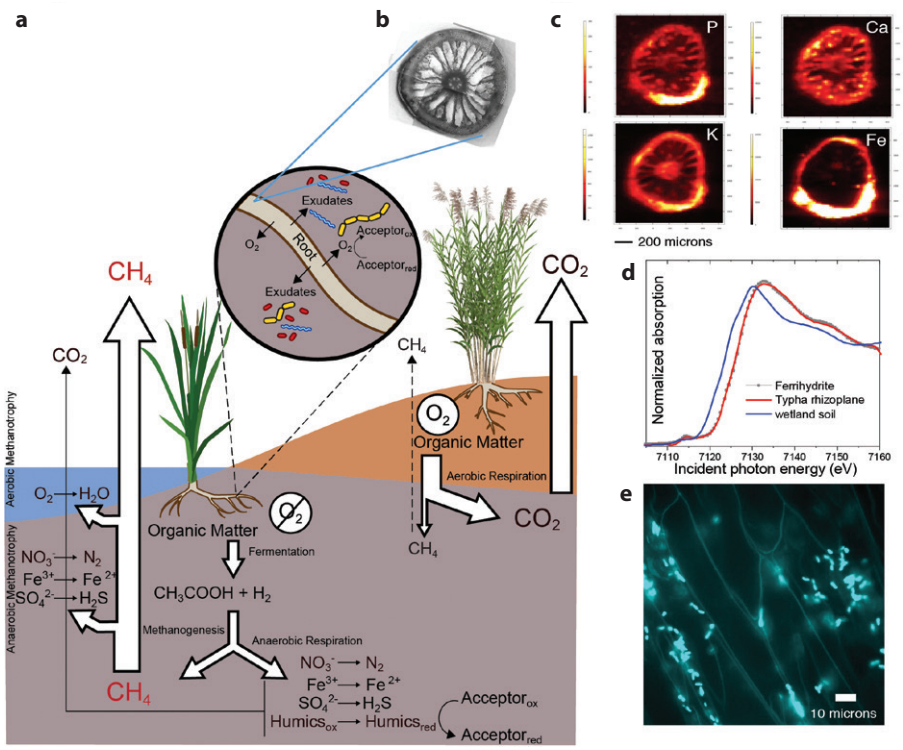


Fig. 4.1. Coupled Biogeochemical Cycles. (a) Routes of carbon turnover in anoxic and oxic interface zones. Arrows with gas flux to the atmosphere represent potential emission routes. **(b)** Optical microscopy cross-section image of wetland plant aerenchyma. **(c)** Corresponding X-ray fluorescence image of nutrient elemental distributions within aerenchyma cross section in **b**. **(d)** Spatially resolved (25-micrometer resolution) iron X-ray absorption spectroscopy measurements of iron valence state in rhizoplane of wetland plant aerenchyma [red Fe(III)] and surrounding sediment [blue Fe(II)]. **(e)** Confocal microscopy image of cytoplasmic fluorescence-labeled rhizobacteria (*Pseudomonas fluorescens* SBW25) colonizing plant root cells. [Image courtesy Ken Kemner, Argonne National Laboratory.]

and mineralogy, to millimeter- and meter-scale community biogeochemical activities, to ecosystem-scale processes. This research requires experimentally quantified molecular to mesoscale (nanometer- to meter-length) measurements of soil carbon and nutrient flow, related biogeochemical processes, and multiomic characterization of microbial dynamics, along with integration of these data into predictive microbial metabolic and systems-ecology models.

A research grand challenge would be to image the distribution and measure the dynamics of elements, microbes, metabolisms, and metabolites within a small (e.g. 1 mm^3) volume of soil and develop a mechanistic understanding and model of its biogeochemical function and control of environmental phenomena within that soil aggregate such as redox gradients and carbon mineralization. Projected 100- to 1,000-fold increases in X-ray brilliance over the next 5 years at diffraction-limited X-ray sources, such as the Advanced Photon Source, will make achieving these objectives possible for a 1 mm^3 volume. However, imaging a volume beyond that, 1 cm^3 with 10 nm^3 voxel resolution, for example, would remain an enormous challenge.

Imaging a Biogeochemical System with Multiple Technologies

Besides being opaque, environmental media such as sediments and soils are dynamic and spatially complex. Technology is needed to provide dynamic imaging of the biogeochemical processes and elemental transformations occurring within these complex environmental media. Moreover, because these processes often are catalyzed by living organisms to obtain energy or nutrients, technology also is needed to enable imaging of inter- and intracellular biological processes. A combined and holistic use of various dynamic imaging and characterization probes, coupled with omics and modeling

approaches, is necessary to bridge across spatial and temporal scales in biogeochemical systems to better understand their role in key environmental processes. The probes used comprise photons with wavelengths across the electromagnetic spectrum, including visible, infrared, ultraviolet, and soft and hard X-rays; charged particles such as electrons and ions; and neutrons. Examples of how some of these characterization probes can be used to image a biogeochemical system such as wetland sediments and plant roots are discussed in Chapter 5 (see Fig. 5.1, p. 38). In addition to approaches for imaging in the laboratory, where many microscopies may be more routinely available, also needed is the ability for continuous imaging of unmodified, untreated natural soils and sediments. Described below in this section are some of the improvements and developments needed to advance spatial and temporal imaging approaches to gain a better mechanistic understanding of biogeochemical processes from the molecular to mesoscale.

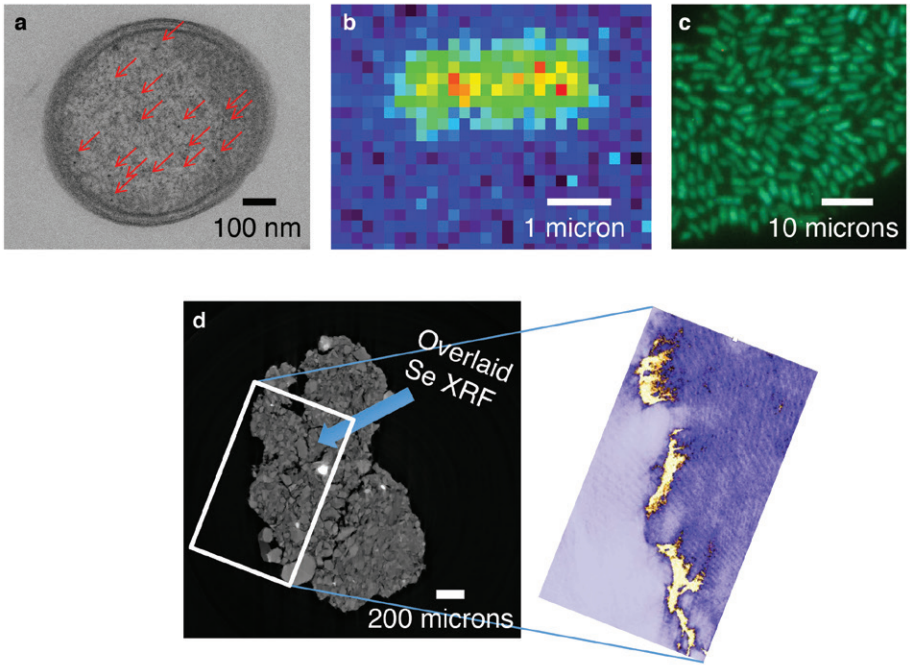


Fig. 4.2. New Contrast Agents for Biogeochemical Analyses. The development of new contrast agents for use in imaging probes should enable the study of biogeochemical systems at spatial scales appropriate to the systems interrogated and questions asked. Cadmium-selenium (CdSe) and zinc-sulfur (ZnS; core/shell) quantum dots can be useful contrast agents for electron–, X-ray–, and optics–based imaging approaches. **(a)** As demonstrated with transmission electron microscopy imaging, CdSe and ZnS quantum dots (identified by red arrows) conjugated to an amino acid enable uptake of quantum dots by *Pseudomonas protogens*. **(b)** The X-ray fluorescence from selenium atoms within the quantum dots can be used to image a similar bacterium with a submicrometer-sized hard X-ray microbeam. **(c)** The visible light resulting from the quantum confinement effects emitted by quantum dots enable bacterial imaging with optical wavelength microscopy. **(d)** A combination of X-ray transmission microtomography (left) and X-ray fluorescence microtomography (right) can be used to create tomographic reconstructions of pore structure (left) and the spatial distribution of quantum dot–containing bacteria within the pores (right) of a visibly opaque, millimeter-sized soil aggregate. [Image courtesy Ken Kemner, Argonne National Laboratory.]

Development of Contrast Agents and Tags. Whole-cell labels, protein tags, and gene expression or biosensor reporters have led to numerous insights in protein function, cell architecture, and cellular processes. Tags and contrast agents for optical, X-ray, charged-particle, and neutron imaging are needed to advance the use of these probes for optimal understanding of molecular to mesoscale processes in biogeochemical systems (see Fig. 4.2, this page).

Visible Light Microscopy. Intracellular transport of nutrients and metabolites in and across the cytoplasm—as well as intercellular transport across cell walls, microbial membranes, and among

cells—occurs in all three dimensions and can be characterized as a complicated multiscale combination of stochastic processes along with active transport. Accurate measurements of intracellular and intercellular transport and structure that are free from temporal and spatial artifacts require data with high three-dimensional (3D) spatial resolution and simultaneous volumetric acquisition. Estimating an object's 3D structure is one of the most fundamental, yet challenging, tasks in optical microscopy, and making this task difficult is the lack of imaging sensors that can directly detect 3D information [e.g., instead, 2D charge-coupled device (CCD) array detectors are used].

Infrared Spectral Microscopy. Key to understanding biogeochemical cycling are *in situ* and real-time measurements of how microorganisms interact with terrestrial and aquatic environments through their wide range of metabolic capabilities. These measurements are important because the coupling of abiotic geochemical cycling of elements (e.g., iron, sulfur, carbon, nitrogen, oxygen, and manganese) with biotic microbial activities occurs primarily at the microbe-substrate interface, and their interfacial properties are quite variable spatiotemporally. At a spatial scale of several to hundreds of micrometers, the coupling could be dominated by clusters of microorganisms that differ in their metabolic activity and by spatial variations in reactive molecules of metal oxides and organic molecules. Several synchrotron Fourier transform infrared (sFTIR) spectromicroscopy tools are available for label-free, real-time molecular measurements. sFTIR can provide detailed molecular information for the functional groups of biomolecules, metabolites, and geological organic and inorganic materials, as well as continuous measurements of *in situ* biogeochemical processes. Scientists also apply sFTIR to make real-time sequential measurements of microbial or biogeochemical processes at resolutions of micrometers (i.e., spatial) and minutes to hours (i.e., temporal) (Holman et al. 1999; Holman et al. 2002b; Holman et al. 2009a; Holman et al. 2009b). Further development of sFTIR would improve chemical sensitivity and spatial and temporal resolution, providing new insights into biogeochemistry at the microbe-environment interface.

Charged-Particle Microscopy. Imaging technologies using electrons and charged particles can provide an opportunity for increased spatial resolution and, hence, visualization of subcellular components within microorganisms and plant cells. In addition to 2D scanning and transmission electron microscopies, electron-based imaging has been revolutionized by electron tomography. Although these approaches often require chemical fixation or quick freezing of biogeochemical samples, thus hindering direct study of biogeochemical process kinetics, they do enable 3D imaging of intercellular and intracellular structures with nanometer-scale resolution. In addition, judicious sample preparation approaches, ensuring that a sample and its reactions are suspended at the appropriate time, can enable reconstruction of dynamic biogeochemical processes after microscopy data collection. For holistic imaging and better understanding of complex biological systems, advancement and standardization are needed in electron-based imaging technologies, including their coupling with other imaging approaches, and those based on optics, X-rays, and charged particles.

X-Ray Microscopy. Further advancements are needed for imaging techniques that enable 3D imaging of cells as well as metabolic and biogeochemical processes within opaque environmental media. Soil pore structure affects water and oxygen diffusion, introduction of microbes and their reactants, and root structure (Kemner et al. 1998; Bailey et al. 2013). While laboratory-based X-ray tomography can image the pore structure of soil aggregates, highly coherent X-ray sources provided by synchrotrons enable imaging of root cell structure with opaque soils (see Fig. 4.3, p. 35). In addition, microbial metabolic activities are the primary drivers of biogeochemical processes controlling the terrestrial carbon cycle, plant nutrient availability, contaminant remediation, and other ecosystem services. However, understanding how microbes and microbial metabolism are distributed

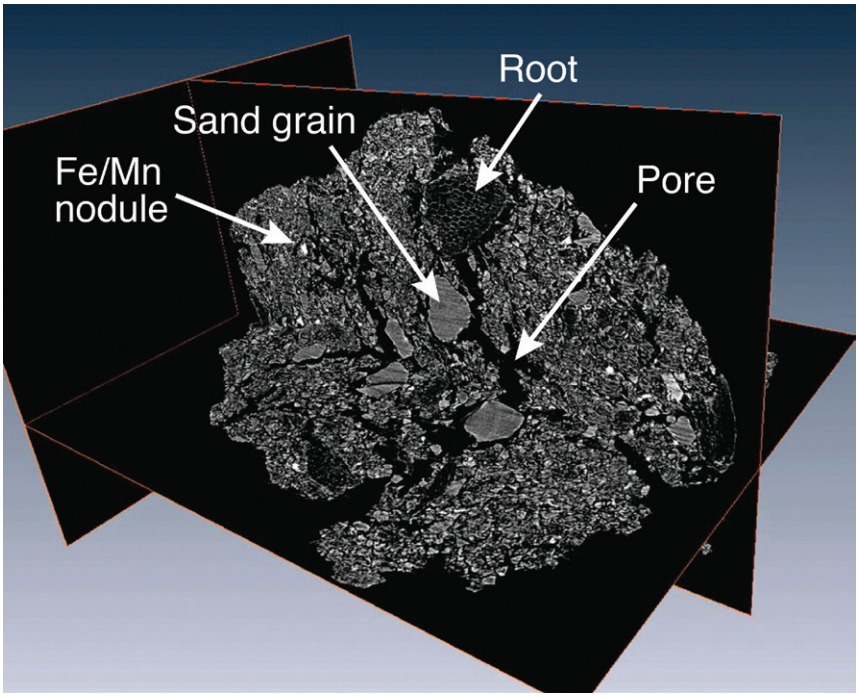


Fig. 4.3. Soil Pore Structure and Root Cell Structure Within a Soil Aggregate. Reconstructed transverse and vertical planes from a transmission X-ray microtomography measurement of a soil aggregate ~1 millimeter in size. Highly coherent X-rays provided by the Advanced Photon Source at Argonne National Laboratory enable imaging of pores, sand grains, iron-manganese nodules, and cell walls from a root within the soil aggregate. [Image courtesy Ken Kemner, Argonne National Laboratory.]

throughout biogeochemical systems such as soil aggregates is limited because no standard techniques are available to routinely image the soil pore network together with the life that inhabits it. Light- or mass spectrometry–based imaging approaches [e.g., fluorescence *in situ* hybridization (FISH) and nanoscale secondary ion mass spectrometry (nanoSIMS)] can detect the location of organisms in relation to their biogeochemical environment, while electron-based imaging approaches can determine with high spatial resolution the location of heavy metal–tagged cells. However, these techniques are unable to go beyond the surface of an opaque sample without sample sectioning. X-ray tomographic approaches (Flannery et al. 1987) are unmatched for producing nondestructive 3D images of rocks without any prior preparation. X-ray microtomography can provide highly detailed 3D renderings of soil mineral and pore structure (Kemner et al. 1998; Young and Crawford 2004; Nunan et al. 2006; Bailey et al 2013), but these approaches cannot distinguish biological cells from other electron-light materials such as air or water. Advances in the coupled development of X-ray contrast agents and X-ray microtomographic imaging approaches are needed to distinguish cells from pores, organic matter, and minerals within opaque environmental media. Planned upgrades to synchrotron and free-electron laser X-ray sources will provide many orders of magnitude increases in X-ray brilliance. These facilities are expected to offer new modes of X-ray–based imaging and increased measurement throughput with existing techniques.

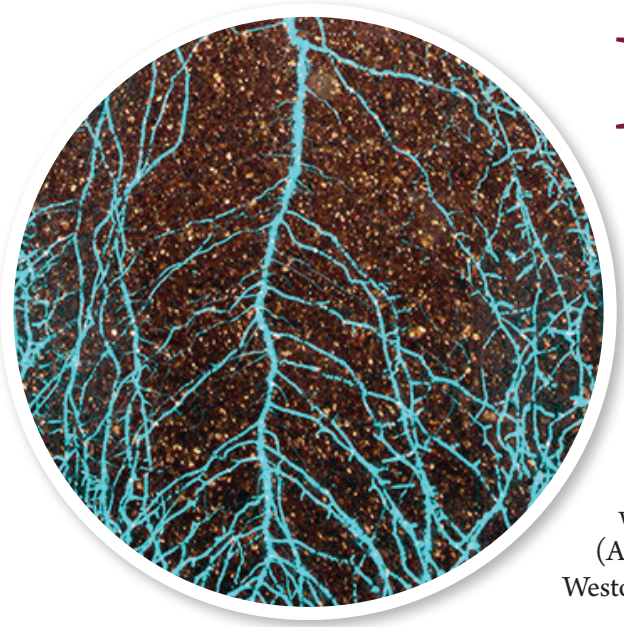
4.3 Improving Model Predictions of Biological and Environmental Systems

Spatial and temporal imaging of biogeochemical system dynamics from the molecular to mesoscale will enable the development of more sophisticated and precise models to predict (1) ecosystem, regional, and global greenhouse gas production; (2) nutrient and contaminant movement and availability; (3) sustainable biomass production; and (4) downgradient water quality. These improved models also will offer better predictions of how microbial

and environmental drivers will respond to future environmental change (i.e., changing atmospheric CO_2 concentrations and aquatic pH, changing temperature, water inundation and dry spells related to episodic climatic events, and different types of minerals and carbon in soils and sediments).

Chapter 5

Metabolic Pathways in Plants, Microbes, and Fungi



Plant and microbial metabolic pathways have long been harnessed for making bioproducts. Scientists have traditionally relied on breeding methods in plants or random mutation methods in microbes to select for the genes that improve these processes. This has led to the concept of biological production of fuels and other compounds, which is central to the mission of the Biological Systems Science Division of the Office of Biological and Environmental Research. It is well established that plant and microbial communities exchange metabolites in an "economy" that ultimately determines the rate of nutrient and water extraction from soil and cycling of soil carbon (Anderson et al. 2003; Johnson and Nielsen 2012; Weston et al. 2012; Song et al. 2015).

A deeper understanding of the mechanisms that these communal organisms use to interact in the environment, as well as the metabolic pathways and specific molecules involved in this exchange, will enable the modification of such pathways to improve nutrient-use efficiency and soil-carbon performance (see Fig. 5.1, p. 38). Ultimately, this knowledge is essential for arriving at optimal plants, microbes and pathways for the biological production of useful molecules that would otherwise be derived from fossil fuels.

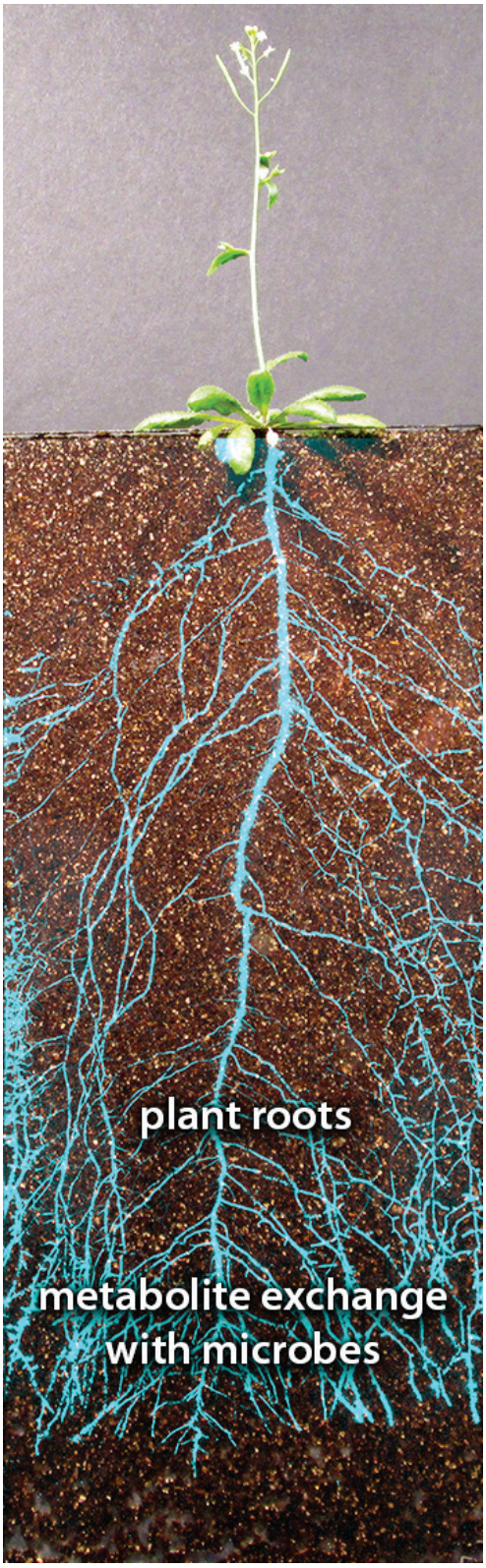


Fig. 5.1. Defining Metabolic Interactions at the Plant-Soil-Microbe-Interface. Growth and Luminescence Observatory for Roots (GLO-Roots) image of soil-grown root system. [Image courtesy Rubén Rellán-Álvarez and José R. Dinneny, Carnegie Institute for Science.]

5.1 Current Science and Technologies

Predicting Metabolic Networks from Genomes and Proteomes

Metabolic pathways comprise enzymes encoded by genes, substrates endogenously produced or exogenously supplied, and products generated through a catalyzed reaction. The metabolome is the full complement of metabolites present in an organism along with a complement of proteins (the proteome) that are available to perform chemical reactions. Understanding how the proteome influences the metabolome requires knowledge of (1) the genes present in the genome, where the proteins are encoded, and (2) how the sequences of proteins relate to their enzymatic function. Model organism studies have determined the enzymatic function of a large number of proteins (Bassel et al. 2012). Using sequence homology, researchers can predict which genes in other species might encode enzymes with similar substrate specificities and products, but the extent to which sequence-based methods result in false-positive or -negative predictions is not clear. Large numbers of microbial and eukaryotic genomes are being sequenced, but little experimental data exist for the protein families they express. Thus, there is a pressing need for functional characterization of specific DOE mission–related enzymes.

All the enzymes that perform known metabolic reactions in an organism can be predicted through sequence homology to characterized enzymes (Mueller et al. 2003; Chae et al. 2014). Many of the enzymes involved in central metabolism have been well characterized and provide a useful resource for comparing these core processes across species. Specialized metabolism, on the other hand, is necessary for an organism to generate the chemicals

Metabolic Pathways in Plants, Microbes, and Fungi Needs

Using Advanced Engineering to Study Metabolomes and Design Organisms

Detection methods and analysis for identifying metabolites that vary in biologically interesting and DOE mission–relevant ways.

- Methods to determine reference spectra for known chemicals, and algorithms that more accurately deconvolve the spectra.
- Methods compatible with in-field measurements to detect metabolites.
- Efforts to establish fabricated ecosystems.

Developing Resources and Higher-Resolution Tools for Localization and Measurement

Capabilities that detect changes in the dynamic systems of plants, microbes, and fungi for finding important functional differences that affect physiology and growth, such as a metabolite flux highly tuned to suit particular environmental conditions.

- Tools to measure the dynamic nature of metabolite fluxes within and among organisms, combined with new higher-resolution approaches to determine spatiotemporal localization and the mechanisms responsible for metabolite synthesis, transport, and degradation.
- Tools to detect internal and external temporal changes in cellular environments, along with cell type–specific differences in metabolites.

that perform often species-specific functions, conferring many of the unique adaptive traits needed for the survival within its specific environmental niche (O'Connor 2015). The plethora of chemicals naturally produced by plants and fungi affect biotic interactions and, for example, form the basis of pharmacological activities used in traditional and modern medicine (Anarat-Cappillino and Sattely 2014). Recent work has shown that the number of genes predicted to encode specialized metabolic enzymes appears to have expanded to a greater extent during plant evolution, compared to genes involved in other metabolic pathways. This difference suggests that the specialized genes were important for the success of plants in land colonization (Chae et al. 2014). The co-localization of genes involved in the same metabolic pathway along chromosomes appears to be a unique property of genes linked to specialized metabolism. This finding suggests that co-regulation may be important for ensuring that any specialized metabolic compounds that are potentially reactive or toxic remain relegated to specific tissues or environments during synthesis.

Detecting Metabolites and Measuring Metabolic Flux

Detecting the activity of an enzyme or flux through a metabolic pathway requires the ability to quantitatively detect the substrate, product, or both. Metabolite detection can be "direct" or "indirect." Direct detection methods usually rely on knowledge of the compound's exact mass, which can be determined precisely by mass spectrometry (MS). Modern instruments can discriminate thousands of different chemicals based on separation through chromatography, ionization, and mass determination (Kusano et al. 2015). Indirect detection of known metabolites can be facilitated by using chemicals that react specifically with the compound of interest. Protein-based sensors, for example, may involve an antibody specific to the metabolite of interest, and immunohistochemistry can be used for visualization. More recently, a promising development concerns proteins whose fluorescent properties change when binding to specific chemicals, potentially expanding the spatial and temporal scales at which metabolic pathways can be interrogated (see

next section, "Quantifying Metabolites with Direct and Indirect *In Situ* Methods," this page; Jones et al. 2013). Importantly, each of these direct and indirect methods requires precise knowledge of the chemical composition of the particular compound under study. Untargeted detection of the metabolome also can lead to a chemical fingerprint–based analysis in which quantitative differences in the abundance of chemicals with specific masses are used to compare an organism's overall metabolic state (da Silva et al. 2015). Such analyses can help to identify metabolites that vary in biologically interesting ways, thus justifying time- and resource-consuming approaches that enable exact determination of the chemical structure.

Quantifying Metabolites with Direct and Indirect In Situ *Methods*

Tissue-specific methods for characterizing plant transcriptomes and proteomes have revealed that most biological pathways are highly regulated at the spatiotemporal level (Mueller et al. 2003; Fukushima et al. 2009). The relative flux through a metabolic pathway can be estimated by integrating these omics datasets with knowledge of metabolic networks.

Direct Detection. However, direct detection of metabolites using cell-specific methods also is necessary because substrate availability and regulation of enzymes at the post-translational level influence flux through a pathway. Current methods to determine a metabolite's *in situ* abundance use both *direct* and *indirect* detection methods. Direct detection includes techniques such as fluorescence-activated cell sorting (FACS), which isolates a marked population of cells, or laser capture microdissection (LCM), which isolates material from sectioned tissue (Rogers et al. 2012; Moussaieff et al. 2013; Dong et al. 2016). More recently, direct ionization and detection of metabolites from thin sections of tissue have enabled spatial maps of dozens of metabolites from various plant tissues (Dong et al. 2016). Coupling such methods with metabolic labeling using radioisotopes has enabled direct detection of flux through particular metabolic pathways in individual cells (Kopf et al. 2016).

Indirect Detection. These methods include genetically encoded fluorescent protein–based sensors that enable the use of light-based microscopy to study *in situ* spatiotemporal dynamics of metabolite flux (Okumoto et al. 2012). Sensors based on Förster resonance energy transfer (FRET), another indirect detection method, encode proteins that change their conformation when bound to the metabolite of interest, inducing a change in the efficiency of energy transfer between fluorophores. Other indirect detectors are nonratiometric sensors, which simplify imaging requirements, as well as fluorescent sensors that can detect the actual activity of transporters moving small metabolites and flux across the plasma membrane (De Michele et al. 2013). The major limitation in using fluorescent sensors is the extensive investment required to design and test such proteins.

Re-Engineering Metabolic Pathways

Engineering metabolic pathways in microbes and plants has seen some success in reconstructing pathways necessary for drug synthesis (Lau et al. 2014; Lau and Sattely 2015). Introducing the morphine pathway into yeast, for example, required not only the expression of plant enzymes involved in the biosynthesis of this compound, but also ensured the correct localization of proteins and transport of intermediate substrates among cellular compartments (Thodey et al. 2014; Galanie et al. 2015). Thus, understanding how cells subcompartmentalize biosynthetic reactions between cells and their subcompartments is needed to reconstruct or redesign such synthetic pathways. The ability to localize specific enzymatic steps into these discrete organism domains may prove useful in avoiding side reactions that reduce biosynthetic efficiency or lead to toxic byproducts. The methods developed for transferring a plant metabolic pathway into yeast can be translated to plant-based bioproduct production.

5.2 Major Challenges in Detecting Metabolic Pathways

Metabolic networks often are illustrated as static wiring diagrams, but dynamic traffic maps may be a

more accurate representation. Current knowledge is limited to a small number of reactions revealed by assays lacking sufficient spatial or temporal resolution to understand the regulatory mechanisms that determine flux through these pathways. The undiscovered metabolome and associated enzymes and transporters that regulate the metabolites prevent an accurate understanding of how most high-value industrial and pharmacological bioproducts are made (Lau et al. 2014).

Uncharacterized metabolites can be considered biology's "dark matter." MS has enabled the direct detection of hundreds of compounds for which only the mass is known. Determining the exact chemical structure of these compounds remains an arduous task. Estimates are that an untargeted metabolomics experiment can be annotated for only about 1.8% of spectra (da Silva et al. 2015). Interactions between organisms occur in large part through the exchange of metabolites that determine whether such exchanges are symbiotic, parasitic, or pathogenic (Baran et al. 2015). The interface between organisms offers a rich domain on which to focus efforts aimed at understanding the significant role that metabolic pathways play in controlling ecosystem-level processes that affect agricultural productivity and sustainability. Engineering metabolic pathways likely will involve a combination of manipulating the activity and expressing endogenously encoded genes, as well as introducing genes from other species. Important for facilitating pathway engineering will be advancing the understanding of how metabolites themselves affect enzymatic activity and how the flux through a metabolic pathway is regulated in space and time (Hackett et al. 2016).

Developing Technologies for Studying Metabolite Fluxes

Plants, microbes, and fungi are dynamic systems in which the flux of a metabolite may be highly tuned to suit particular environmental conditions. Detected changes will lead to finding important functional differences that affect physiology and growth.

Discovery of Key Metabolomes for Organisms Relevant to Bioenergy. Research resources and consortia are needed for a coordinated effort to catalogue the metabolites of organisms important for bioenergy development. Included are biosynthetic pathways associated with biomass production, nutrient cycling, and carbon fixation, but equally important are key metabolites in plant-microbe interactions that are poorly understood. Characterizing plant-microbe properties requires technology development in several areas, including computational methods, and measurement methods at both the cellular and *in situ* field levels. Computational methods being developed will enable the determination of a chemical fingerprint for uncharacterized spectra using tandem MS, in which metabolites are fragmented to determine the mass of the fragments. Using available databases of compounds and their known or predicted fragmentation patterns, novel small molecules can be categorized based on their chemical constituents (Duhrkop et al. 2015). Determining the exact chemical structure can involve years of focused effort, but such computational methods may identify simpler mass features that can be elucidated, especially those seemingly contained in a large number of unknown metabolites.

Metabolite Measurement at the Cellular Level (Sensors and Direct Detection). Characterizing a system's metabolite flux is limited by constraints in the spatial and temporal resolution of detection. Several methods enable direct detection of metabolites, but access to the needed equipment is limiting. Nuclear magnetic resonance (NMR) spectroscopy has been used on plant tissue to characterize the flux of simple metabolites such as sugar and amino acids across plant stems (Kockenberger et al. 2004; Dong et al. 2016). Greater use of this or similar methods is needed to characterize a broader range of complex molecules in plants and microbes. Limiting the use of such methods are the spatial resolution and significant constraints on the size of objects that can be imaged. Raman and infrared spectroscopy hold promise for enabling *in situ* determination of the

chemical spectra of cells (Heredia-Guerrero et al. 2014; Dong et al. 2016). While Raman imaging methods use shifts in the wavelength of light scattered from objects, infrared imaging methods measure energy exchange between the infrared photons and vibrational motions of atoms in the functional groups of molecules. Nevertheless, both methods measure the nature of the chemical bonds present. Deconvoluting emission spectra often is challenging because most biological samples comprise hundreds to thousands of molecules. Determining the spectra of known chemicals and developing algorithms that more accurately deconvolve such data will be needed to make these imaging modalities accessible for *in situ* metabolite measurements. Acquiring additional downstream, spatially resolved MS measurements can further elucidate the chemical constituents.

FRET sensors (see Fig. 5.2, this page) can be created using a tool kit of DNA parts that enable assembly of dozens of different sensors. These designs can be rapidly evaluated using expression systems such as yeast or *Escherichia coli*. FACS and high-throughput sequencing enable rapid identification of constructs with sufficient expression levels and exhibit a strong change in FRET efficiency when binding to the chemical of interest (Jones et al. 2014). Currently, most FRET sensors are designed to detect a specific chemical. Also possible, however, are libraries of potential FRET sensors that use enzyme sequences or receptor domains. These sensor libraries could be screened for their responsiveness to a chemical library, enabling further optimization of experimentally useful sensors with appropriate binding affinities for *in vivo* measurements.

Metabolite Mechanisms of Synthesis and Transport. The identification of novel chemicals and enhanced ability to detect their spatiotemporal localization will be important for determining the mechanisms responsible for their synthesis, transport, degradation, and perception. Methods to rapidly screen mutant library collections of microbes

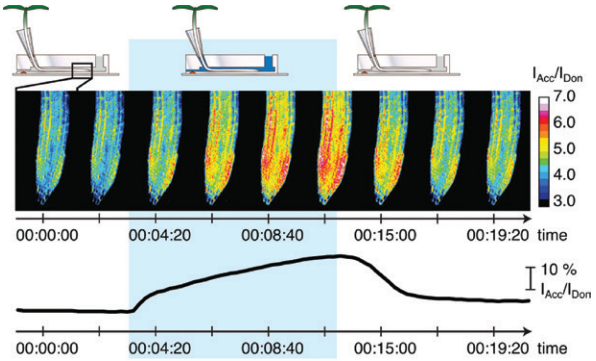


Fig. 5.2. Growth of Plant Seedlings in a Perfusion Chamber (RootChip) Environment, Allowing Highly Controlled Changes in the Root's Chemical Environment. The graphic illustration shows a fluorescence biosensor for glucose that is encoded by the plant and imaged. Changes in the biosensor's fluorescent property indicate when glucose enters the root cells. [Reprinted from Jones, A. M., et al. 2013. "*In Vivo* Biochemistry: Applications for Small Molecule Biosensors in Plant Biology," *Current Opinion in Plant Biology* **16**(3), 389–95. © 2013, with permission from Elsevier.]

and, more recently, algae and plants will facilitate the identification of genetic components for these pathways (Zhang et al. 2014). Use of chemical genetic screens has been successful in identifying signal transduction components (Park et al. 2009). Discovery of these molecular mechanisms will enable the transfer or reconstitution of biosynthetic pathways and greater control over metabolic pathways of interest to BER.

Field Deployment of Sensor Technologies. The physiological status of plants and microorganisms is highly context dependent. The influence of climate and the chemical environment on biological systems can be underestimated, especially when experiments are conducted under highly controlled conditions. Experimental systems must be designed based on actual environmental conditions that an organism might experience and that can vary across similar parameters (Rellán-Álvarez et al. 2016). Furthermore, organisms rarely grow in isolation from other organisms; thus, interactions among

plants, microbes, fungi, and animals also will exist in the context of a changing environment, and all these influences will affect the metabolites present in cells. Two complementary approaches are needed to ensure that observations of biological systems are physiologically relevant: (1) development of metabolite-detection methods that are compatible with in-field measurements and (2) establishment of fabricated ecosystems that enable controlled manipulation of complex synthetic communities and environments that simulate aspects of natural environments.

5.3 Predictively Understanding the Metabolome Across Scales, from Organisms to Ecosystems

Predictive Metabolome Understanding to Enable Synthetic Pathway Construction and Optimization. Understanding metabolomes and applying improved metabolome discovery tools to a diverse collection of organisms will aid in the development of a model for predicting how an organism's total chemical constituency is determined. Also making this model development possible will be a mechanistic understanding of the proteins that control metabolite biosynthesis, localization, degradation, and activity. The determination of genome sequence, along with transcriptomic and proteomic datasets, will enable predictions of how a cell's chemical environment may change with its environment or vary between genotypes. Ultimately, there is the possibility of predicting organisms within a system capable of synthesizing chemicals that are never actually detected. Cataloguing the genome sequences of diverse organisms within a system may then enable a predictive catalogue of the chemical diversity of life that preserves knowledge of biosynthetic pathways present in extinct or rare organisms.

Metabolite Economy: Whole-Ecosystem Accounting of Carbon and Nutrient Cycling. Matter cannot be created or destroyed, and the same is true for metabolites in an ecosystem. A full accounting of the carbon and nutrients entering an ecosystem is necessary to understand how terrestrial life influences the global carbon cycle and climate. Development of detection systems for *in situ* measurements and the establishment of realistic experimental systems will enable relevant measurements of metabolic exchanges at the interface between organisms. The identification of dependencies essential for maintaining diversity in ecosystems will also be enabled. Such metabolic exchange models may explain how pathogens or environmental contaminants cause changes in ecosystem diversity and may highlight ways to engineer these systems through targeted manipulation via chemicals or the introduction of alternative community members.

Chapter 6

Biosystems Design

Advances in high-throughput biology and biotechnology have led to an array of biological insights in diverse organisms, the environment, medicine, agriculture, and energy supply. The enormous breadth of genomic diversity that enables organisms to adapt to diverse environments also endows them with rich biosynthetic potential (Venter et al. 2004; Tringe et al. 2005). Harnessing this potential (e.g., by generating biologically derived chemicals, fuels, and materials to ensure environmental sustainability) could enable solutions to global challenges (Way et al. 2014). Achieving such goals, however, requires a thorough understanding of biological systems, as well as their safe and programmable control. In this regard, the ability to (1) modify genetic and biochemical molecules in cells, (2) apply advanced computation to design and analyze engineered biosystems, and (3) isolate engineered cells and communities with desired function remain defining challenges.

Fueled by concurrent advances in systems biology and the emerging field of synthetic biology, biosystems design provides a valuable approach for probing, studying, and introducing new functions into biological systems. Synthetic biology combines principles from biology, chemistry, physics, mathematics, and engineering to assemble the biological tools necessary to redesign biological systems. More specifically, synthetic biology employs engineering principles to reduce genetic information into DNA "parts," so that those parts can be understood in isolation and reassembled into new biological parts, devices, and whole systems to build desired or expanded functions in living

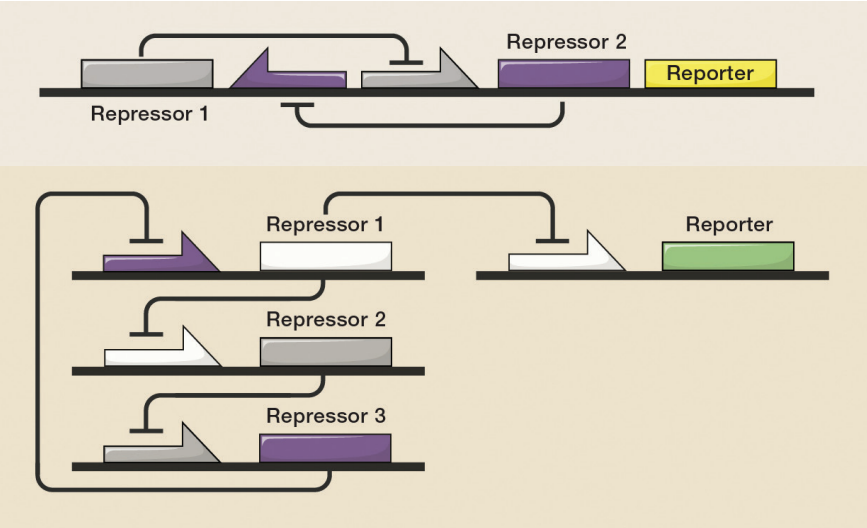


Fig. 6.1. Synthetic Biological Devices Based on Gene Regulatory Networks. (Top) The "toggle switch" is a co-repressive gene network in which two repressors turn each other off, permitting bistability of gene expression and memory. **(Bottom)** The "repressilator" comprises a network of three repressors that sequentially turn each other off, leading to oscillations in gene expression over time. [Reprinted from Way, J. C., et al. 2014. "Integrating Biological Redesign: Where Synthetic Biology Came from and Where It Needs to Go," *Cell* **157**(1), 151–61. © 2014, with permission from Elsevier.]

cells. In many respects, synthetic organic chemistry serves as a model for the nascent field of synthetic biology. As a discipline, synthetic organic chemistry shows that organic synthesis could be made into a rational science. This is a new paradigm for biology, in which a desired biological function is conceived, designed, and constructed to work as predicted, reliably and robustly using well-defined parts (see Fig. 6.1, this page). The field has rapidly advanced over the past 15 years with the use of (1) modular and well-characterized biological parts; (2) computer-aided design (CAD) software customized for biological systems; (3) metabolism re-engineering for production of high-value compounds (e.g., butanol and artemisinin); (4) large-scale gene synthesis; and (5) designer genomes containing refactored pathways and minimal or recoded genomes. Given the profusion of these advances, synthetic biology is poised to make fundamental breakthroughs in predictive understanding of complex biological systems and actualize an array of impactful applications that address global challenges in food and energy supply, environmental health, and medicine.

6.1 Current Science and Technologies

Applying Genome Technologies: DNA Sequencing, Gene Synthesis, and Genome Engineering

Next-generation DNA sequencing has revealed the complete genome sequences of numerous organisms, establishing a fundamental and growing understanding of genetic variation and phenotypic diversity. DNA sequencing technologies (that "read" genomes) have undergone revolutionary advances during the past decade, resulting in a precipitous drop in sequencing costs and coinciding with dramatic increases in throughput. Among many active genomics research areas, sequencing has advanced, leading to unprecedented insights into genetic variants associated with an observed phenotype, metagenomics, genome structure, and genome evolution. DNA sequencing and genome engineering are synergistic technologies that establish a platform for pursuing many goals in biological design. To date, reading capacity in the biological sciences has far outpaced the capability for "writing genomes," but recent developments underscore

Biosystems Design Needs

Designing New Tools for Biosystems Study and Safety Measures

Capabilities for modifying genetic and biochemical molecules in cells, applying advanced computation to the design and analysis of engineered biosystems and isolating the cells and communities having desired functions; and for addressing the primary safety concerns involving up to billions of genotypic variants of engineered biosystems.

- New tools to apply biosystems design methods to a broader range of organisms, including unusual microbes, environmental isolates, algae, and plants, as well as microbial communities and plant-microbe interfaces.
- Improved methods to explore extensive phenotype landscapes and link genotype to phenotype.
- Tools to study intractable, eukaryotic, and multicellular organisms.
- Methodologies to focus on biocontainment and genetic isolation.

Using Genomics Technologies, Computational Models, and Omics Studies

Technologies and more efficient tools for genome and epigenetic engineering adaptable to diverse organisms for conducting precise, error-free, and high-throughput genome manipulation, enabling comprehensive, pertinent biosystems design.

- Tools to combine with improvements in steady-state and time-variant computational modeling and omics studies, including flux-omics and metabolomics.
- Methods to rapidly assay function and fitness, together with technologies to manage the containment of engineered biosystems and isolate products of engineered pathways.
- Further development of organisms that have more orthogonal and alternative genetic codes to facilitate their implementation in model systems, diverse species, or even communities.

dramatic progress in this area (see Fig. 6.2, p. 48; Boeke et al. 2016).

Technologies for genome engineering (i.e., writing genomes) have lagged behind reading genomes, so the quest to introduce precise genome modifications at multiple genetic loci remains a defining challenge. The development of recombinant DNA (rDNA) technologies in the 1970s laid the foundation for current genome engineering technologies. Recombinant DNA insertions typically involve the transformation of plasmids—the insertion of circular rDNA molecules capable of replicating alongside the host's genome or direct insertion into the host's genome using homologous recombination—and typically are limited to single or few gene modifications.

The confluence of three complementary technologies has fueled advances in genome engineering (see Fig. 6.3, p. 49). First, *de novo* DNA synthesis technologies enable tailored construction of user-defined, double–stranded DNA segments. Similar to advances in DNA sequencing, DNA synthesis has undergone logarithmic improvements in scale, cost, and throughput. For example, large-scale DNA microchip–based synthesis methods permit high-density synthesis of about 10^5 customized single-stranded DNA (oligonucleotides). Gene synthesis (using overlapping oligonucleotides), combined with improvements in DNA error-correction methods, has enabled high-quality and cheap construction of designer synthetic genes. Second, and complementary to DNA synthesis, large-scale DNA assembly *in vivo* and *in vitro* methods permit

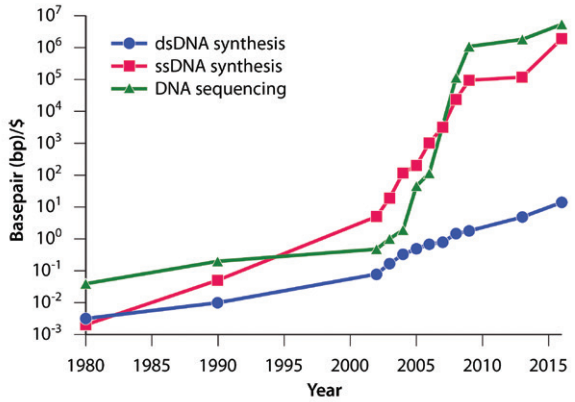


Fig. 6.2. Temporal Advances in DNA Sequencing and Synthesis Technologies. Costs ($/base pair) of DNA sequencing (green), single-stranded DNA (ssDNA) synthesis (red), and double-stranded DNA (dsDNA) synthesis (blue) are plotted for the past 35 years. [From Boeke, J. D., et al. 2016. "The Genome Project-Write," *Science* **353**(6295), 126–27. Reprinted with permission from AAAS.]

the precise assembly of individual synthetic genes into higher-order combinations at the network and whole-genome scales. Advanced recombination and transplantation techniques are being developed to improve the efficiency of introducing synthesized genes, pathways, and genomes into target organisms. Finally, genome editing permits targeted changes directly in the chromosomes of living cells. Many genome-editing technologies generate DNA double-strand breaks at targeted loci to introduce genomic modifications.

There are four main classes of sequence-specific nucleases (SSNs) used to introduce genome modifications (reviewed in Voytas and Gao 2014): (1) meganucleases, (2) zinc finger nucleases (ZFNs), (3) transcription activator-like effector nucleases (TALENs), and (4) a nuclease system consisting of clustered regularly interspaced palindromic repeats (CRISPRs) and an RNA-guided protein called Cas9 (CRISPR-associated protein-9). Meganucleases were discovered first, followed by ZFNs and TALENs. ZFNs and TALENs recognize specific DNA sequences through protein-DNA interactions and use the *FokI* nuclease domain to introduce double-strand breaks at genomic loci. Construction of functional ZFNs and TALENs with desired DNA specificity, however, remains laborious, costly, and primarily limited to modifications at a single genetic locus. CRISPR-Cas9 has been broadly adopted for multiplexed targeting of genomic modifications because the CRISPR nuclease Cas9 uses a short-guide RNA to recognize the target DNA via Watson-Crick base pairing and has been shown to function in many organisms (Jinek et al. 2012; Cong et al. 2013; Jinek et al. 2013; Mali et al. 2013). The CRISPR-Cas9 nuclease system is an innate bacterial defense mechanism against viruses and plasmids that uses RNA-guided nucleases to cut foreign DNA sequences, thereby disabling them. Scientists have re-engineered the CRISPR-Cas9 system so that a single RNA (the guide RNA) can create the Cas9-mediated cut of a target sequence in a genome. This system's ease of design, specificity, and simplicity have made it a popular technique for generating future biotechnological products, but certain features (e.g., the creation of double-strand DNA breaks) are cytotoxic and limiting. Other genome-editing approaches are still emerging. For example, multiplex automated genome engineering (MAGE) permits multisite and combinatorial genome modifications through hybridization of synthetic oligonucleotides during the process of DNA replication (Wang et al. 2009). Genome-engineering technologies have enabled organism engineering at the gene, network, and genome levels, resulting in the *de novo* synthesis of refactored biosynthetic pathways (Smanski et al. 2014), whole-genome synthesis of bacterial genomes (Gibson et al. 2010; Hutchison et al. 2016), and whole-genome recoding (see Fig. 6.4, p. 50; Isaacs et al. 2011; Lajoie et al. 2013; Ostrov et al. 2016).

Biological Applications

Scientists have documented several recent success stories of synthetic biology applications to biotechnological products and processes. A prominent

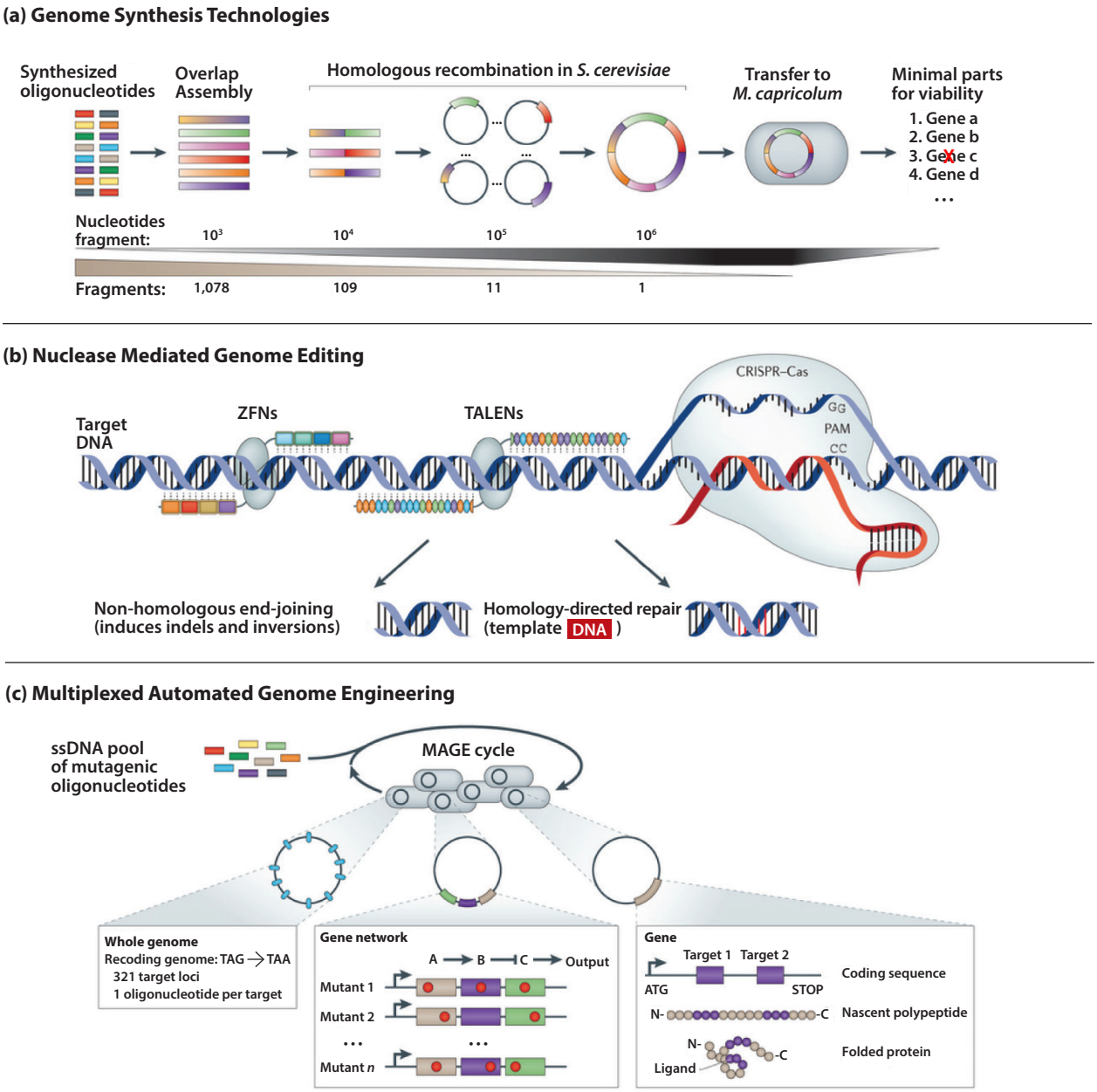


Fig. 6.3. Genome Engineering Enabling New Possibilities in Biosystems Design. The rapid emergence of genome engineering is driving unprecedented advances in biosystems design. Genome engineering establishes the ability to generate precise chromosome modifications, elucidate causal links between genotype and phenotype, and enable the design and reprogramming of organisms. **(a)** *De novo* DNA synthesis permits customized construction of synthetic DNA fragments. Large-scale DNA assembly methods assemble synthetic genes into pathway- and genome-scale constructs. Genome-editing technologies introduce modifications at targeted genetic loci of diverse organisms. **(b)** Nuclease-mediated genome-editing technologies generate DNA double-strand breaks at targeted loci to introduce genomic modifications. In particular, the CRISPR-Cas9 system has been broadly adopted for multiplexed targeting across genomic loci in many organisms. **(c)** Multiplex automated genome engineering (MAGE) introduces multisite and combinatorial genome modifications through hybridization of synthetic oligonucleotides during replication. MAGE technologies permit modifications at the gene, pathway, and genome levels, leading to refactored biosynthetic pathways, whole-genome synthesis of bacterial genomes, and recoding of whole genomes. [Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics. Haimovich, A. D., P. Muir, and F. J. Isaacs. 2015. "Genomes by Design," *Nature Reviews Genetics* **16**, 501–16. © 2015.]

Key: CRISPR, clustered regularly interspaced short palindromic repeats; **TALENS**, transcription activator-like effector nucleases, **ZFN**, zinc finger nucleases.

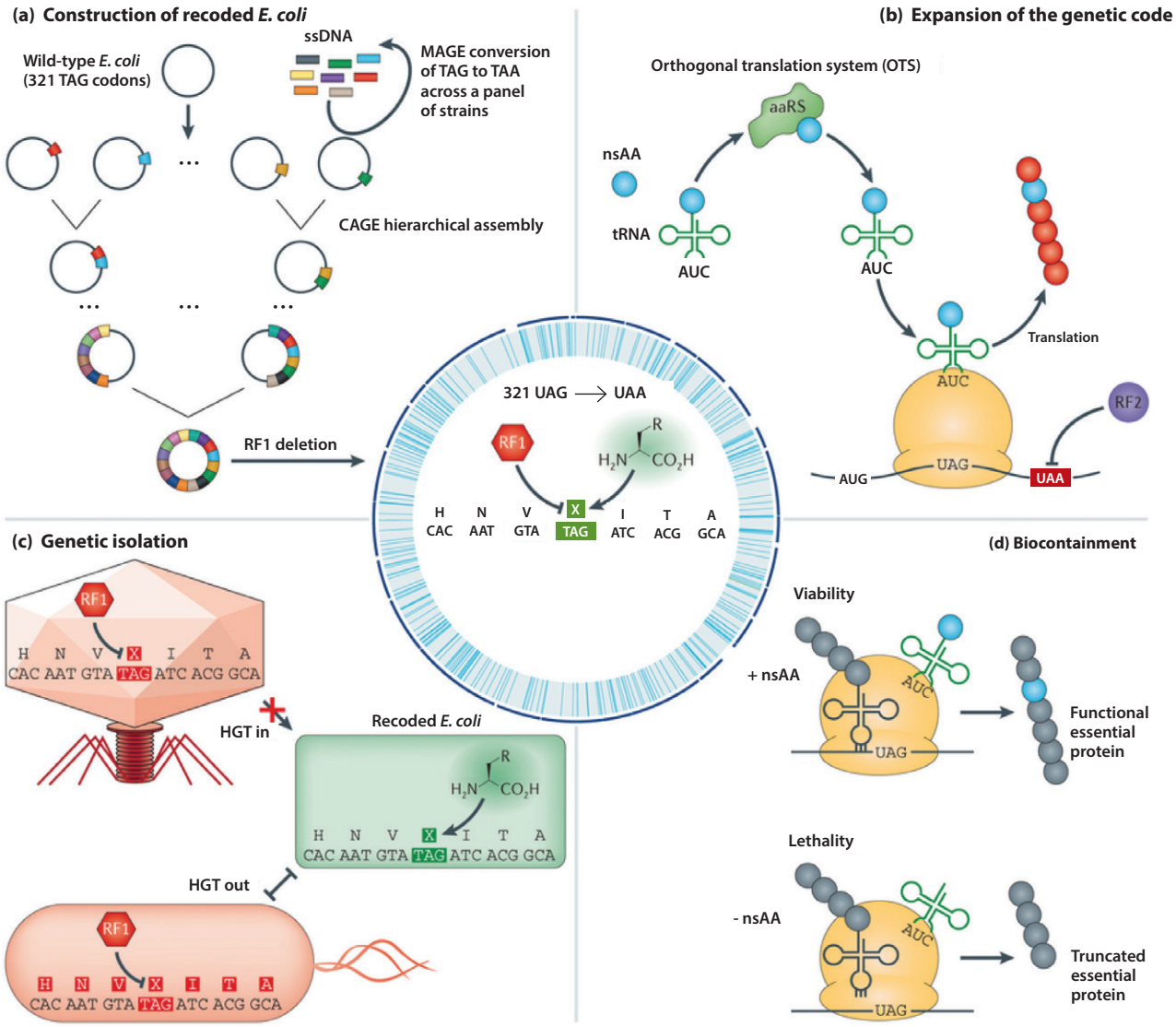


Fig. 6.4. Genomically Recoded Organisms (GROs): "A Living Foundry for Making New Materials." (a) The GRO was constructed using multiplex automated genome engineering (MAGE) to convert nucleotide triplet TAG stop codons to TAAs in a panel of strains. Conjugative assembly genome engineering (CAGE) then was used for the hierarchical assembly of these partially recoded genomes into a fully recoded genome in which all 321 TAG stop codons are reassigned. Release factor 1 (RF1) then was deleted to establish TAG as an open (instead of terminating) codon for incorporating nonstandard amino acids (nsAAs). **(b)** Adding an orthogonal translation system capable of changing an nsAA to a TAG-recognizing transfer RNA (tRNA) enables the genetic encoding of nsAA-incorporating proteins. **(c)** The reassignment of TAG from a stop to an open codon can render TAG-containing or TAG-terminating genes nonfunctional in a different organism, reducing the probability of horizontal gene transfer (HGT) between recoded and natural organisms. **(d)** Incorporating TAG codons in essential genes provides a method for the biocontainment of recoded organisms by linking their viability to the presence of an exogenous nsAA. aaRS, aminoacyl-tRNA synthetase. [Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics. Haimovich, A. D., P. Muir, and F. J. Isaacs. 2015. "Genomes by Design," *Nature Reviews Genetics* **16**, 501–16. © 2015.]

example is the microbial synthesis of artemisinic acid, a precursor to the antimalarial therapeutic artemisinin (see Fig. 6.5, p. 52). Artemisinin is naturally produced by the plant *Artemisia annua* in relatively miniscule quantities, and both synthetic biology and organic synthesis were explored as ways to amplify production. Much effort on the synthetic biology front resulted in artemisinic acid titers as high as 25 g per liter in yeast (Paddon et al. 2013). This effort required overcoming multiple technical barriers, such as optimizing enzyme expression to fine-tune metabolic flux toward artemisinic acid as well as selecting and using the directed evolution of enzymes that oxidized precursor molecules to artemisinic acid (Paddon and Keasling 2014). In another example, high-throughput genome engineering was used to rapidly identify genetic variants that overproduced the phytochemical lycopene. Specifically, MAGE resulted in the daily generation of billions of variants, and those variants that overproduced lycopene fivefold were identified within days (Wang et al. 2009). Another example of genome engineering employed combinatorial design to reorganize the gene cluster responsible for nitrogen fixation in the bacterium *Klebsiella oxytoca* (Smanski et al. 2014), a process whose enhancement may have significant promise for agriculture (see Fig. 6.6, p. 53). This process resulted in more than 100 synthetic variants of a complex gene cluster, from which an enzyme assay and RNA sequencing were used to identify variants with significant activity (Smanski et al. 2014). Other notable examples are too numerous to list here. At the same time, several computational developments are making valuable contributions to biosystems design: *in silico* modeling (e.g., Fiore et al. 2016), model-guided design of biosystems (e.g., Schreiber et al. 2016), and software tools for biosystems design (e.g., Oberortner and Densmore 2015).

Overall, industrial biotechnological research and development of genetically modified (GM) organisms have been major contributors to the U.S. economy. Biotechnology revenues in 2012 totaled more than $300 billion and were apportioned roughly equally among drugs synthesized by GM organisms; GM crops; and industrial molecules including fuels, chemicals, and biotechnological food products and feedstocks (Carlson 2016). A remarkable trend is that bioproducts, and not biofuels, are now the largest group of industrial biotechnological products. Biochemicals are estimated to contribute about 0.4% of U.S. gross domestic product, compared with roughly 3% for petroleum-derived chemicals (Carlson 2016).

6.2 Major Challenges in Design of Engineered Biosystems

Synthetic Biology Applied to Intractable, Eukaryotic, and Multicellular Organisms. Most synthetic biology applications to date have focused on model prokaryotic bacteria or simple eukaryotes such as yeast. Biosystems design relevant to mission areas of the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) warrants the development of tools for intractable, eukaryotic, and multicellular organisms. These organisms include, but are not restricted to, phototrophs (e.g., cyanobacteria), archaea, extremophiles, microbes isolated from biofilm mats, algae to herbaceous plants, and woody plants (e.g., trees). Such systems present new challenges for synthetic biology applications compared to those for model prokaryotes. Another challenge is epigenetic regulation of gene expression, a phenomenon infeasible in prokaryotic biosystems but often observed in eukaryotes (Rodriguez-Escamilla et al. 2016).

Engineered Communities of Microorganisms and Microbe-Plant Interfaces. Biosystems consisting of multiple, interacting organisms can use the metabolic conversion capabilities of the constituent organisms to yield a more efficient process than is feasible in a single engineered organism. Already, synthetic biology approaches reportedly have been used on fungi-bacteria (e.g., Minty et al. 2013) and bacteria-yeast (e.g., Hu et al. 2016) consortia. New tools that target microbial consortia are needed to

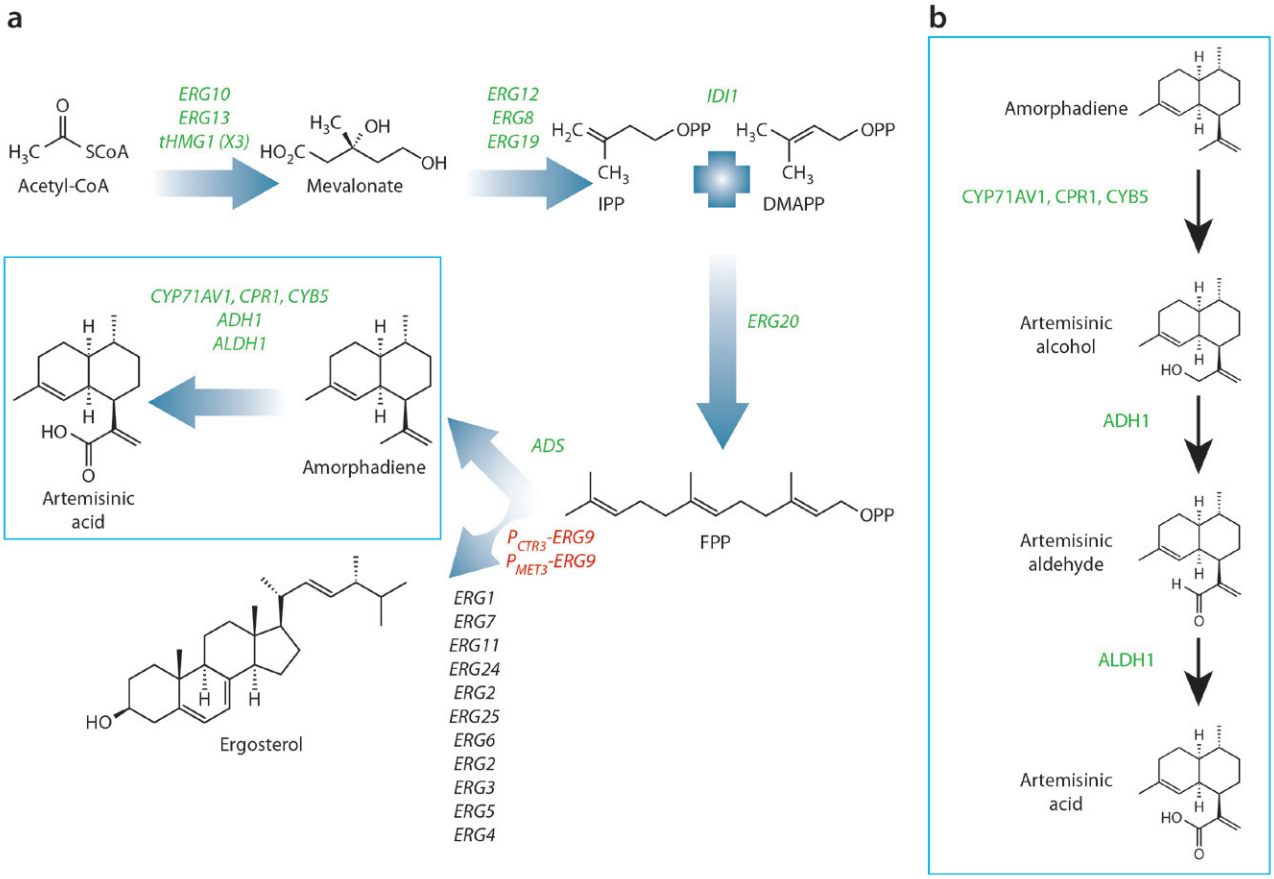


Fig. 6.5. Artemisinic Acid Production Pathway in *Saccharomyces cerevisiae*. (a) Overview of the artemisinic acid production pathway. Galactose (GAL)-induced genes for overexpression (green). Copper- or methionine-repressed squalene synthase (ERG9) (red). (DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; IPP, isopentenyl diphosphate) tHMG1 encodes the truncated HMG-CoA reductase. **(b)** The full three-step oxidation of amorphadiene to artemisinic acid from *Artemisia annua* expressed in *Saccharamyces cerevisiae*. CYP71AV1, CPR1, and CYB5 oxidize amorphadiene to artemisinic alcohol; ADH1 oxidizes artemisinic alcohol to artemisinic aldehyde; ALDH1 oxidizes artemisinic aldehyde to artemisinic acid. [Reprinted by permission from Macmillan Publishers Ltd: Nature. Paddon, C. J., et al. 2013. "High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin," *Nature* **496**(7446), 528–32. © 2013.]

advance research in this area. Similarly, engineering metabolite exchange between soil microbes (e.g., rhizobia) and plant roots, such as in symbiotic nitrogen fixation, requires genetic tools optimized for plant-microbe interfaces.

Exploration of Phenotype Landscapes Resulting from Genome Engineering. Genome engineering, through *de novo* DNA synthesis and *in vivo* genome editing, can produce hundreds to billions of genetic variants of a natural biosystem. Because phenotypes are fundamentally unpredictable, exploring the combinatorial space of genotypes is essential, rather than focusing on a single, static design. The scale of phenotype exploration is a factor of two variables—ease of genotypic diversification and power of selection. Combining design, *de novo* synthesis, and *in vivo* evolution will lead to the creation of large and

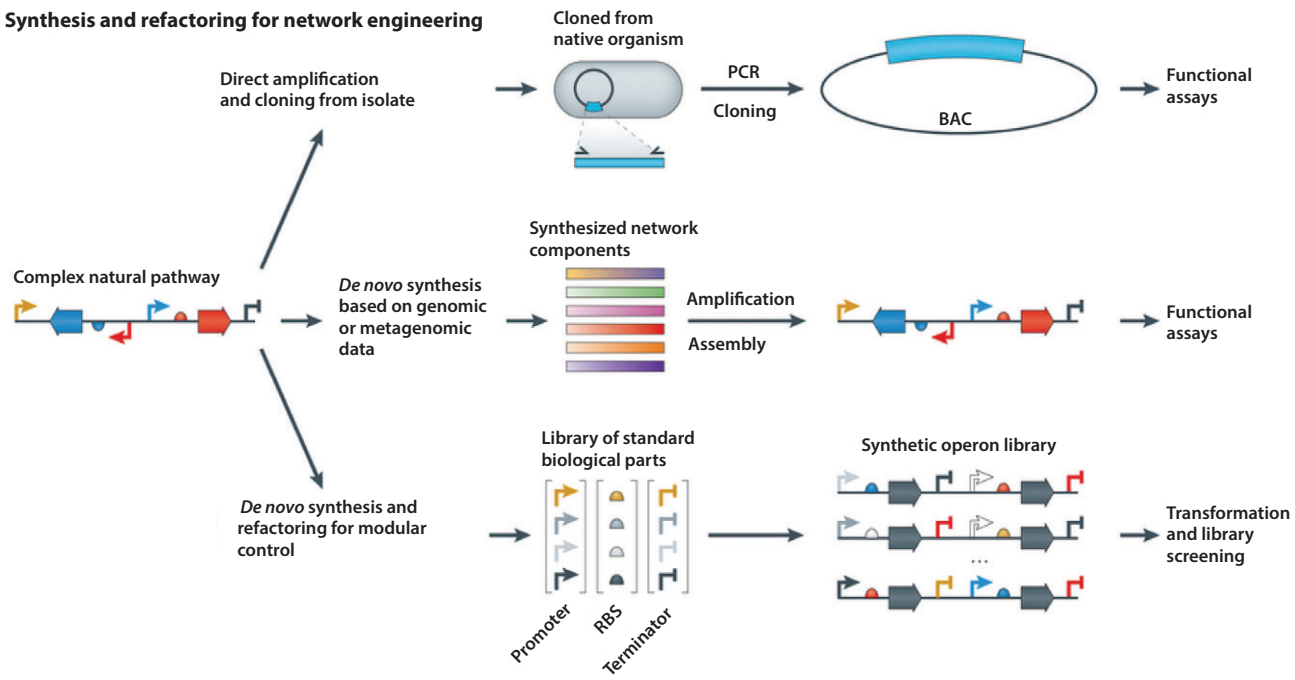


Fig. 6.6. Synthesis and Refactoring of Biosynthetic Pathways. Numerous approaches have been pursued to isolate, transfer, and characterize gene clusters. Biosynthetic clusters can be cloned from their native organism and transferred to another organism for heterologous expression. For organisms that are recalcitrant to laboratory culture techniques, gene sequences of interest can be identified from genomic or metagenomic databases and subsequently synthesized and introduced into a production organism. The native regulatory elements of a heterologous pathway may not be optimal for expression in the host organism. Refactoring approaches seek to rebuild the pathways using well-characterized modular regulatory elements and removing all native regulation. [Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics. Haimovich, A. D., P. Muir, and F. J. Isaacs. 2015. "Genomes by Design," *Nature Reviews Genetics* **16**, 501–16. © 2015.]

Key: BAC, bacterial artificial chromosome; **PCR**, polymerase chain reaction.

diverse populations, though both computation and phenotype selections will be key limitations.

Isolation of Engineered Organisms with Desired Functions. Biosystems design goals often seek to engineer a prescribed function (e.g., maximized production of a chemical or biofuel). Front-end design technologies (e.g., CAD software), genome engineering, and omics-level datasets have enabled the generation of many engineered variants of biosystems and also established the next phase of challenges. Prominent among these challenges is the development of biomolecular technologies capable of isolating engineered organisms with desired functions. Genome-engineering technologies are now establishing unprecedented and powerful capabilities to enact precise genome modifications across a growing suite of organisms, though the task of characterizing and isolating optimal variant(s) with a sought-after function is daunting. Typical efforts involve costly, laborious, and time-consuming omics screening of $\geq 10^2$ clones, generating large genomic, metabolomic, and proteomic datasets that require analysis.

Safeguarding Engineered Biosystems. Over the past decade, synthetic biology has fueled the emergence of GM organisms with increased sophistication as common and valued solutions in clinical,

industrial, and environmental settings (Pieper and Reineke 2000; Steidler 2003; Way et al. 2014). This advancement necessitated safety and security measures, which were first outlined in the 1975 Asilomar conference on recombinant DNA (Berg et al. 1975). These include the development of robust solutions to intrinsic biocontainment and genetic isolation. While guidelines for physical containment and safe use of organisms have been widely adopted, intrinsic biocontainment—biological barriers limiting the spread and survival of microorganisms in natural environments—remains a defining challenge. Safeguards have been designed (1) to control cell growth by essential gene regulation (Kong et al. 2008), (2) as inducible toxin switches (Szafranski et al. 1997), and (3) as engineered auxotrophies (Steidler et al. 2013). Yet these approaches are compromised by cross-feeding of essential metabolites, leaked expression of essential genes, or genetic mutations (Jensen et al. 1993; Ronchel and Ramos 2001). Similarly, genetic isolation of microorganisms by preventing the functional exchange of DNA through horizontal gene-transfer events will help stabilize engineered biosystems from environmental threats. Recent advances over the past 2 years have presented numerous safeguard solutions, but they remain limited to model microorganisms (e.g., *Escherichia coli* and *Saccharomyces cerevisiae*). Key challenges remain for their implementation in both closed- and open-system applications (e.g., remediation), as well as their development in nonmodel organisms, multicellular species (e.g., plants), and across communities.

Identifying Technological Needs to Support Predictive Biology

Genome-Engineering Technologies Adapted to Diverse Organisms and Addressing Epigenetic Engineering. As the efficiencies of genome-engineering technologies continue to improve, their function increases. They expand to adapt to diverse species, requiring new tools to complement these technological advances. Specifically, new biomolecular tools (e.g., greater numbers of engineered Cas9 protein variants) are needed to (1) improve genomic targeting capable of enacting precise, base pair–level precision; (2) reduce or eliminate unintended off-target genome modifications; and (3) increase multiplex genome-editing capabilities that generate combinatorial variation. Additionally, these capabilities must expand beyond genome modifications to other important forms of regulation (e.g., epigenetics, transcription, and translation) to fully enable comprehensive biosystems engineering of cells with sophisticated functions.

Computational Modeling Interfaced with Omics and Imaging Measurements to Develop Integrated Prediction Tools. Synthetic biosystems ideally are developed via a design-build-test-learn cycle (see Fig. 6.7, p. 55) (Paddon and Keasling 2014; Petzold et al. 2015). Analytical tools involved in this process include computational toolboxes for initial biosystem design, omics and imaging measurements for testing the biosystem, and further computational analysis to interpret the large-scale measurements that link genotype to phenotype. Design tools include flux balance analysis (FBA), which can be used to predict carbon rerouting through large metabolic networks at steady state (Orth et al. 2010) and is available for community-developed systems (e.g., OptForce; Ranganathan et al. 2010). Successful applications of such tools to genetically engineered systems have been reported (e.g., Ranganathan et al. 2010). Design tools also include those that predict the dynamics of time-variant biosystems (Lomnitz and Savageau 2016). Omics tools ranging from transcriptomics (e.g., RNA-seq), proteomics, phosphoproteomics, metabolomics (e.g., discussed in Chapter 5. Metabolic Pathways in Plants, Microbes, and Fungi, p. 37), and isotope-based metabolic flux analysis (MFA) are enormously beneficial for assessing the impact of gene or genome engineering. Of these, transcriptomics and metabolomics are high throughput and mature, whereas proteomics and phosphoproteomics are still undergoing development. Isotope-based MFA (reviewed in Antoniewicz 2015) is a powerful method to measure

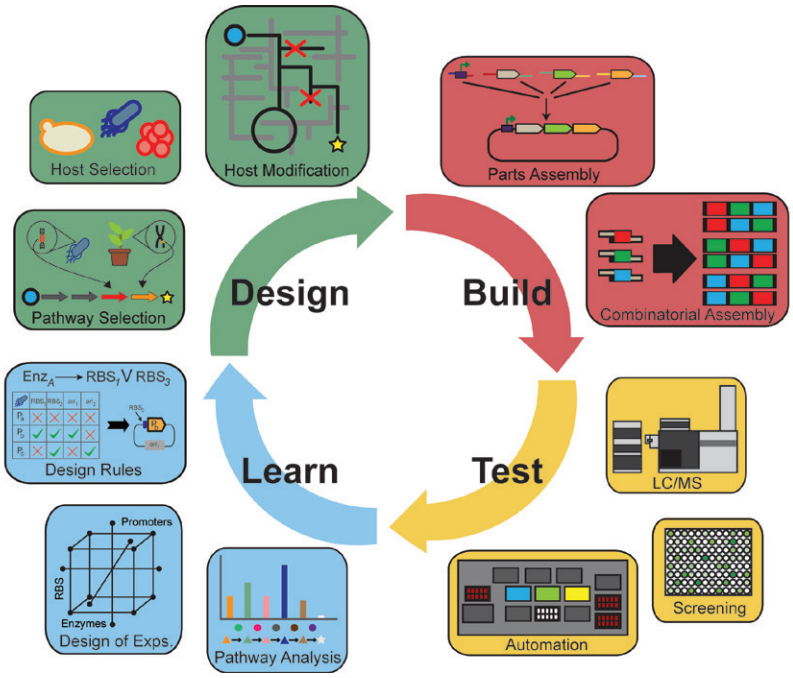


Fig. 6.7. The Design-Build-Test-Learn (DBTL) Cycle. DBTL is a basic approach for efficient problem solving that has been adapted for engineering biological systems. The **design** component defines the problem, establishes an approach to solve the problem, and identifies the biological components needed to build or modify. The **build** component synthesizes, assembles, or edits (or all three) the components of the engineered biological system. The **test** component characterizes the different biological systems and identifies the variants with the prescribed behavior. The **learn** component analyzes the test data and informs subsequent iterations of the cycle. [Reprinted under a Creative Commons Attribution License from Petzold, Chan, Nhan, and Adams.]

carbon flow and partitioning in metabolically engineered biosystems. It has been applied to complex systems including plants (e.g., Nargund et al. 2014) and simple organismal consortia (Ghosh et al. 2014; Mandy et al. 2014). The extension of such tools to complex consortia prevalent in nature is important for testing the performance of these biosystems. Meaningful interpretation of the raw data generated by omics tools requires extensive modeling and computation. Interfacing the ensuing results with further computation ultimately will enable successful prediction of regulation and genotype-phenotype relationships. Although predictive tools along these lines already have been reported, substantially more development and integration are essential before routine, accurate predictions are achieved for complex, multiorganismal biosystems.

Selection Assays. With the rapid advent of powerful biosystems design, diversification, and manipulation (e.g., genome engineering) and screening technologies, "selection" is the missing and necessary component for achieving maximum phenotypic benefit and functionality. Thus, the next key technological challenge to overcome is isolating the cell(s) among a large (~10^9) population that yields a desired phenotype. The development of biomolecular sensors and selectors capable of reporting and selecting cells expressing desired biomolecules from engineered microorganisms could address this challenge. Drawing inspiration from biology, such solutions could establish artificial Darwinian selection systems that integrate investigator-defined genetic diversification, selection, and isolation of optimal strains.

Biocontainment and Genetic Isolation: Auxotrophies, Alternative Genetic Codes, and Microcompartments. Recent studies in biocontainment (Cai et al. 2015; Gallagher et al. 2015; Mandell et al. 2015; Rovner et al. 2015; Chan et al. 2016) and genetic isolation (Lajoie et al. 2013; Ma and Isaacs 2016) have developed significantly improved solutions for safeguarding engineered biosystems.

Even so, many challenges and technological needs remain unmet. They include (1) testing of existing safeguards in real-world applications (e.g., bioremediation), (2) translating safeguards to nonmodel organisms and communities, and (3) scale-up of testing and implementation in industrial settings to determine optimal utility and identify failure points. The development of new safeguard solutions remains a high priority. For example, promising avenues are restricting growth to synthetic molecules or engineering microcompartments, even in prokaryotes (Sargent et al. 2013), to sequester proteins or bring enzymes and substrates into close proximity (Choudhary et al. 2012). Potentially, these structures can become biocontainment solutions by sequestering toxic products that otherwise would endanger the host or its environment. Additionally, a set of notable papers that describe the development of genomically recoded organisms (GROs) has established promising and robust solutions for biocontainment and genetic isolation through the design and construction of organisms possessing alternative genetic codes (Lajoie et al. 2013; Mandell et al. 2015; Rovner et al. 2015; Ma and Isaacs 2016). Further development of organisms with more orthogonal and alternative genetic codes in model systems, as well as their implementation in diverse species or even communities, remains an unmet technological area whose realization could have a big impact on DOE goals and applications in energy supply and the environment.

6.3 Leveraging Synthetic Biology for Future Energy and Resource Needs

Achieving the advances in biosystems design outlined in this chapter is expected to widely benefit DOE's mission of planning for future energy and resource needs. Harnessing the novel biological and biochemical functions of biology-based sources that use fermentative and photosynthetic organisms entails the engineering of plants, microbes, and organismal communities. Plants must be engineered to (1) improve photosynthetic rate and yield by incorporating carbon-concentrating mechanisms and alleviating photorespiratory carbon losses, (2) enhance water- and nutrient-use efficiency (especially nitrogen and phosphorus), (3) combat abiotic environmental stresses, and, most importantly, (4) facilitate readily degradable cell walls. Microbial biosystems design must leverage fermentative and photosynthetic species to target their high-efficiency deconstruction of cellulose and other cell wall components and synthesis of biofuels or bioproducts, coupled with engineering for high-yield and end-product productivity as well as improved tolerance to end-product toxicity. Finally, engineered microbial biosystems must be able to synthesize compounds not attainable via conventional organic chemical synthesis, as well as a wide variety of similar compounds (e.g., higher carbon content alcohols or hydrocarbons) with product variety that can be controlled according to need and produced on demand.

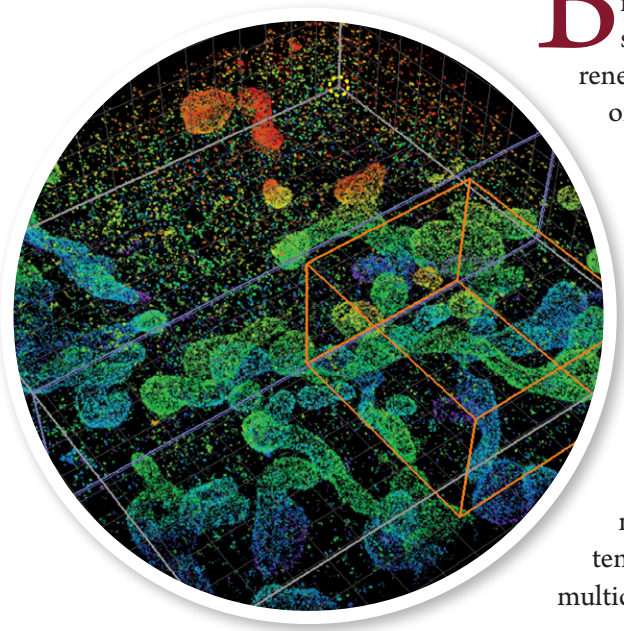
Highly desirable applications of complex engineered biosystems such as microbial and microbe-plant communities include solutions to ecological challenges such as bioremediation, carbon dioxide (CO_2) capture and storage, and nitrogen fixation, all of which are energy-intensive processes. In this context, researchers envision using prokaryotes for carbon capture (Hicks et al. 2017) and plant-invading bacteria for symbiotic nitrogen fixation (Hicks et al. 2017). Engineering for CO_2 sequestration and reduction may benefit from the decoupling of electron carrier (e.g., nicotinamide adenine dinucleotide phosphate or NADPH) regeneration from carbon assimilation. Microbes that perform electrosynthesis [i.e., directly use electrons from an electrode toward metabolism (Rabaey and Rozendal 2010; Nevin et al. 2011)] may be incorporated into biosystems that harness electrical energy. Engineering to facilitate nitrogen fixation may benefit from a detailed understanding of carbon-nitrogen interactions in metabolic networks of individual or symbiotic nitrogen-fixing biosystems.

Accomplishing these goals may enable synthetic biology to move from the engineering of individual microbes that produce specific molecules to engineering of microbial communities and microbe-plant symbioses (as in the rhizosphere). This trajectory will require the simultaneous development of imaging, metabolomic, fluxomic, and other omics tools that can measure metabolism and molecular exchange among communities of organisms. Progress is being made toward developing a broader set of foundational technologies to engineer model and undomesticated microbes, plants, and microbial communities. Thus, the new tools discussed in this report—particularly, imaging, measurement, and characterization techniques—will be needed to test designer biosystems in organisms.

These new technologies are expected to (1) enable the precise probing and manipulation of genetic and biomolecular processes, (2) aid the optimization of biosystems design by providing an understanding of the spatiotemporal dependencies of engineered pathways, (3) enable identification of bottlenecks in metabolic networks, and (4) aid in gaining an understanding of how incorporating new functions produces unintended impacts on the metabolism of individual organisms or an organismal consortium. Ultimately, the iteration between the design and large-scale phenotypic characterization of biosystems will lead to an understanding of how genomes translate into function, another goal of BER's Biological Systems Science Division.

Chapter 7

Cellular Ultrastructure and Physiology



Biological systems can be metabolically manipulated for a variety of beneficial purposes, including the sustainable production of fuels and chemicals from renewable plant biomass. Maximizing yields from native or engineered biological systems while preventing unintended consequences to other systems or the environment takes significant effort and is a major focus for researchers supported by the Biological Systems Science Division within the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER). Achieving the depth of understanding needed to control biosystem output and productivity is a daunting challenge that will require characterization of processes occurring among vastly diverse molecular species and across a range of spatial and temporal scales—from the atomic to the cellular and multicellular levels.

7.1 Science and Technologies

This chapter reviews current technologies for characterizing biological systems on the atomic to cellular scales and identifies technological shortfalls responsible for knowledge gaps in key research areas. Given the breadth of science covered and the brevity of this report, needed capabilities are discussed in terms of the primary data types generated and their optimal spatial, temporal, and chemical capabilities.

Using Structural Imaging to Characterize a Biosystem's Organization

Many different biophysical measurements can provide spatial and temporal information, but structural imaging is the most direct means of characterizing a biosystem's organization. The bioimaging community has a multitude of different techniques at their disposal, each of which operates optimally over a well-defined range of specimen sizes and spatiotemporal resolutions.

Atomic and Molecular-Level Structure and Dynamics

The ability to probe biological phenomena at the nanoscale and relevant temporal scales has been transformed by new developments in technologies: cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryoET), nuclear magnetic resonance (NMR), neutron macromolecular crystallography (NX), small-angle neutron scattering (SANS), small-angle X-ray scattering (SAXS), and macromolecular X-ray crystallography (MX). DOE investments in user facilities have accelerated the application of many of these foundational tools while fostering a broad user base. Despite such advances, development in these areas is still needed to improve resolution, sampling, and throughput; bridge the remaining technology gaps; and push the frontiers of structural imaging.

Transmission Electron Microscopy (TEM). Technologies employing TEM (see box, Electron Microscopy Terms, this page) produce three-dimensional (3D) images of biological specimens in the macromolecular to cellular scale range. The recent revolution in single-particle cryoEM technologies and approaches has dramatically increased the spatial resolution potential for macromolecule and macromolecular complex samples. This can be attributed to the development of (1) direct electron detectors (Li et al. 2013; Wang et al. 2014), (2) computationally assisted data collection in modern electron microscopes (Lyumkis et al. 2013), and (3) image-processing software developed in academic laboratories (DiMaio and Chiu 2016; Ludtke 2016; Scheres 2016). Continuing development of better recording devices and electron optics, such as phase plate technology, promises to deliver better resolving power and will generate even larger datasets. Advances already achieved in those areas have led to the structural determination of increasingly complex molecules (see Fig. 7.1, p. 62), including (1) multicomponent molecular machines, at near-atomic to atomic resolution (Li et al. 2013; Wang et al. 2014; Glaeser 2016) and (2) the ability to view such molecular machines on and inside cells (Guerrero-Ferreira et al. 2011; Hu et al. 2013; Mattei et al. 2016). Despite the dynamics in vitrified specimens being limited to time scales longer than the process of freezing (~0.1 milliseconds), the samples' natural heterogeneity reflects the conformational landscape of several macromolecular complexes and has enabled some exciting discoveries (Amunts et al. 2015; Bai et al. 2015; Nguyen et al. 2016). Continued improvements, such as streamlining the methodology for steps ranging from sample preparation to data acquisition and computational reconstruction, are needed to empower wider adoption and increase throughput of atomic-resolution structure determination. In addition, emerging new TEM capabilities on the millisecond to nanosecond time

Electron Microscopy Terms

SEM – Scanning electron microscopy

TEM – Transmission electron microscopy

Cryo – Vitrification of aqueous samples at liquid nitrogen or lower temperatures

Single-particle cryoEM – Many individual projections are combined to form a 3D image of a biomolecule

Tomography – Many images taken from the same cell or cell section as it is being tilted are computationally recombined to create a 3D volume

FIB – Focused ion beam milling, used to prepare thin samples for tomography or combined with SEM

Phase plate – Technology for increasing image contrast

Cellular Ultrastructure and Physiology Needs

Probing Biosystem Dynamics with Experimental and Computational Methods

Understanding of contributors to biosystem dynamics, such as enzyme function incurred by structure, membrane composition, and localization of components within a cell, to provide the foundation for functional models that bridge chemistry and biology.

- Advanced experimental methods, combined with computation, to probe atomic and molecular structure.
- Streamlined transmission electron microcrospy methodologies, from sample preparation to data acquisition, and computational reconstruction improvements in light-source peak brightness.
- Methodologies to permit atomic-resolution, pump-probe imaging of two-dimensional crystals and, ultimately, single-particle macromolecular complexes.
- Advances and rejuvenations to "mature techniques" that can benefit from insights gained from new technologies such as X-ray free-electron lasers.

Imaging Whole Organisms

Advancements to understand cells, the functional units of most biosystems, and their intracellular organization; and how cellular phenotypes respond to variations in environmental conditions and genetic and chemical modifications.

- New specimen-preserving methods for *in situ*, dynamic, or cryogenic imaging of whole organisms to determine the highly heterogeneous organization of cells with nanoscale resolution for theory and model refinement.
- Methods to enable *in situ* analysis of hydrated and living specimens.
- Advancements in visual proteomics to pinpoint the location of all proteins and complexes of known structure within three-dimensional reconstructions from whole-cell electron tomography.

Imaging Chemical Events: Gradients, Transformation, and Fate

New imaging methods to understand the series of highly orchestrated chemical events occurring at defined locations within cells for revealing the network of molecular interactions.

- New methods to image the localization of enzyme reactions and the flow of chemicals and macromolecules within and between cells to unravel the complex network of molecular interactions defining cell phenotypes and functions.
- Improved approaches to conduct correlative imaging of intact or living plants.

scales with atomic to near-atomic resolution (Evans and Browning 2013) would extend this approach to spatiotemporal regimes relevant to motions of protein secondary structure elements—domains and subunits central to regulating protein activity. The ability to visualize all conformational states of a given enzyme during the continuum of a reaction would help advance isozyme engineering for more efficient biofuel production.

Light Sources. The commissioning of fourth-generation X-ray free-electron laser (XFEL) light sources such as the Linear Coherent Light Source at DOE's SLAC National Accelerator Laboratory has created the new paradigm of "diffract-before-destroy" imaging (Spence and Hawkes 2008). These femtosecond X-ray sources empower new research using MX and absorption and emission spectroscopy (Kern et al. 2013).

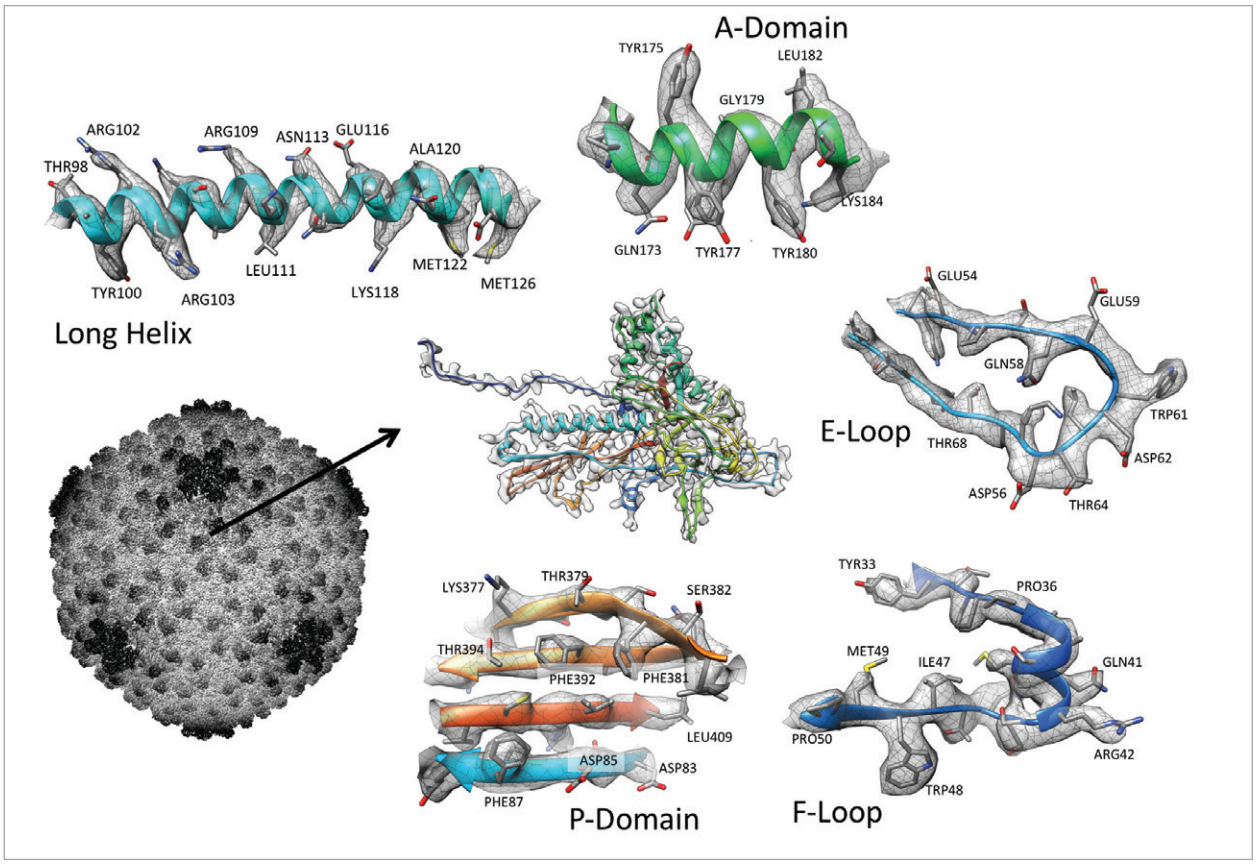


Fig. 7.1. Atomic Model for the P22 Virus Capsid Obtained from Single-Particle Cryo-Electron Microscopy. Key regions in the capsid protein are shown with the corresponding electron microscopy density. [Reprinted from Hryc, C. F., et al. 2017. "Accurate Model Annotation of a Near-Atomic Resolution Cryo-EM Map," *Proceedings of the National Academy of Sciences of the USA* **114**(12), 3103–08.]

XFEL light sources enable analysis of certain samples—for example, 3D nanocrystals (Aquila et al. 2012), 2D crystals (Frank et al. 2014), and large single-particle molecular complexes (Saldin et al. 2011)—not amenable to investigation at synchrotron light sources. Both XFELs and synchrotrons can perform pump-probe dynamic experiments on 3D crystals on ultrafast time scales: up to tens of femtoseconds for XFELs and up to hundreds of picoseconds for synchrotrons (Levantino et al. 2015; Young et al. 2016). Beyond these advances are further needs, including light source improvements in peak brightness coupled with new methodologies to permit atomic-resolution, pump-probe imaging of 2D crystals and, ultimately, single-particle macromolecular complexes. Improvements in current synchrotron sources will also make new measurement methods more widely available, such as fluctuation X-ray scattering to more accurately probe the structures of molecules in solution (Chen et al. 2013; Donatelli et al. 2015), diffractive imaging of single molecules (Donatelli et al. 2017), and ptychographic imaging (Chapman 2010) of microbes and microbial communities (Zhu et al. 2016). These new light source capabilities will greatly benefit efforts to understand the structural dynamics of a wide range of BER-relevant targets in near-native geometries and environments.

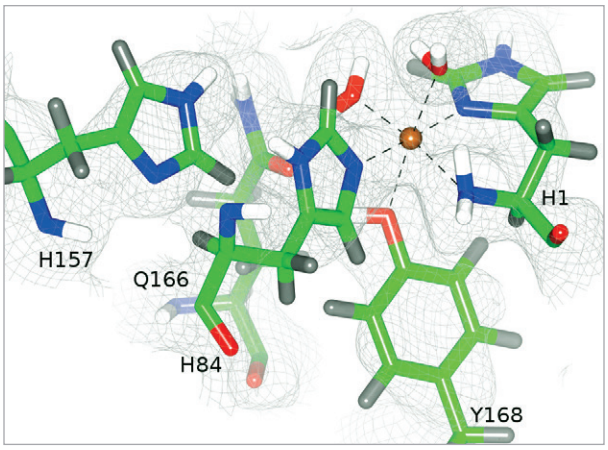


Fig. 7.2. Insight into Oxygen Activation in Lytic Polysaccharide Mono-Oxygenase (LPMO). A structural description of molecular oxygen activation at the Cu active site by a fungal LPMO was obtained using X-ray and neutron crystallography and density functional theory calculations. X-ray crystallography reveals the positions of heavier elements, while neutron crystallography is uniquely capable of visualizing positions of the critical hydrogen atoms. [Reprinted from O'Dell, B., P. K. Agarwal, and F. Meilleur. 2017. "Oxygen Activation at the Active Site of a Fungal Lytic Polysaccharide Monooxygenase," *Angewandte Chemie International Edition* **56**(3), 767–70. © The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.]

Mature Technologies. As these newly developed light source technologies are providing novel insights into more complex systems and at smaller time scales, mature technologies such as MX, NX, SAXS, SANS, and NMR spectroscopy continue to evolve and fulfill new roles in biosystem imaging and analysis. Synchrotron-based MX remains unparalleled in providing high-throughput, high-resolution macromolecular structures, with recent advances extending these results to more challenging and flexible systems (Berman et al. 2000; Moult et al. 2014). The MX workflow is particularly well adapted to conduct expansive surveys and rapidly test predictions. Recently, the methods for sample introduction, data collection, and data processing developed for XFELs (and in some cases cryoEM) are being implemented at the synchrotrons, enabling new science applications. SAXS provides high-throughput, moderate-resolution structural information (Brunette et al. 2015; Bale et al. 2016; Boyken et al. 2016). NMR, SANS, and NX enable altered contrast in heterogeneous materials and provide direct insights into the role of hydrogen atoms (see Fig. 7.2, this page) that can only be inferred by other techniques (Blakeley et al. 2008; Le et al. 2014; Frederick et al. 2015). In all cases, support for these critical capabilities and their continued innovation is needed to enhance the understanding of biosystem complexity and unravel mechanisms central to biological and environmental processes.

Subcellular Localization and Organization with Minimal Specimen Perturbation

Understanding biological system architecture requires the ability to locate specific molecules and complexes in the milieu of a cell, community, biofilm, or tissue with only minimal perturbation to the system. Electron, ion, neutron, optical, and X-ray microscopy approaches can image various scales and sample states. In many cases, a combination of these approaches can provide synergistic imaging modes that enable visualizing a sample across scales.

Most biological systems are damaged when exposed to ionizing radiation and exhibit photobleaching effects under high-light exposure. In addition to physical damage, such exposure can alter the chemical nature of the sample if the generated free electrons or radicals created during beam and sample interactions diffuse freely through the medium and react with the sample. Researchers are developing new approaches to limit such damage during *in situ* analysis, but most experiments currently mitigate beam-induced damage using preanalysis cryopreservation or chemical fixation. Rapid cryopreservation is considered the "gold standard" for retaining the specimen in its native state. This technique immobilizes water in the specimen as amorphous (vitreous) ice rather than in crystalline form that can damage ultrastructure. In comparison, chemical fixation usually alters the true native state

of a specimen's function or ultrastructure, but it does provide a means to identify structures or locate specific molecules in optical, electron, and X-ray images via the use of stains or antibody-conjugated probes. Unfortunately, the process of chemically fixing or cryogenically freezing samples removes the experimenter's ability to track fast dynamic changes that occur on the submillisecond and faster time scales. Probing the dynamics in this temporal range, therefore, requires new methods that enable *in situ* analysis of hydrated and living specimens, even if only under a single-shot, pump-probe regime to avoid damage effects.

Optical Microscopies. These approaches can provide static imaging at high spatial resolution for both fixed and cryogenic samples as well as images depicting dynamics in living cells. Conventional widefield epifluorescence, total internal reflection fluorescence (TIRF), and confocal microscopies have contributed significantly to the understanding of subcellular organization. Researchers have used these techniques to (1) localize fluorescently labeled components inside living cells; (2) identify and quantify the size and shape of organelles and their relative positions in 3D; and (3) discern the relative positions of various cells and cell types within multicellular organisms, plants, or microbial communities. Data can be acquired as a continuous movie at imaging rates of 1 to 1,000 frames per second or via time-lapsed acquisition with a series of time-resolved images collected over many hours. For these conventional optical microscopy approaches, the spatial resolution typically is limited by diffraction to ~200 nanometers (nm) at best. However, the development of "super-resolution" methods has led to a technology-driven resurgence of fluorescence microscopy (see Fig. 7.3, this page). These methods—including photoactivated localization microscopy (PALM; Betzig et al. 2006), stochastic optical reconstruction microscopy (STORM; Rust et al. 2006), and stimulated emission depletion (STED; Klar et al. 1999)—have been bolstered by other techniques, such as light sheet microscopy. These techniques enable optical sectioning through thicker specimens (Zanacchi et al. 2011). Although these instruments and methods have pushed the achievable spatial resolution of fluorescence imaging far beyond what once was

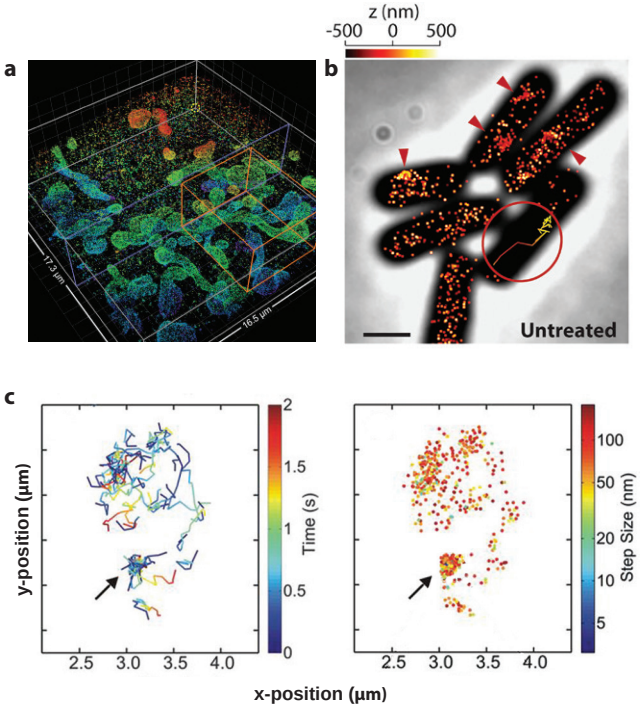


Fig. 7.3. Super-Resolution Microscopy Enabling Nanometer-Scale Imaging, Real-Time Subcellular Particle Tracking, and Three-Dimensional (3D) Microscopy in Cells. (a) A mitochondrial network is visualized by ultrahigh-resolution 3D imaging of whole cells. **(b)** 3D diffusion and dwelling behavior of DNA polymerase is resolved in live *Bacillus subtilis* cells. Scale bar: 1 micrometer (µm). **(c)** Subcellular positions (top) and trajectories (bottom) of outer-membrane enzymes reveal the interactions of proteins with starch during catabolism by living microbial gut symbionts. [**(a)** Reprinted under a Creative Commons Attribution License and modified from Huang, F., et al. 2016. "Ultra-High Resolution 3D Imaging of Whole Cells," *Cell* **166**(4), 1028–40. **(b)** Reprinted from Liao, Y., et al. 2016. "Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells," *Biophysical Journal* **111**(12), 2562–69 with permission from Elsevier. **(c)** Reproduced from Karunatilaka, K. S., et al. 2014. "Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts," *mBio* **5**(6), 1–10 with permission from the American Society for Microbiology under a Creative Commons Attribution NonCommercial ShareAlike license.]

thought possible (i.e., precision down to 20 nm in some instances), super-resolution fluorescence microscopy is limited to imaging fluorescently labeled molecules and not the surrounding unlabeled cell environment. As is the case with most cutting-edge instrumentation, trade-offs exist, and the increase in spatial resolution beyond the diffraction limit for super-resolution optical microscopy tends to result in lower temporal resolution.

Alternate, Combined Approaches. In contrast, combining optical microscopy with vibrational spectroscopy methods such as infrared microspectroscopy and, more recently, nanospectroscopy enables the imaging of subcellular components and chemical distributions in a label-free manner (Centrone 2015; Muller et al. 2015). Coherent anti-Stokes Raman scattering (CARS; Evans and Xie 2008) and stimulated Raman scattering (Freudiger et al. 2008) microspectroscopies are nonlinear optical approaches that can achieve the same objective. These alternate approaches can image subcellular organization with chemical information at a molecular level. Because the resulting infrared or coherent Raman signals depend on interactions between light and the resonant modes of the functional groups of target molecules, detection could be limited to certain classes of molecular components. For a complete picture of cellular structure and dynamics, fluorescence and CARS imaging data often are complemented with structural data from electron or X-ray microscopy using correlative microscopy workflows.

Tomography and Scanning TEM for 3D Reconstructions. Transmission electron microscopy (TEM) combined with tomography produces 2D images of specimens at very high spatial resolution (nanometers to Angstroms) that can be back-projected to reconstruct 3D ultrastructure and cellular organization (see Fig. 7.4, p. 66). However, the high spatial resolution provided by TEM comes with a trade-off related to tolerable specimen thickness. In general, the thicker the specimen, the lower the achievable resolution. For conventional microscopes, biological specimens ideally should be less than 200 nm thick. Current knowledge of organelle structures is almost entirely a result of decades' worth of ultramicrotome sample sectioning in combination with room-temperature or cryoEM. Though such reconstructions have been immensely informative, they are fraught with unavoidable image artifacts.

New advances in phase plate technology (see box, Electron Microscopy Terms, p. 60; Danev and Nagayama 2001; Danev et al. 2014) and scanning TEM (Wolf et al. 2014) have allowed tomographic reconstructions of intact bacterial and algal cells more than 1 μm thick (Henderson et al. 2007; Dai et al. 2013; Guerrero-Ferreira and Wright 2014) with a spatial resolution of 2 to 10 nm without cell sectioning or staining. Yet visualizing organisms larger than 1 μm in 3D using EM involves a technically difficult, time-consuming workflow—(1) cutting consecutive physical sections of a cell thin enough for electron transparency, (2) imaging and reconstructing the sections separately, and (3) stitching them together computationally. Such slice-by-slice imaging methods include serial section TEM, slice-and-view scanning EM (Heymann et al. 2006), bi-directional ion milling to thin a specimen already sitting on a TEM grid (Marko et al. 2007; Chaikeeratisak et al. 2017), or using a focused ion beam (FIB) to selectively lift out a specific region from a specimen and attach it to a TEM grid (Mahamid et al. 2015). The medical research community is using these methods to image large, montaged volumes of brain tissue for connectomics (the creation of maps showing the connections between neurons), an approach which holds great promise for analyzing small ecosystem communities such as biofilms or samples isolated from the rhizosphere. Similarly, new advancements in visual proteomics (Dai et al. 2013; Mahamid et al. 2016) are needed to pinpoint the location of all proteins and complexes of known structure within 3D reconstructions from whole-cell electron tomography. These highly detailed maps would refine insights into how cellular and subcellular organization varies

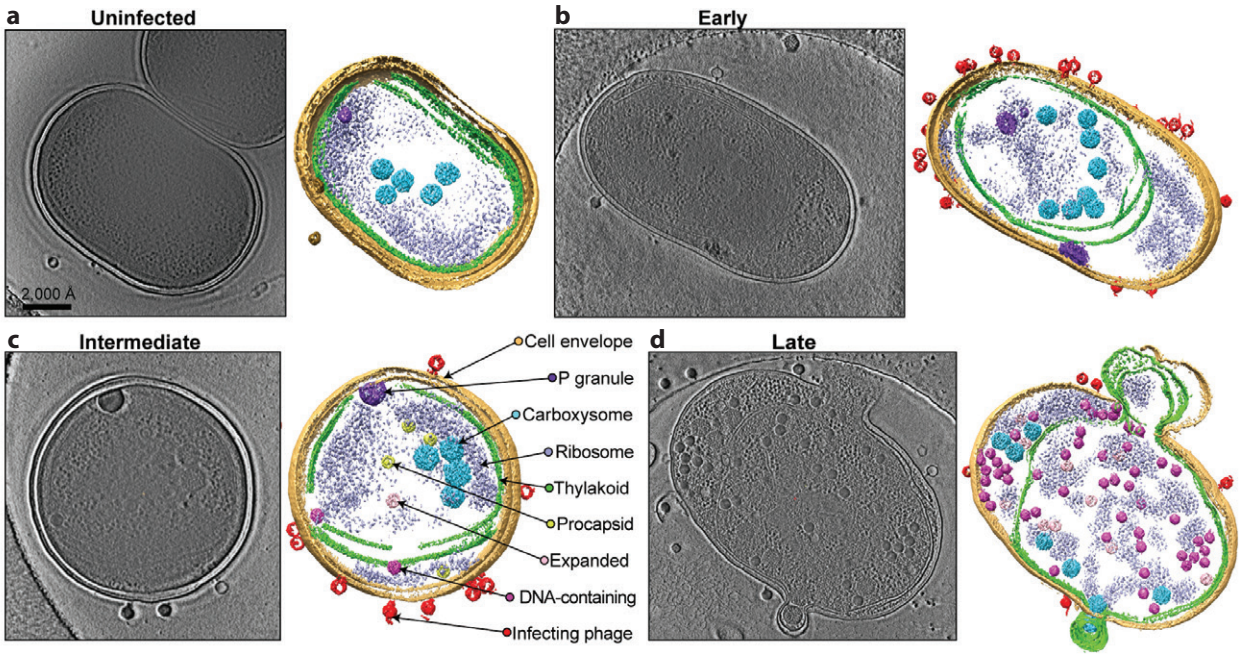


Fig. 7.4. Zernike Phase Contrast Cryo-Electron Tomography of Syn5-Infected Marine Cyanobacteria at Various Stages of Infection. Views (**a–d: left**, tomogram slice view; **right**, annotated tomographic slice view) of 54 Angstrom sections of Cyanobacteria, *Synechococcus* sp. Strain WH8109, cells during the process of infection by cyanophage Syn5. Sections are computationally derived from the middle of tomograms of the whole cells. **(a)** Uninfected cell, **(b–d)** early, intermediate, and late stages of infection. Cellular components and phages are colored and labeled in the annotated view in **c**. Phage progeny can be separated into three types based on size, shape, and internal density: procapsid, yellow; expanded capsid, pink; and DNA-containing capsid, magenta. [Reprinted by permission from Macmillan Publishers Ltd: Nature. Dai, W. et al. 2013. "Visualizing Virus Assembly Intermediates Inside Marine Cyanobacteria," *Nature* **502**(7473), 707–10. © 2013.]

between phenotypes or in response to environmental changes (see Fig. 7.4, this page).

Transmission X-Ray Microscopes (TXMs). A complementary or alternative technology to electron microscopy, TXMs can also image the internal structure of cells (Leis et al. 2009; Larabell and Nugent 2010). TXMs use soft X-rays to image samples at inherently lower spatial resolution than electron microscopy (10 to 50 nm compared to 1 to 5 nm, respectively). However, soft X-rays can image cryopreserved specimens up to 15 μm thick, providing higher sampling rates and sample numbers that can be statistically analyzed for most unicellular and some multicellular organisms without requiring physical sectioning (Le Gros et al. 2012; McDermott et al. 2012). Similar to neutron imaging, hard X-ray TXMs can probe even thicker samples (greater than centimeters thick) to gain insights into plant biology, microbe-soil interactions, or microbial communities and biofilms. In addition to imaging the spatial arrangement in whole cells, TXMs, in combination with linear absorption coefficients, enable direct identification of subcellular structures (see Fig. 7.5, p. 67; Smith et al. 2014). This capability is possible because image contrast in TXMs comes from the attenuation of the incident light by biomolecules in the specimen, obviating the need for contrast agents or stains (Do et al. 2015). Water is relatively transparent (compared to the absorption

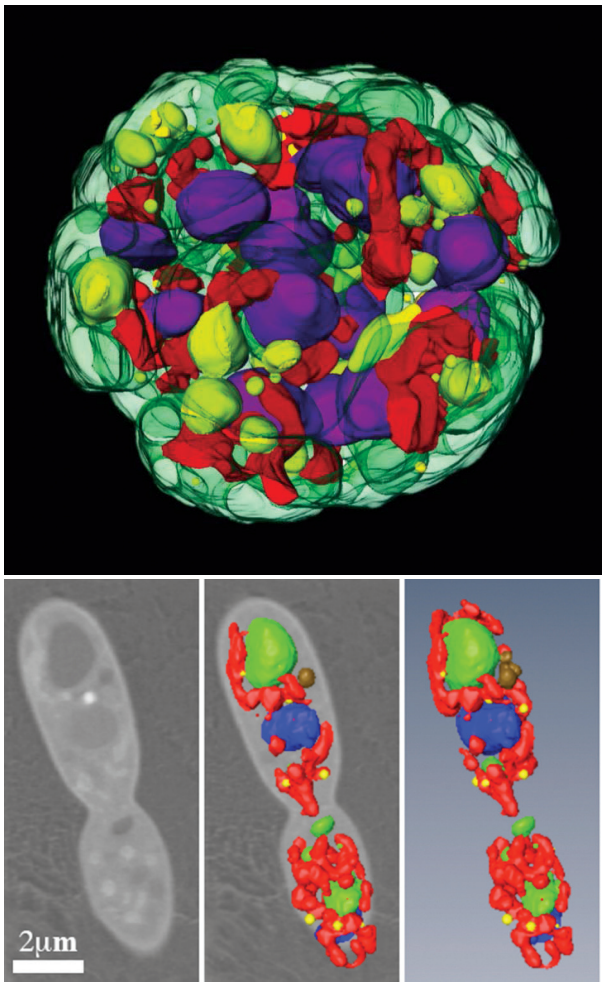


Fig. 7.5. Cryogenic Soft X-Ray Nanotomography of Cellular Systems. (Top) Cryo-soft X-ray tomography of a reconstructed *Chromochloris zofingiensis* cell with segmented nucleus (purple), chloroplast (green), mitochondria (red), lipids (yellow), and starch granules within the chloroplast (blue). **(Bottom)** Frozen-hydrated whole yeast cells imaged at the Advanced Light Source. Organelles within the cell can be directly modeled and assigned based on their linear absorption coefficients. The central slice of the three-dimensional tomogram is shown on the left and final model on the right. Vacuoles, nucleus, mitochondria, and lipid droplets are colored green, blue, red, and yellow, respectively. [Top image courtesy National Center for X-ray Tomography. Reprinted from Roth, M. S. et al. 2017. "Chromosome-Level Genome Assembly and Transcriptome of the Green Alga *Chromochloris zofingiensis* Illuminates Astaxanthin Production," *Proceedings of the National Academy of Sciences of the USA* **114**(21), E4296–4305. Bottom image courtesy James Evans, Chuck Smallwood, and Erin Bredeweg, Pacific Northwest National Laboratory.]

by carbon-containing molecules) within the so-called soft X-ray "water-window" (Parkinson et al. 2013). Consequently, soft X-ray microscopes provide an accurate measurement of the cell's carbon content, an important factor in calculating a biosystem's energy balance at the cellular level.

Imaging Chemical Events: Spatial and Multiscale Mapping

Biological systems rely on chemistry taking place at the correct time and in the right place (Le Gros et al. 2012). Tracking the location and concentration of molecules over time is thus a key characterization step for biological investigations. Determining location and concentration is relatively easy if the molecule of interest is produced at high levels or distributed over large areas, but accurately characterizing small numbers of molecules produced by native or engineered biosystems remains challenging, particularly for single-cell measurements. NMR, Fourier transform infrared (FTIR) spectroscopy, and mass spectrometry (MS) are approaches well suited to identifying chemical composition. MS imaging, such as with Nano-Secondary Ion Mass Spectrometry (NanoSIMS), allows part-per-million detection of ions and isotopes with up to ~50 nm spatial resolution, but it is a static and destructive method best suited to imaging chemical gradients and cellular assimilation. In comparison, NMR approaches have a lower spatial resolution but can image living cells, providing information on native states of their lipids, nucleic acids, polysaccharides, and proteins. Researchers have recently applied neutron scattering techniques in combination with isotopic labeling to investigate lipid domains in membranes and protein dynamics *in vivo*, opening up new and exciting avenues for live-cell imaging (Anunciado et al. 2017). Recently, nano-FTIR spectromicroscopy has provided chemical information on organic samples with a spatial resolution of 30 nm (Amenabar et al. 2017), although the spectral range is significantly limited by the use of laser light sources. Further development of each technique (i.e., imaging MS, NMR, and FTIR) is needed to

simultaneously improve chemical sensitivity and spatial resolution. These advances would help reveal new insights into biological compartmentalization and its role in regulating molecular interactions.

Synchrotron-based X-ray spectroscopy or microspectroscopy can generate elemental maps of specimens, but the inherent damage from beams can limit these approaches to static observations from nearly equivalent samples. In contrast, synchrotron-based broadband FTIR (sFTIR) spectroscopy or spectromicroscopy does not introduce detectable effects on biological materials (Holman et al. 2002a), thereby permitting analysis of live cells and observation of chemical events in the same cell or sample. sFTIR spectromicroscopy (Holman et al. 2010; Probst et al. 2013) and time-resolved microfluidic sFTIR spectromicroscopy (Holman et al. 2009b; Holman et al. 2010; Loutherback et al. 2016) are now matured techniques for measuring chemical changes in live microbial specimens with diffraction-limited spatial resolution of 2 to 10 μm and temporal resolution of several to tens of seconds.

Use of midinfrared wavelengths for sFTIR inherently precludes their use in obtaining nanoscale chemical information. Synchrotron infrared nanospectroscopy (SINS; Bechtel et al. 2014) overcomes this spatial limitation by combining high spatial resolution atomic force microscopy with broadband (4,000 to 650 per centimeter) synchrotron infrared radiation, enabling the measurement of chemical changes down to 20 nm or finer spatial resolution. SINS, however, is limited to measuring chemical events in water-free environments because water induces viscous damping that severely decreases the interpretability of the data. Thus, the next key technological challenge to overcome is the ability to probe live chemical events in living cells at the nanoscale. Worth noting is that all FTIR approaches are limited to detecting infrared-active molecules. Therefore, integration of sFTIR imaging with spatially resolved ambient MS (O'Brien et al. 2015) is a promising approach to improving the chemical specificity for unique compound identification.

Correlative Imaging. Because localizing chemical information is fundamental to understanding biological systems, the ability to overlay spectroscopy and structural data from the same specimen onto information from other imaging approaches creates a highly informative composite view of the specimen. The field of correlative imaging has expanded rapidly in recent years, leading to many novel insights not possible with a single technique (see Fig. 7.6, p. 69). Of particular note is the considerable correlative power gained by combining fluorescence and electron microscopy to visualize the localization of labeled protein complexes within a whole-cell context using cryoET (Briegel et al. 2008; Hampton et al. 2017) or cryo-soft TXM. Challenges remain regarding common sample geometries, amenable sample thickness, image indexing and registration, and computational visualization of data at different spatial resolutions. Continued advancements in these areas will help make such holistic structural and chemical imaging approaches more widely adopted and useful for BER applications, including understanding (1) the impact of biofuel production on microbes, (2) how microbes communicate to form biofilms, and (3) the cellular components responsible for biogeochemical cycling. Correlative imaging of plants and plant-microbe interactions is another area ripe for development, but its large spatial range—covering macromolecules (~10 nm), single microbes (~2 μm), single plant cells (~10 μm), and a whole plant (millimeters to meters)—complicates workflows for both data acquisition and downstream analysis and visualization. Although efforts to image root systems have advanced significantly in recent years, correlative imaging of intact or living plants remains challenging because of the inherent opaqueness of soils.

Root system architecture has a dramatic bearing on plant viability and crop productivity in given soil conditions. New developments in dynamic, static, or cryogenic multiscale imaging using electron,

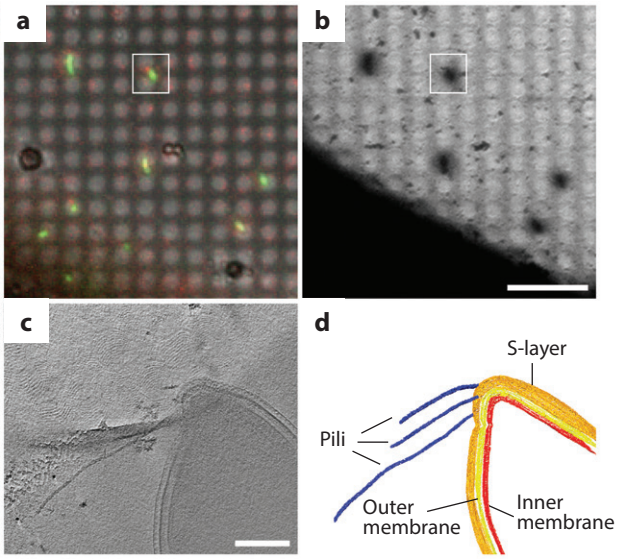


Fig 7.6. Correlative Imaging for Structure-Function Studies. Cryo-correlative light and electron microscopy (cryoCLEM) combines spatiotemporal information from fluorescence light microscopy (fLM) with high-resolution structural data from cryo-electron tomography (cryoET). CryoCLEM enabled the identification of fluorescently labeled piliated *Caulobacter crescentus* swarmer cells dispersed on an electron microscopy grid. The combined approach facilitated structure-function investigations of the pilus filament and pilus assembly complex during secretion and retraction stages. **(a)** CryofLM of frozen-hydrated *Caulobacter* cells in which the pilus filament was fluorescently labeled (red). **(b)** Cryo-electron microscopy (cryoEM) montage magnified 100x. **(c)** Tomographic slice through the *Caulobacter* cell with the labeled pili. **(d)** Segmented *Caulobacter* cell volume. Pili (blue), s-layer (gold), outer membrane (yellow), and inner membrane (red). White box in **(a)** corresponds to white box in **(b)** and tomogram of *Caulobacter* cell **(c)**. Scale bars: **(b)** 10 mm, and **(c)** 100 nm. [Image courtesy Rebecca S. Dillard, Emory University; Cheri M. Hampton, Emory University; Courtney K. Ellison, Indiana University; Yves V. Brun, Indiana University; and Elizabeth R. Wright, Emory University.]

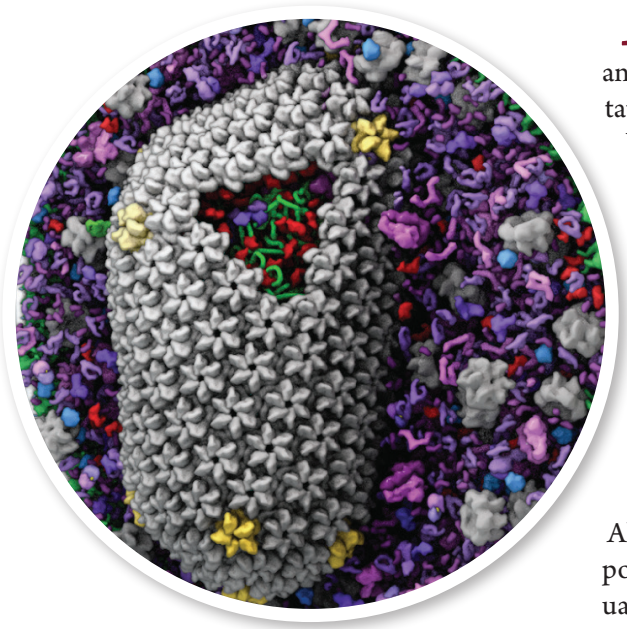
neutron, optical, infrared, or X-ray approaches are needed to reveal the unperturbed, overall 3D architecture and chemistry of a root system and specific cell surface–mediated interactions between root cells and microbes. Given the potentially significant increase in crop yields that could result from efforts to match plant species to optimal soil type and texture, increase their drought tolerance, or leverage interactions with beneficial microbes, developing new technologies for imaging plants from roots to leaves could have dramatic environmental, economic, and humanitarian effects.

7.2 Understanding Genotype-Phenotype Interactions to Improve Biofuel Feedstocks

Comprehensive biosystem characterization requires understanding how the system senses environmental perturbations and the mechanisms that cells use to make changes in response to these cues. Developing new bioimaging and spectroscopy capabilities spanning electron, ion, optical, neutron, and X-ray modalities and fusing them with new advances in systems biology, computation, and automation can provide a deeper understanding of the link between genome and phenotype. Exploiting and controlling those linkages would radically enhance the ability to design and harness biosystems for cheaper, more efficient biofuels and bioproducts, increased crop yields, and biological routes to soil remediation.

Chapter 8

Data Integration and Analysis



The streams of data produced by the techniques described in the previous chapters can be difficult to interpret unless the data are combined and synthesized into visual libraries or, ideally, computational models. Better methods are needed for assembling realistic models of complex biological systems. These methods must include the ability to connect to and seamlessly integrate multiple sources of experimental data, from genomics and proteomics to structural data at different scales (atomic to cellular, and beyond).

8.1 Current Science and Technologies

Integrating Analyses Across Scales

Although integrative analysis across scales can be very powerful (see Fig. 8.1, p. 72), it is often performed manually. Showing great promise is the recent development of methods that integrate information from multiple experimental sources to create models or ensembles of models, consistent with the data (Russel et al. 2012). Still, these methods fall short of generating models that cross multiple length scales, and they do not yet address the time domain. Fortunately, researchers are discussing data formats and relationships that would support integrative models emerging from multiple experimental approaches (Sali et al. 2015) that also include dynamics. These new classes of models potentially can provide rich information across the spatial and temporal scales that were the subject of the workshop.

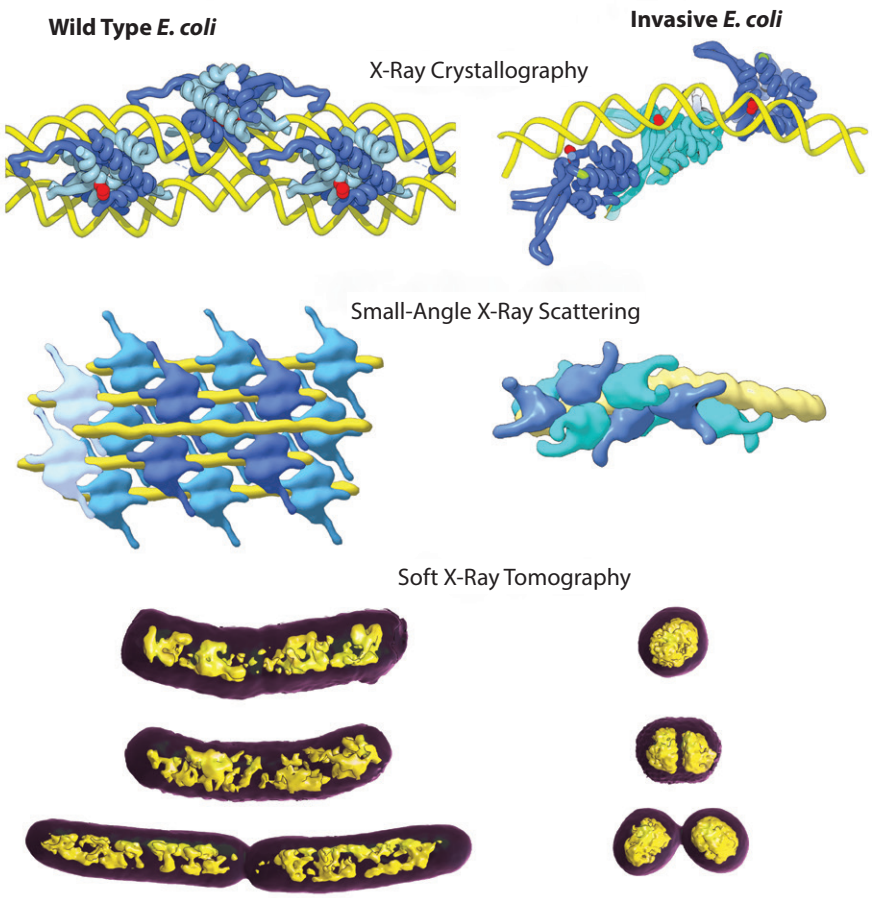


Fig. 8.1. Examining the Impacts of DNA Compaction in *Escherichia coli* Using Multiple Techniques: An Example Use Case. Applying multiple experimental techniques enables elucidation of the molecular mechanisms underlying the role of DNA compaction in gene expression in *E. coli*. The histone (HU) proteins were imaged at different resolutions and stages using X-ray crystallography and small-angle X-ray scattering (SAXS) capabilities. The crystallography provides atomic-level details of how the HU proteins interact with the bacterial DNA, while SAXS shows how the HU proteins assemble and affect the longer DNA strands in a solution. X-ray tomography reveals the natural contrast in organic material in as close to a living state as possible and provides quantitative comparisons of how compacted the chromosomes are in pathogenic and normal *E. coli* strains. [Image courtesy Michal Hammel, Lawrence Berkeley National Laboratory.]

The creation of integrative models can reach such a scale and scope that the data they generate rapidly become a peta- to exascale computing challenge, especially when coupled with data from the simulation of dynamics. Data handling, visualization, and computation, including integrative modeling, will benefit from access to state-of-the-art supercomputing facilities typically housed in national laboratories. Investigators supported by the Biological Systems Science Division within the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) are studying a large landscape of temporal dynamics ranging from picoseconds to microseconds (for probing enzyme activity) to days or even weeks (for understanding subcellular reorganization). Advances in both hardware and software will be required to empower the modeling of whole biosystems with fine-grained resolution. For example, these models would simulate the motion of individual molecules and macromolecules over milliseconds to hours. Such simulations could be used to better understand the flow of metabolites between cells, the active transport of small molecules across cell membranes, dynamics of the molecular machines responsible for cycling

Data Integration and Analysis Needs

Conducting Multiple, Systematic Tests with Control over Variables

Systematic testing to gather statistics, multiple angles of attack to confirm direct linkages, and the ability to control variables for assembling data needed for realistic biological models.

- Biological experiments to test hypotheses of biosystem function.
- Integration and automation methods for sampling, acquisition, processing, and modeling bioimaging data to efficiently optimize phenotypes.
- Data and management methods to realistically model complex biological systems.
- Improved tools to automate visualization tools.

New Data Analysis Approaches

Mathematical and computational approaches for analyzing large datasets to provide models of complex systems with fundamental details at atomic and molecular scales.

- New mathematical approaches to employ the publicly accessible resources of large, raw datasets.
- Integration of molecular with genomic data to elucidate sequence-structure-function relationships.
- Technologies to make use of emerging insights, such as the productive use of complete X-ray diffraction and diffuse scattering data.
- New algorithms and standards to automate extraction, classification, and visualization of features available in large tomographic datasets.

of elements, and other fundamental processes that cannot be measured directly by experiment.

Big Data and High-Performance Computing

Many of the imaging techniques described in this report are entering the era of big data, where the sheer volume of acquired data and variables requiring analysis have become intractable for rapid processing with personal computers. Consequently, the research community's continued access to high-performance computing (HPC) is essential, as is the development of new informatics resources that can support automated integration of the hundreds of disparate data types generated by new technologies. These technologies include graph database representations of scientific concepts guided by researcher expertise and community models, along with real-time machine learning and inference. Also needed to support scientific discovery are new advances in automated processing, registering, overlaying, visualizing, and interpreting multimodal or multivariable data.

Currently, access to instrumentation, experts, or computational resources is rapidly becoming the limiting factor for many bioimaging experiments. As these capabilities become widely available or more efficient for fast data acquisition, the net result will be a continuation of near-exponential growth in data flow. Concomitant needs arising are improved data handling, archiving, and compressing. Ultimately, the real strength of improved HPC for data analysis and modeling will come from developing an iterative workflow in which experiment refines theory and theory drives next-level experimentation to enhance the efficiency of biosystem optimization. The following sections summarize the computational challenges currently faced in multiscale simulations of large biological systems and in data

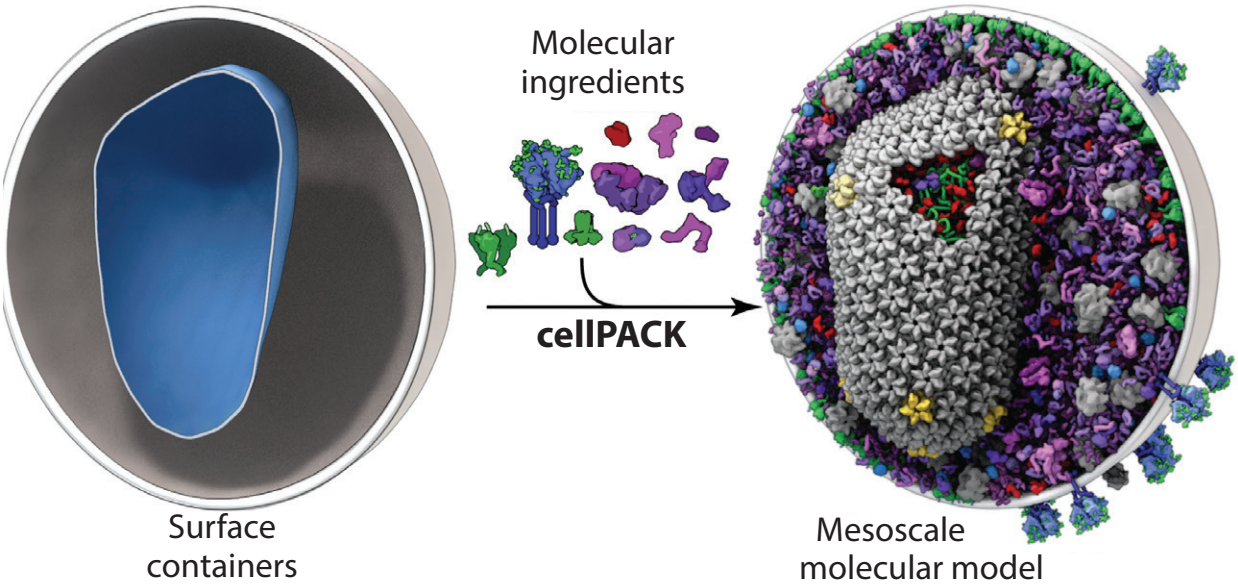


Fig. 8.2. Data-Centric, Data-Driven Modeling Environment. CellPACK enables assembly of models that are consistent with many different types of datasets, such as proteomics and genomics, fluorescence microscopy, electron microscopy (EM), and X-ray crystallography. The total system volume can be divided into different subvolumes using surface definitions from, for example, EM data. Formulas for different system subcompartments can be parameterized and used to generate model ensembles consistent with observed data. [Reprinted from Macmillan Publishers Ltd: Nature Methods. Johnson, G. T., et al. 2015. "cellPACK: A Virtual Mesoscope to Model and Visualize Structural Systems Biology," *Nature Methods* **12**(1), 85–95. © 2015.]

analysis via atomic-resolution crystallography and microscopy methods.

8.2 Major Challenges in Data and Computation

Clearly, atomic-scale modeling is approaching the mesoscale, and within 5 years the first atomic-scale cell simulation likely will become reality. New tools in this arena, such as cellPACK (see Fig. 8.2, this page; Johnson et al. 2015) and LipidWrapper (Durrant and Amaro 2014) automate and enable the construction of molecular models of subcellular environments. However, with the petascale computing available on DOE machines during preparation of this report, simulations of such models [e.g., roughly 180 million atoms, using classical, Newtonian, and molecular dynamics (MD) simulations] are able to achieve only dynamic time scales on the order of hundreds of nanoseconds (Zhao et al. 2013; Yu et al. 2016).

Anticipating Exascale Simulation of Cellular Systems

A major need is the development of exascale computing architectures that allow for massively parallelizable MD codes. Typically, these codes spatially decompose the biosystems in the parallelization scheme and therefore scale well with processor count, but only if the computing platforms have fast interprocessor communication. With exascale computational architectures, the sampling of longer and more relevant biological time scales (milliseconds) for these large-scale systems will shed light on myriad biological processes in realistic *in silico* environments. HPC systems on the order of billions of atoms (micrometers in dimension; e.g., a full-sized cell) will be able to scale to millions of cores.

Exascale machines should be equipped with large memory nodes to facilitate data processing and analysis directly at the computing site, obviating the need for moving petabyte-scale datasets between sites. Also needed are powerful visualization capabilities that allow researchers to interact with their data remotely, both during and after a simulation, all without having to transfer the data to their own institutions.

Using Enhanced Sampling and Multiscale Approaches

Along with exascale simulations of image systems, researchers can extract great value from long time scale simulation of smaller molecular and macromolecular components with classical MD and with so-called enhanced sampling approaches (Bernardi et al. 2015). Further development of new theoretical or statistical frameworks for simulation analysis presents a different set of challenges and requirements. For example, the use of Markov state models (Malmstrom et al. 2015; Plattner and Noe 2015; Meng et al. 2016) or milestoning (Faradjian and Elber 2004) now enable the extraction of long time scale dynamics (milliseconds to seconds) from many short time scale simulations for single-molecule biosystems. Such methods benefit from execution on independent graphical processing units (GPUs) where interprocessor communication is less of a bottleneck compared to the previously described large spatial systems. Multiscale methods embody another set of approaches that can seamlessly combine different levels of theory or resolution, for example, quantum mechanics and molecular mechanics (Chudyk et al. 2014) or MD and Brownian dynamics (Votapka and Amaro 2015), to increase accuracy or access longer time scales and larger length scales. Another multiscale method class is represented by the hierarchical integration of sets of approaches carried out at different scales, where the development of one cohesive model is obtained through the interchange of key parameters across model scales (Boras et al. 2015; Mih et al. 2016; Yu and Bagheri 2016). Capabilities on exascale platforms will bring seconds-long simulations of macromolecules into routine practice, enabling connections to experiments at orders-of-magnitude longer time scales than is possible today (i.e., milliseconds, routinely).

Atomic-Scale Spatial and Temporal Data

Crystallography, nuclear magnetic resonance (NMR), and, more recently, single-particle cryoEM provide some of the highest-resolution information at atomic and molecular scales. This information usually is interpreted with individual structural models but is increasingly useful for supplying the fundamental atomic detail for multiscale, integrative models. While the isolated structures are highly useful, such as in enzyme engineering (Campbell et al. 2016), high-resolution structural approaches potentially could be used to help "divide and conquer" larger, more complicated systems (Chiu et al. 2006; Ward et al. 2013). Additionally, new light sources, such as the X-ray free-electron laser (XFEL) at DOE's Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory, are enabling resolution of atomic details as a function of time (Levantino et al. 2015). The internal motions of macromolecules are key to their function, so researchers could use these new capabilities to visualize how natural and engineered enzyme variants differ, potentially improving the further rational design or interpretation of mutations (Bhabha et al. 2015). In addition, the acquisition of time-resolved data at near-atomic resolution is valuable for improving simulation force fields, which, in turn, will enable the development of more predictive and dynamic models of biosystems (van den Bedem and Fraser 2015; Kupitz et al. 2017). As atomic-scale information becomes richer through pushing the boundaries of spatial and temporal resolution, several computational challenges must be addressed.

The LCLS pulse of 120 hertz frequency enables image collection at that rate (White et al. 2015). Larger detectors will be needed to capture higher-resolution details, with a large dynamic range to properly measure low-resolution data.

Serial Crystallography

This important new crystallography approach was developed at X-ray free-electron lasers (XFELs), which exploit the diffract-before-destroy phenomenon made possible by very powerful, short (femtosecond) X-ray pulses. More recently, the method has been translated to high-brightness synchrotron X-ray sources.

Typical data collection runs currently generate many terabytes (TB) of data. Future pulse rates can increase this data rate by more than 2 orders of magnitude. At this scale, data storing, transferring, and processing represent a significant challenge. With the use of imperfect sample introduction tools resulting in some collected images that are seemingly devoid of data, current emphasis is on sorting blank images from those that contain "meaningful information." This has driven development of new algorithms and software (Ginn et al. 2016), which will need further refinement. Additionally, sample introduction tools are improving, so hit rates for serial crystallography experiments (see box, Serial Crystallography, this page) are increasing (Fromme et al. 2015; Oghbaey et al. 2016; Fuller et al. 2017). If all images with any non-background intensity should be saved, then new raw data repositories likely will be needed.

The current best practices require deposition only of data that are used to calculate structures. Repositor examples include the SBGrid's Structural Biology Data Grid (Meyer et al. 2016), Coherent X-ray Imaging Data Bank (Maia and Hajdu 2016), and others (Kroon-Batenburg et al. 2017). However, this practice may limit downstream research to mine data not included in the original analysis. Looking to the future, the ability to store and provide global access to original datasets could enable DOE researchers to return to these datasets and extract additional value, unseen at the time of acquisition.

The most exciting experiments will rely on real-time data processing to optimize serial crystallography experiments on the fly (Lyubimov et al. 2016), requiring new capabilities and algorithms. Moreover, several of the most exciting directions in X-ray data processing involve interpreting the weakest signals, that is, diffuse (Wall et al. 2014) or continuous (Ayyer et al. 2016) intensities, which are much harder to distinguish than sharply peaked Bragg intensities (Van Benschoten et al. 2016); see Fig. 8.3, p. 77). The ability to perform this interpretation efficiently as data are collected has great ramifications, as the volume of data needing transfer and maintenance potentially can be reduced massively on site.

Major challenges in data representation involve accurate representation of multiple conformations (Woldeyes et al. 2014), data uncertainty (Terwilliger et al. 2007; Lang, P. T., et al. 2014), and time-dependent evolution of the conformational ensemble (Aranda et al. 2006). Particularly appealing is the potential for additional data (1) in time-resolved experiments (Schmidt et al. 2013), (2) in multiple temperatures (Keedy et al. 2015b), or (3) from diffuse scattering (Van Benschoten et al. 2016) to constrain the modeling of multiple conformations (Burnley et al. 2012; Keedy et al. 2015a).

Determining Biological Processes with Microscopy and Tomography

Single-particle cryo-electron microscopy (cryoEM) has advanced to the point where full atomic models can be obtained for protein components in viruses, membrane proteins, filaments, and molecular machines (Liao et al. 2013; Zhang et al. 2013; Amunts et al. 2014; Bartesaghi et al. 2014; Galkin et al. 2014; Lu et al. 2014; Wang et al. 2014). Additionally, cryo-electron tomography (cryoET) of frozen, hydrated cells and organelles (Li and Jensen 2009; Davies et al. 2012) has enabled subnanometer-resolution structural snapshots of molecular machines in different functional states inside the cell under true *in situ* conditions (Mahamid et al. 2016). These recent advances in

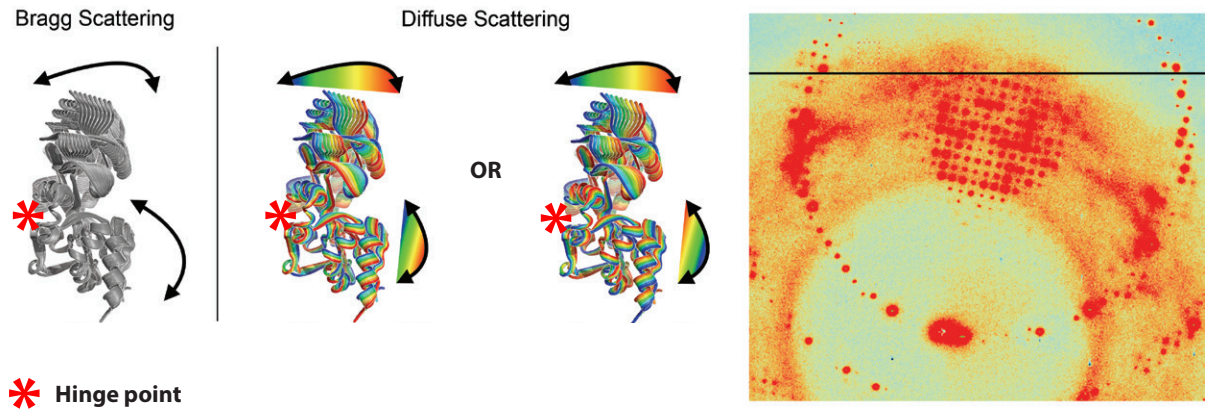


Fig. 8.3. Diffuse Scattering and Bragg Scattering. (Left) A protein molecule is shown in ribbon format, and contains two domains connected by a hinge (marked with an asterisk). The intensities of crystallographic Bragg scattering provide information about the mean squared displacements of atoms, which can be interpreted as the motion of these domains (arrows indicate the extent of the motion for each domain). However, the correlations between these motions are lost in the Bragg scattering. **(Middle)** Fortunately, the diffuse scattering between the Bragg peaks contains information about the correlations. The rainbow colors indicate the sequence of conformations traced as the domains move. In the left model, when one domain moves down so does the other; whereas in the right model, the domains move toward and then away from each other in a closing and opening motion. The functional implications of these two modes of motion may be profound. **(Right)** The diffuse scattering between and around the sharp Bragg peaks (deep orange) is shown in a representative diffraction image. [Image courtesy James Fraser, University of California San Francisco.]

both cryoEM and cryoET can provide answers to mechanistic questions related to the structural dynamics of macromolecules key to many biological processes in gene duplication, translation, protein trafficking, protein homeostasis, metabolite transport, signal transduction, and responses to environmental stresses.

Improved Image Processing and CryoEM. To obtain a complete description of these biological processes, determining multiple structures at different steps of the processes at different times will be necessary. The resulting high-quality data explosion resulting from cryoEM experiments (as well as serial crystallography experiments at XFELs and synchrotrons) represents an immediate challenge for data management, processing, and visualization. For example, with the current generation of detectors and reasonable measures for compression, a typical cryoEM exposure for a single-particle grid area occupies 2 to 4 gigabytes of disk space. At 1,000 exposures per day, that total amounts to roughly 4 to 5 TB of collected raw data per day. Inevitably, with single-particle data having multiple conformations, use of sophisticated algorithms to sort out heterogeneous data will become the norm rather than the exception. Such data processing is CPU intensive and requires more efficient computational algorithms to reduce data processing time and to include structure validation and error measurements. Such sophisticated and improved image processing has not yet been developed. Final structures, which generally are represented by atomic models, must be validated and annotated, with their positional accuracy quantified. A consensus of metrics to report cryoEM structure has not yet emerged from the cryoEM community.

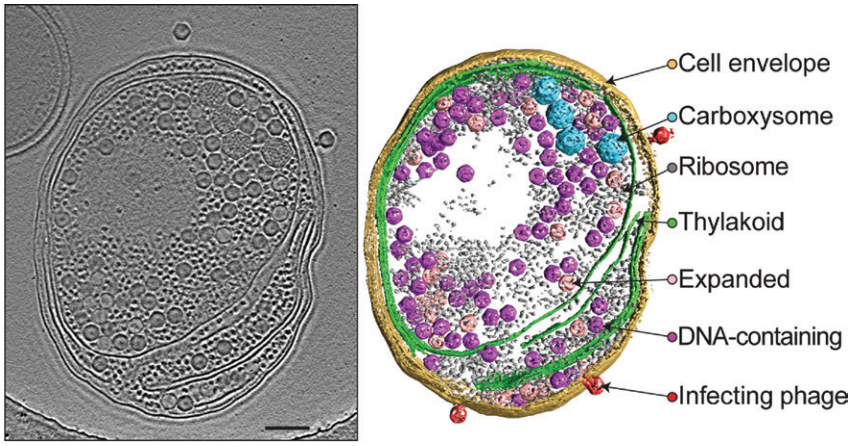


Fig. 8.4. Annotation (currently performed manually) of Cryo-Electron Tomography of a Cyanobacterium Infected with a Syn5 Cyanophage. (a) Slice of a tomogram. **(b)** Annotated subcellular components in the three-dimensional tomogram. [Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols. Dai, W., et al. 2017. "Zernike Phase-Contrast Electron Cryotomography Applied to Marine Cyanobacteria Infected with Cyanophages," *Nature Protocols* **9**(11), 2630–42. © 2014.]

CryoET and Improved Visualization. For cell tomograms obtained by cryoET, the data interpretation challenge is to annotate and segment the three-dimensional (3D) tomograms of crowded subcellular features (see Fig. 8.4, this page). The incomplete data resulting from limited specimen tilt angles present challenges in confidently recognizing subcellular components. Automated algorithms must be developed to (1) identify features of interest to avoid subjectivity and increase throughput; (2) extract subvolumes for classification, alignment, image correction, and averaging; and (3) assess the reliability and resolution of the resulting structures. In addition to zooming in on specific subcellular features, cryoET provides an opportunity to generate new hypotheses regarding structure and function relationships of subcellular organization, as suggested by direct visualization. Similar challenges are confronting the study of plant tissues or biofilm imaging by Fourier transform infrared microtomography (FTIR μTomography; Martin et al. 2013). Therefore, developing user-friendly visualization tools will be key to enabling scrutinization of the complex and crowded density maps enriched with various features. Feature discovery and conformational dynamics in the cellular context will require both statistical and visualization tools to reveal them. Furthermore, disseminating such multidimensional data to other scientists is currently challenging; therefore, new standards for data representation and protocols for data transfer are needed.

Other Structural Bioinformatics. In addition to the structure data generated by cryoEM and cryoET, or chemical data generated by FTIR μTomography, a timely consideration is the correlation of the spatial and temporal information with other structural bioinformatics such as crystal structures and fluorescence imaging, simulation dynamics, and proteomic and genomic data. Integrated structural information coupled with genomics certainly will be a driving force for biology in the coming decades.

8.3 Advancing Computational Platforms for Large-Scale Data Processing and Analysis

Continued development of imaging and microscopy tools that enable general, highly detailed 3D images of cells via serial block face (i.e., volume) imaging (Denk and Horstmann 2004) currently are and will continue to present data and computing challenges at scale. In addition, developments in biophotonics and image capture will enable, in time, the extension of these volume images as the fourth dimension. Large-scale image processing capabilities (e.g., to visualize microbial communities

and their interactions and growth over time) will be needed to handle volumes of single-image data on the order of 10 TB and greater. These capabilities will be necessary to train machine-learning algorithms to select key biological features from the images in an automated fashion, thus accurately and rapidly identifying and segmenting the molecular, macromolecular, and subcellular components and, ultimately, allowing realistic 3D or 4D cell imaging. The potential application of deep-learning algorithms for automated, unsupervised learning and subsequent data processing is likely to be a key area of relevance for BER researchers. These efforts will require (1) access to multicore, large-memory machines with ultrafast input/output capabilities; (2) low-latency and high-bandwidth communication, GPUs, or many integrated cores (MICs) for fast, large-matrix mathematics; and (3) faster, larger-memory hierarchies for data and model parameter staging to drive developments.

A related set of challenges involves the intelligent processing of data from experimental sources (i.e., mainly microscopy), where advances in applied mathematics are needed to address gaps. Current algorithms for biological cryoEM of macromolecular components scale to that of Extreme Science and Engineering Discovery Environment (XSEDE) resources but with inefficient algorithms that are incapable of handling sparse datasets. New mathematical methods and advances are needed for microscopy related to the cell's macromolecular components, as well as for larger-scale (i.e., volume) imaging and other imaging techniques that currently could scale to exascale computing but do so inefficiently. Investments in these foundational components can maximize output from exascale platforms.

Chapter 9

Summary and Conclusions

Within the U.S. Department of Energy (DOE), the Office of Biological and Environmental Research (BER) supports fundamental research on complex biological processes to address some of the most challenging, long-term energy and environmental issues relevant to DOE BER's mission. The approach to this research, managed by BER's Biological Systems Science Division (BSSD), seeks to understand the processes controlling the translation of information encoded in the genomes of plants and microbes into expressed phenotypes. Gaining an understanding of these mechanisms will lead to improved understanding, manipulation, and design of biological processes for a range of beneficial purposes. The study of plants and their microbial communities exemplifies the complexity of this type of research (see Fig. 9.1, p. 82). Reaching a comprehensive understanding of plants requires research from the molecular to the organismal scale, over time periods spanning picoseconds to years. Structural and characterization studies are needed at vastly different length scales, and they must be coupled with an understanding of genomic, proteomic, metabolomic, and phenotypic variations.

The presentations and discussions at the Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and Environment workshop encompassed very diverse areas of research relevant to BER missions. However, these participant interactions brought to light several common themes. Perhaps not surprisingly, there is a general enthusiasm for increasing the spatial resolution of analysis methods, consistent with expressed research needs. Similarly, there is increased awareness of the need for methods to probe the time dimension in biology. Measuring and following the dynamics of biological systems are challenging but key to

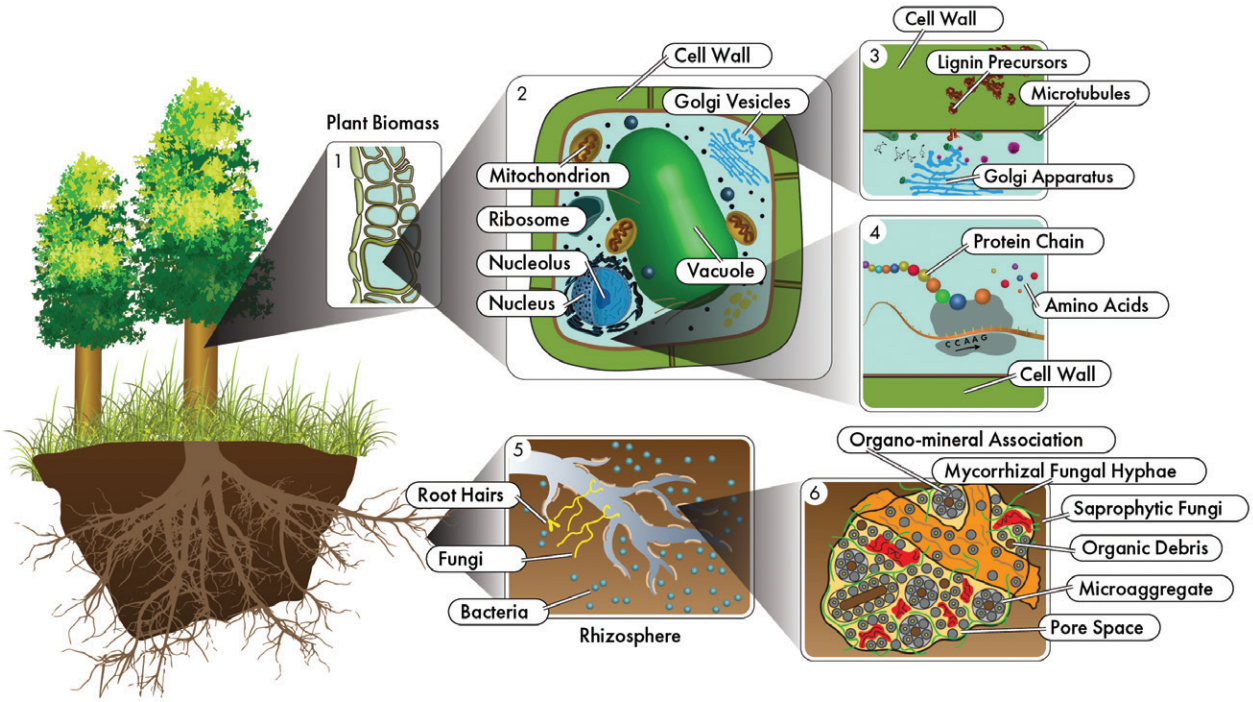


Fig. 9.1. Multiscale Studies of Plant and Microbial Systems and Their Interactions. This illustration highlights the significance of understanding plant and microbial systems and their interactions across hierarchically complex levels of organization from the whole plant and tissue levels to the cellular, subcellular, and microbial community levels. **Aboveground system: (1)** Lignocellulosic biomass consists of heterogeneously composed stem biomass. **(2)** Cell growth and cell wall biosynthesis and remodeling are tightly regulated and coordinated across various subcellular compartments in response to internal and external cues. **(3)** The trans-Golgi network plays a central role in cell wall biosynthesis via transport of proteins (i.e., cellulose synthase complex and wall glycoproteins) and precursors (i.e., hemicellulose, pectin, and lignin) involved in wall polymer synthesis. **(4)** Understanding the effect of nucleotide sequence variation on protein structure and activity (i.e., sequence-function continuum) is critical for understanding and predicting functional outcomes of feedstock improvement programs based on advanced breeding or biosystems design strategies. **Belowground system: (5)** Plant roots and their associated microbial communities (in the rhizosphere) interact for mutual benefit. Microbial communities acquire sugars and other materials from plant roots and assist plant hosts with nutrient and water acquisition, provide pathogen defense, and mediate transformation of root tissues into soil organic matter. **(6)** From this plant-microbial association, soils are further developed into groups of soil particles (i.e., aggregates) that consist of plant and fungal debris; microbially derived organo-mineral associations; mycorrhizal fungal hyphae; organic matter colonized by saprophytic fungi, clay, and polysaccharides; and other pore-space chemicals. [Image courtesy Udaya Kalluri, Oak Ridge National Laboratory. Soil aggregate illustration (6) modified from Jastrow and Miller 1998.]

understanding how microbes and plants function and react to their environment and each other. Another theme spanning the research topics was the ability to measure not only structure (i.e., *where* things are), but also chemistry (i.e., *what* things are and *how* they function). A particular challenge is being able to simultaneously measure chemistry and structure at high resolution. Finally, for many BER-relevant science issues, an important challenge is the ability to conduct measurements *in situ* as well as in the laboratory. Figure 9.2, p. 83, charts the broad range of imaging methods, as well as their length and time scales, that are (1) mature but being developed further, (2) under development now, or (3) need to be developed in the near future. These technologies either are now available to DOE BER researchers or could be in the future after they have been developed further.

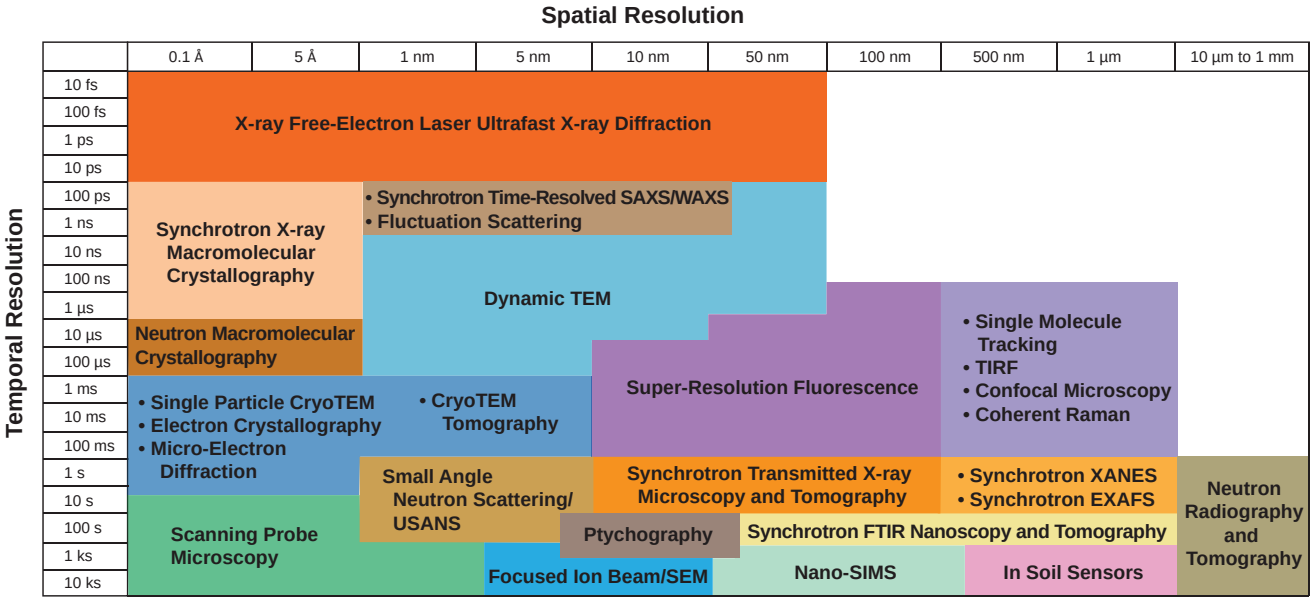


Fig. 9.2. Spatial and Temporal Resolutions of Imaging Technologies. Summary of imaging and other selected measurement technologies discussed at the workshop. Mature technologies that will benefit from further development include X-ray and neutron crystallography and scattering, scanning probe microscopies, X-ray tomography, synchrotron spectroscopy, and confocal microscopy. Techniques undergoing rapid development and with potential application to the mission of the U.S. Department of Energy's Office of Biological and Environmental Research (BER) include X-ray free-electron laser ultrafast diffraction, time-resolved X-ray scattering, cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryoET), infrared imaging methods, super-resolution fluorescence imaging. New technologies not yet fully developed and applied to research supported by BER's Biological Systems Science Division include dynamic EM, fluctuation scattering, ptychography, and in soil sensors.

Key: Å, Angstrom; **EXAFS**, extended X-ray absorption fine structure; **fs**, femtosecond; **ks**, kilosecond; **µm**, micrometer; **µs**, microsecond; **mm**, millimeter; **ms**, millisecond; **nm**, nanometer; **ns**, nanosecond; **ps**, picosecond; **s**, second; **SAXS**, small-angle X-ray scattering; **SEM**, scanning electron microscopy; **TEM**, transmission electron microscopy; **TIRF**, total internal reflection fluorescence; **USANS**, ultrasmall-angle neutron scattering; **WAXS**, wide-angle X-ray scattering; **XANES**, X-ray absorption near edge structure.

Workshop participants agreed that multimodal methods will be required for many of the research needs discussed. Increasingly, as structural biology and imaging methods are converging, multiscale, multidisciplinary approaches to plant and microbial cell biology are emerging. This evolution will be important, particularly for approaches that couple structural and chemical imaging to provide comprehensive characterization. Ideally, the combination of structural imaging and labeling methods that pinpoint the location of specific molecules has the potential, over time, to dramatically change the understanding of microbes, microbial communities, and plants. The ability to accurately track molecules in unfixed biological systems with single-molecule precision will greatly enhance the modeling of molecular function in the cell, especially when coupling that process with higher-resolution studies of enzymes.

BSSD has a long history of developing and supporting the use of genomic characterization of biological systems. This, in part, has led to the high-throughput DNA sequencing technology available to researchers worldwide. In more recent years, genome sequencing has vastly outpaced genome writing and the ability to

interpret genome function. Methods for characterizing biological processes at the molecular and cellular level, important for inferring function, certainly are less mature, and there is a great opportunity for their further development to address issues of genome-coded function. Therefore, to enable more comprehensive systems biology–based approaches, which typically require measurements be made on many samples, workshop participants highlighted the need for development of highly sensitive methods to provide accurate measurements from small sample volumes. In addition, methods need to be operable in high-throughput or highly parallel modes to enable rapid query of biological systems for data that then can be fed iteratively back into the experiment. Achieving these two goals is critical to enabling predictive engineering of biological systems. Although much of the discussion at the workshop centered on measurement, participants also noted that further development of manipulation technologies is critical. Biosystems design has the potential to revolutionize the way biology is exploited to produce economically valuable molecules; however, this will be possible only if high-throughput measurement technologies are combined with tools for precise genetic manipulation and computational algorithms are devised for accurate prediction of phenotype resulting from genome manipulation.

Recent technology developments are expected to address some of the needs raised at the workshop. For photon sources, the successful implementation of the X-ray free-electron laser (XFEL) technology, first implemented at the Linac Coherent Light Source at SLAC National Accelerator Laboratory, has made available unprecedented time scale measurements. The unique properties of XFEL sources have generated several new approaches for studying biological systems—for example, serial femtosecond X-ray crystallography (SFX), fluctuation X-ray scattering, and diffractive imaging. In the future, upgrades to these XFEL facilities will enable the exploitation of lower photon energies, higher repetition rates, and even shorter pulse lengths. Some methods developed at XFELs can be implemented at the DOE synchrotron photon sources, particularly the National Synchrotron Light Source II at Brookhaven National Laboratory with its double-bend achromat lattice, where bright and almost coherent life science beamlines were recently commissioned. Planned upgrades at the Advanced Photon Source at Argonne National Laboratory and Advanced Light Source at Lawrence Berkeley National Laboratory will incorporate multibend achromatic technologies to provide researchers with diffraction-limited photon sources that produce extremely bright and coherent hard and soft X-rays, respectively. This will enable (1) high-brightness experiments that currently are only possible at FEL sources, such as SFX and fluctuation scattering, and (2) completely new methods, such as long-wavelength ptychography that can use X-ray scattering to image both structure and chemistry.

Nuclear magnetic resonance (NMR) approaches have a lower spatial resolution but can be performed on living cells to provide information on native states of cellular lipids, nucleic acids, polysaccharides, and proteins. The development of higher-field magnets will enable the accurate measurement of increasingly dilute samples. Instruments with increased sensitivity and resolution will enhance the ability of NMR to quantify metabolic changes and identify intracellular processes. Ultrahigh fields will provide access to multinuclear NMR, with sufficient sensitivity and resolution to analyze multicomponent systems with isotopic labeling. These fields also will enable the analysis of intrinsically disordered macromolecules and low-population transition states. The ability of NMR to probe the dynamics of molecules in solution is of particular importance for the modeling and simulation of biological complexes.

Mass spectrometry (MS) technologies continue to evolve rapidly, with a strong focus on improved spatial and temporal resolution. Desorption ionization techniques such as nanoscale secondary ion mass spectrometry (nanoSIMS) and matrix-assisted laser

desorption ionization (MALDI) are now routinely used for imaging complex samples such as tissue sections. These technologies are moving into the subcellular regime, enabling quantitative interrogation of single cells and organelles *in situ* (Steinhauser and Lechene 2013; Caprioli 2016). Software tools such as Skyline have made targeted MS a viable alternative for rapid, highly reproducible quantitative analysis of the temporal dynamics of complex systems (i.e., proteomics and metabolomics; MacLean et al. 2010; Glukhova et al. 2013). Top down approaches to analysis of intact proteins provide an integrative view of sequence variations, post-translational modifications, and three-dimensional (3D) structure, enabling a tighter coupling between quantitative functional studies and structural biology (Durbin et al. 2016). The development of new ion-separation technologies, such as structures for lossless ion manipulations (SLIM; Deng et al. 2017) will increase resolution and sensitivity while substantially simplifying the design of high-performance MS systems.

The development of new technologies first must be tightly coupled with driving research questions and, second, must be disseminated to the broad research community in easy-to-access forms. The former suggests continued efforts to build multidisciplinary research programs that bring together biological researchers and technology developers; while the latter requires consideration of a variety of different mechanisms for technology dissemination and the practical, hands-on training of biologists. The facility model has been remarkably successful for many technologies—DOE-supported synchrotrons, XFEL, and neutron sources have led the way in efficiently providing expert technical resources to a diverse research community. This same model is expected to be very successful for emerging techniques such as single-particle cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryoET). The dissemination of imaging technologies often has been more *ad hoc* than facility-based structural biology methods, with researchers initially needing to build their own instruments on the basis of prototypes developed by others. Only after significant long-term efforts do these new technologies mature to the point of being commercially available. DOE BER–supported user facilities, the Environmental Molecular Sciences Laboratory and Joint Genome Institute, have provided researchers with access to newly commercialized as well as further customized technologies through their user programs. The Howard Hughes Medical Institute has taken this concept one step further through the creation of the Advanced Imaging Center at Janelia Farms, which makes imaging technologies it developed available to the scientific community before the instruments are available commercially. Finally, there is the challenge of introducing technologies to the broader BER research community. The BER Facilities Integrating Collaborations for User Science (FICUS) program provides an excellent example of how providing access to multiple technologies through an organized program can stimulate new science (Solomon et al. 2016). Extension of this program to include other user facilities, such as the synchrotron light and neutron sources, potentially can introduce many researchers to technologies that will enable their research.

Throughout the workshop, the important role of computation was discussed many times. The need for new measurement and manipulation technologies to enable future research was clear, but this is accompanied by the need for improved computational methods. Without new developments in computational methods and automation, researchers will not be able to fully exploit the wealth of information that current and new technologies will provide. Beyond the need for sufficient compute and data resources, new algorithms are needed to analyze experimental data to extract weak-signal or new information previously undetected. One particularly challenging problem is that of automated interpretation of 4D single-molecule imaging datasets, which can be terabytes in size and show the movements of molecules inside the cell over time. These datasets rise to the

level of big data and will require new computational approaches for automated feature extraction, dimensionality reduction, and interpretation. In general, automated annotation of complex structural and chemical imaging data remains an unsolved problem. This gap is compounded by the need for new algorithms that can integrate multiple data sources to provide an integrated picture of a biological system. This becomes a very challenging limitation when studying heterogeneous systems, such as microbes and plants, in which no two cells are exactly the same. New computational methods are needed to extract features and map them onto mathematical models for further analysis and simulation.

In conclusion, the workshop highlighted several BER research needs that would be greatly enhanced by further improvement of existing technologies and the development of new technologies for measuring and manipulating biological processes. Such innovative development is needed both for instruments at large-scale user facilities as well as for smaller laboratory-scale imaging instruments, techniques, and methods. Fully realizing the potential of these improved or new developments also will require new computational algorithms and the necessary computing resources.

Appendix A

Workshop Participants

Paul Adams (Chair)
Lawrence Berkeley
National Laboratory

Elizabeth Wright (Chair)
Emory University

Rommie Amaro
University of California,
San Diego

Parastoo Azadi
Complex Carbohydrate
Research Center
University of Georgia

Philip Benfey
Duke University

Joerg Bewersdorf
Yale University

Julie Biteen
University of Michigan

Wah Chiu
Baylor College of Medicine
(Stanford University at time
of publication)

Bob Cottingham
Oak Ridge
National Laboratory

Shi-you Ding
Michigan State University

Jose Dinneny
Carnegie Institution
for Science

James Evans
Pacific Northwest
National Laboratory

Matthew Fields
Montana State University

Brian Fox
University of Wisconsin
Madison

James Fraser
University of California
San Francisco

Britt Hedman
Stanford University

Hoi-Ying Holman
Lawrence Berkeley
National Laboratory

Greg Hura
Lawrence Berkeley
National Laboratory

Farren Isaacs
Yale University

Andrzej Joachimiak
Argonne National Laboratory

Udaya Kalluri
Oak Ridge
National Laboratory

Ken Kemner
Argonne National Laboratory

Carolyn Larabell
University of California
San Francisco

Sean McSweeney
Brookhaven
National Laboratory

Michelle O'Malley
University of California
Santa Barbara

Hugh O'Neill
Oak Ridge
National Laboratory

Jennifer Pett-Ridge
Lawrence Livermore
National Laboratory

Ganesh Sriram
University of Maryland

Elizabeth Villa
University of California
San Diego

Tuan Vo-Dinh
Duke University

Appendix B
Workshop Agenda

Technologies for Characterizing Molecular and Cellular Systems (TCMCS) Relevant to Bioenergy and Environment

September 21–23, 2016
Hilton Washington, D.C./Rockville Hotel

Day 1. Wednesday, September 21

8:30 a.m.	Welcome	Amy Swain
8:40 a.m.	BER/BSSD Background	Todd Anderson
9:05 a.m.	Welcome from Co-Chairs	Paul Adams and Liz Wright
Session 1. BER Enabling Activities – Moderator: Amy Swain		
9:25 a.m.	Facilities Integrating Collaborations for User Science (FICUS) Report	Britt Hedman
9:45 a.m.	TCMCS Bioimaging Principal Investigator (PI) Meeting Report	James Evans
10:05 a.m.	Light/Neutron Source Workshop Report	Sean McSweeney
10:30 a.m.	**Break**	
Keynote Biology Presentation – Moderator: Liz Wright		
10:50 a.m.	The Bioenergy Research Centers: Towards Renewable Lignocellulosic biofuels	Paul Adams
Short Presentations		
11:20 a.m.	Tools for Structural Characterization of Polysaccharides and Glycoproteins	Parastoo Azadi
11:35 a.m.	Functional Annotation of Bioenergy Phylogenetic Space	Brian Fox
11:50 a.m.	Lignin Degradation by Soil Bacteria	Andrzej Joachimiak
12:20 p.m.	**Lunch**	
Keynote Technology Presentations – Moderator: Paul Adams		
1:30 p.m.	Cryo-Electron Microscopy (EM): Investigations from the Molecular to the Mesoscale	Elizabeth Wright
1:50 p.m.	Birth of the Cool: Multitemperature, Multiconformer X-Ray Crystallography and Allosteric Control	Jaime Fraser
Session 2. Cellular Ultrastructure and Physiology – Moderators: Jose Dinneny and Carolyn Larabell		
2:10 p.m.	Signaling Dynamics in Plant Root Cells	Philip Benfey
2:30 p.m.	Imaging subcellular behavior at the molecular scale	Julie Biteen
3:00 p.m.	**Break**	
Short Presentations		
3:30 p.m.	Opening Windows into the Cell: Bringing Structure and Molecular Context to Cell Biology using Cryo-EM	Elizabeth Villa
3:45 p.m.	Fourier Transform Infrared (FTIR) Spectroscopy Imaging: Investigations of Cellular Chemistry from Microns to Nanoscale	Hoi-Ying Holman

4:00 p.m.	Fluorescence Microscopy with Three-Dimensional, Sub-20 nanometer Resolution	Joerg Bewersdorf
Breakout Discussions (Jackson and Monroe Rooms)		
4:30 p.m.	Metabolic Pathways in Plants, Microbes, and Fungi	Jose Dinneny
	Cellular Structure, Organization, Signaling, Networks	Carolyn Larabell
6:00 p.m.	Breakout Session Reports	Paul Adams and Liz Wright
6:30 p.m.	**Adjourn**	

Day 2. Thursday, September 22

Technology Keynote Presentations – Moderator: Liz Wright		
8:30 a.m.	Multiscale Dynamics: Molecules to Cells	Rommie Amaro
8:50 a.m.	Plasmonic Nanoprobe Technology for Monitoring Genomic Biomarkers and Molecular Pathways	Tuan Vo-Dinh
Session 3. Bioenergy and Bioproducts Production – Moderators: Shi-You Ding and Farren Isaacs		
9:10 a.m.	Exploiting Anaerobes for Biomass Breakdown and Sustainable Chemistry	Michelle O'Malley
9:30 a.m.	Fundamental Studies of Biomass Formation to Enable Sustainable Biofuel Production	Udaya Kalluri
10:00 a.m.	**Break**	
Breakout Discussions (Jackson and Monroe Rooms)		
10:20 a.m.	Plant Cell Wall Composition and Degradation	Shi-you Ding
	Synthetic Biology (Biosystems Design)	Farren Isaacs
11:50 p.m.	**Breakout Session Reports – Moderators: Paul Adams and Liz Wright**	
12:20 p.m.	**Lunch**	
Short Presentations		
1:30 p.m.	Neutron Technologies for Characterizing Molecular to Mesoscale Biological Processes	Hugh O'Neill
1:45 p.m.	Combining Small-Angle X-Ray Scattering (SAXS) and High-Resolution Structure for Synthetic Biology	Greg Hura
2:00 p.m.	Instationary Isotope Labeling Dissects Metabolic Cycles in a Photoautotrophic Organism	Ganesh Sriram
Session 4. Environmental Microbiology – Moderator: Ken Kemner		
2:30 p.m.	Microbial Interactions at Different Scales	Matthew Fields
2:50 p.m.	Using Isotopes and Imaging to Track Microbe-Mineral Interactions	Jennifer Pett-Ridge
3:30 p.m.	**Break**	
Breakout Discussions (Jackson and Monroe Rooms)		
4:00 p.m.	Community interactions including rhizosphere	Jennifer Pett-Ridge
	Biogeochemical cycling of elements	Ken Kemner
5:30 p.m.	**Breakout Session Reports – Moderators: Paul Adams and Liz Wright**	

6:00 p.m.	**Day 1 and 2 Summary**	
6:30 p.m.	**Adjourn**	

Day 3. Friday, September 23

Session 5. Community Access to Technology – Moderators: Paul Adams and Liz Wright

8:30 a.m.	Cryo-EM Research Resource and Training to a Broad Community	Wah Chiu
8:50 a.m.	DOE Knowledgebase (KBase) Capabilities and Extension to Address Computation and Data Challenges for Systems Biology Research	Bob Cottingham
9:15 a.m.	**Round Table Discussion of the Charge Questions**	Paul Adams and Liz Wright
10:00 a.m.	**Break**	
10:30 a.m.	Develop Plan for Workshop Report Writing and Action Items	Paul Adams and Liz Wright
12:00 p.m.	**Adjourn**	

Appendix C

Figure Credits

Cover, Figure 1.1., and chapter front images

(clockwise from top)

Biogeochemical cycling. (Chapter 4) Courtesy Ken Kemner, Argonne National Laboratory.

Cell wall composition and degradation. (Chapter 2) Courtesy Udaya Kalluri, Oak Ridge National Laboratory.

Cellular ultrastructure and physiology. (Chapter 7) Overview of the mitochondria network visualized by immunolabeling TOM20 with Alexa Flour 647. Courtesy Huang, F., et al. 2016. "Ultra-High Resolution 3D Imaging of Whole Cells," *Cell* **166**(4), 1028–40.

Biosystems design. (Chapter 6) GLAMM metabolic map. Image courtesy Lawrence Berkeley National Laboratory. Bates, J. T., et al. 2011. "GLAMM: Genome-Linked Application for Metabolic Maps," *Nucleic Acids Research* **39**, W400–05.

Metabolic pathways in plants, microbes, and fungi. (Chapter 5) GLO-Roots image of soil-grown root system. Courtesy Rubén Rellán-Álvarez and José R. Dinneny, Carnegie Institute for Science.

Rhizosphere community interactions. (Chapter 3) Confocal microscope images of bacteria on the surface of poplar roots. Courtesy J. L. Morrell-Falvey, Oak Ridge National Laboratory.

Data integration and analysis. (Chapter 8) Mesoscale molecular model. [Reprinted from Macmillan Publishers Ltd: Nature Methods. Johnson, G. T., et al. 2015. "cellPACK: A Virtual Mesoscope to Model and Visualize Structural Systems Biology," *Nature Methods* **12**(1), 85–95. © 2015.]

Figure 1.2.

Row 1, Image 1. Nitrogenase. Courtesy David S. Goodsell and the RCSB Protein Data Bank.

Row 1, Image 2. PF1205. Reprinted by permission from Macmillan Publishers Ltd.: Hura, G. L., et al. 2009. "Robust, High-Throughput Solution Structural Analyses by Small Angle X-ray Scattering (SAXS)," *Nature Methods* 6, 606–12. © 2009.

Row 1, Image 3. Comprehensive imaging of densely packed transmembrane proteins using photoactivated localization microscopy (PALM). Courtesy Greenfield, D., et al. 2009. "Self-Organization of the *Escherichia coli* Chemotaxis Network Imaged with Super-Resolution Light Microscopy," *PLoS Biology* 7(**6**): e1000137.

Row 1, Image 4. Confocal microscope image of bacteria on the surface of poplar roots. Courtesy J. L. Morrell-Falvey, Oak Ridge National Laboratory.

Row 1, Image 5. Scanning electron microscopy (SEM) image of a sulfate-reducing biofilm obtained from a borehole used for long-term (100+ days) acetate injection during biostimulation activities at DOE's Integrated Field Research Challenge site near Rifle, Colorado. Courtesy A. Dohnalkova, Pacific Northwest National Laboratory. Biofilm material courtesy K. H. Williams, Lawrence Berkeley National Laboratory.

Row 1, Image 6. Positron emission tomography (PET) image showing uptake of radioactive nitrate in a poplar sapling. Courtesy Richard Ferrieri, University of Missouri.

Row 1, Image 7. X-ray computed tomography cross section of an active layer/permafrost soil core from the DOE BER NGEE–Arctic project.

Courtesy Tim Neafsey, Lawrence Berkeley National Laboratory.

Row 2, Image 1. Haem-copper active site in respiratory enzymes. Reprinted by permission from Macmillan Publishers Ltd.: Hura, G.L., et al. 2009. "The Octahaem MccA is a Haem *c*–Copper Sulfite Reductase," *Nature* **520**, 706–09. © 2015.

Row 2, Image 2. Density difference map for the ensemble structure $(mF_{obs} - DF_{model})\exp[i\varphi_{model}]$**.** Reprinted under a Creative Commons Attribution License (CC BY 4.0) from Burnley, B.T., et al. 2012. "Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement," eLIFE 1, e00311.

Row 2, Image 3. Acetate consumption over time of *Psychrobacter aestuarii* and *Geobacter sulfurreducens* in axenic cultures and co-cultures. Courtesy Alice C. Dohnalkova, Pacific Northwest National Laboratory.

Row 2, Image 4. Transmission electron microscope (TEM) image of *Escherichia coli* O157:H7. Courtesy Centers for Disease Control and Prevention.

Row 2, Image 5. Plant root nutrient uptake. Courtesy Jennifer Pett-Ridge and Erin Nuccio, Lawrence Livermore National Laboratory.

Row 2, Image 6. Deconstructed biomass. Courtesy National Renewable Energy Laboratory Biomass Structural Characterization Laboratory.

Row 2, Image 7. Engineered bacteria produce biodiesel. Courtesy Joint BioEnergy Institute, Lawrence Berkeley National Laboratory.

Appendix D

Acronyms and Abbreviations

2D, 3D, 4D	two-, three-, and four-dimensional
Å	Angstrom
AFM	atomic force microscopy
AMF	arbuscular-mycorrhizal fungi
AMS	accelerator mass spectrometry
antiSMASH	antibiotics and Secondary Metabolite Analysis Shell
ASCR	Office of Advanced Scientific Computing Research, DOE Office of Science
BAC	bacterial artificial chromosome
BER	Office of Biological and Environmental Research, DOE Office of Science
BES	Office of Basic Energy Sciences, DOE Office of Science
BSSD	BER Biological Systems Science Division
C5, C6	carbon cellulosic sugar molecules, such as xylose and glucose, respectively
CAGE	conjugative assembly genome engineering
CARD	catalyzed reporter deposition
CARS	coherent anti-Stokes Raman scattering
Cas9	CRISPR-associated protein-9
CEF	cellulose elementary fibril
CesA	cellulose synthase
CH_4	methane
CLEM	correlative light and electron microscopy
cm	centimeter
CME	clathrin-mediated endocytosis
CO_2	carbon dioxide
CRISPR	clustered regularly interspaced short palindromic repeats
CRS	coherent Raman scattering
cryo	cryogenic
cryoEM	cryo-electron microscopy
cryoET	cryo-electron tomography
CSC	cellulose synthase complexes (proteins)
DNA	deoxyribonucleic acid
DGT	diffusive gradients in thin films
DMAPP	dimethylallyl diphosphate
DOE	U.S. Department of Energy
dsDNA	double-stranded DNA
EM	electron microscopy
EMSL	BER Environmental Molecular Sciences Laboratory
EXAFS	extended X-ray absorption fine structure
FEL	free-electron laser
FIB	focused ion beam SEM
FICUS	Facilities Integrating Collaborations for User Science
FISH	fluorescence *in situ* hybridization
fLM	fluorescence light microscopy
FPP	farnesyl diphosphate
FRET	Förster resonance energy transfer
fs	femtosecond
FTIR	Fourier transform infrared microscopy
GC	gas chromatography
GFP	green fluorescent protein
GM	genetically modified
GPR	ground-penetrating radar
GPU	graphical processing units
HGT	horizontal gene transfer
HIM	helium ion microscopy
HIM-SIMS	HIM with SIMS
HPC	high-performance computing
HPFM	hybrid photonic force microscopy or Hybrid Photonic Mode-Synthesizing Atomic Force Microscope
IPP	isopentenyl diphosphate
JGI	DOE Joint Genome Institute
KBase	DOE Systems Biology Knowledgebase
LCLS	DOE Linac Coherent Light Source at SLAC
LPMO	lytic polysaccharide mono-oxygenase

m	meter
MALDI	matrix-assisted laser desorption ionization
MASC	microtubule-associated cellulose synthase compartments
MD	molecular dynamics
MFA	metabolic flux analysis
min	minute
mm	millimeter
MRM	multiple reaction monitoring
ms	millisecond
MS	mass spectrometry
MS/MS	tandem mass spectrometry
µm	micrometer
µs	microsecond
µTomography	microtomography
MX	macromolecular X-ray crystallography
NADPH	nicotinamide adenine dinucleotide phosphate
nanoSIMS	nanoscale secondary ion mass spectrometry
NERSC	National Energy Research Scientific Computing Center
NIMS	nanostructure-initiator mass spectrometry
nm	nanometer
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
ns	nanosecond
NSLS II	National Synchrotron Light Source II, at Brookhaven National Laboratory
NX	neutron macromolecular crystallography
OTS	orthogonal translation system
PALM	photoactivated localization microscopy
PCR	polymerase chain reaction
PET	positron emission tomography
ps	picosecond
RNA	ribonucleic acid
s	second
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SEM	scanning EM
sFTIR	synchrotron FTIR
SIMS	secondary ion mass spectrometry
SINS	synchrotron infrared nanospectroscopy
SIP	stable-isotope probing
SLAC	DOE SLAC National Accelerator Laboratory
SLIM	structures for lossless ion manipulations
SOM	soil organic matter
SRM	selected reaction monitoring
ssDNA	single-stranded DNA
SSN	sequence-specific nuclease
s-SNOM	scanning near-field optical microscopy with scattering
STD	saturation transfer difference (NMR spectroscopy)
STED	stimulated emission depletion
STORM	stochastic optical reconstruction microscopy
STXM	scanning transmission X-ray microscopy
TAA, TAG	nucleotide triplets stop (or terminating) codon
TALENS	transcription activator-like effector nucleases
TEM	transmission electron microscopy
TIRF	total internal reflection fluorescence
TOF-MS	time-of-flight MS
tr-NOESY	transfer-NOESY
tRNA	transfer RNA
TXM	transmission X-ray microscopy
USANS	ultrasmall-angle neutron scattering
WAXS	wide-angle X-ray scattering
XANES	X-ray absorption near edge structure (spectroscopy)
XFEL	X-ray free-electron laser
XSEDE	Extreme Science and Engineering Discovery Environment
ZFN	zinc finger nucleases

Appendix E

References

Amenabar, I., et al. 2017. "Hyperspectral Infrared Nanoimaging of Organic Samples Based on Fourier Transform Infrared Nanospectroscopy," *Nature Communications* **8**, 14402. DOI:10.1038/ncomms14402.

Amunts, A., et al. 2015. "Ribosome. The Structure of the Human Mitochondrial Ribosome," *Science* **348**(6230), 95–98. DOI:10.1126/science.aaa1193.

Amunts, A., et al. 2014. "Structure of the Yeast Mitochondrial Large Ribosomal Subunit," *Science* **343**(6178), 1485–89. DOI:10.1126/science.1249410.

Anarat-Cappillino, G., and E. S. Sattely. 2014. "The Chemical Logic of Plant Natural Product Biosynthesis," *Current Opinion in Plant Biology* **19**, 51–58. DOI:10.1016/j.pbi.2014.03.007.

Anderson, R. T., et al. 2003. "Stimulating the In Situ Activity of *Geobacter* Species to Remove Uranium from the Groundwater of a Uranium-Contaminated Aquifer," *Applied and Environmental Microbiology* **69**(10), 5884–91. DOI: 10.1128/AEM.69.10.5884-5891.2003.

Antoniewicz, M. R. 2015. "Methods and Advances in Metabolic Flux Analysis: A Mini-Review," *Journal of Industrial Microbiology & Biotechnology* **42**(3), 317–25. DOI:10.1007/s10295-015-1585-x.

Antoniewicz, M. R. 2014. "Advances in ^{13}C Metabolic Flux Analysis: Complete-MFA, Co-Culture MFA and Dynamic MFA." In proceedings of *Metabolic Engineering X*, June 15–19, 2014. Vancouver, Canada. American Institute of Chemical Engineers, New York, NY. Printed from e-media with permission by Curran Associates, Inc., Red Hook, NY.

Anunciado, D. B., et al. 2017. "*In Vivo* Protein Dynamics on the Nanometer Length Scale and Nanosecond Time Scale," *Journal of Physical Chemistry Letters* **8**(8), 1899–904. DOI:10.1021/acs.jpclett.7b00399.

Aquila, A., et al. 2012. "Time-Resolved Protein Nanocrystallography Using an X-Ray Free-Electron Laser," *Optics Express* **20**(3), 2706–16. DOI:10.1364/OE.20.002706.

Aranda, R., et al. 2006. "Time-Dependent Atomic Coordinates for the Dissociation of Carbon Monoxide from Myoglobin," *Acta Crystallographica Section D Biological Crystallography* **62**(Pt 7), 776–83. DOI:10.1107/S0907444906017318.

Ayyer, K., et al. 2016. "Macromolecular Diffractive Imaging Using Imperfect Crystals," *Nature* **530**(7589), 202–06. DOI:10.1038/nature16949.

Baelum, J., et al. 2012. "Deep-Sea Bacteria Enriched by Oil and Dispersant from the Deepwater Horizon Spill," *Environmental Microbiology* **14**(9), 2405–16. DOI:10.1111/j.1462-2920.2012.02780.x.

Bai, X. C., et al. 2015. "Sampling the Conformational Space of the Catalytic Subunit of Human γ-Secretase," *eLife* **4**. DOI:10.7554/eLife.11182.

Bailey, V. L., et al. 2013. "Micrometer-Scale Physical Structure and Microbial Composition of Soil Macroaggregates," *Soil Biology and Biochemistry* **65**, 60–68. DOI:10.1016/j.soilbio.2013.02.005.

Bale, J. B., et al. 2016. "Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes," *Science* **353**(6297), 389–94. DOI:10.1126/science.aaf8818.

Banoub, J., et al. 2015. "A Critique on the Structural Analysis of Lignins and Application of Novel Tandem Mass Spectrometric Strategies to Determine Lignin Sequencing," *Journal of Mass Spectrometry* **50**(1), 5–48. DOI:10.1002/jms.3541.

Baran, R., et al. 2015. "Exometabolite Niche Partitioning among Sympatric Soil Bacteria," *Nature Communications* **6**, 8289. DOI:10.1038/ncomms9289.

Bartesaghi, A., et al. 2014. "Structure of β-Galactosidase at 3.2-Å Resolution Obtained by Cryo-Electron Microscopy," *Proceedings of the National Academy of Sciences of the USA* **111**(32), 11709–14. DOI:10.1073/pnas.1402809111.

Barton, C. V., and K. D. Montagu. 2004. "Detection of Tree Roots and Determination of Root Diameters by Ground Penetrating Radar Under Optimal Conditions," *Tree Physiology* **24**(12), 1323–31. DOI:10.1093/treephys/24.12.1323.

Bashline, L., et al. 2014. "The Trafficking of the Cellulose Synthase Complex in Higher Plants," *Annals of Botany* **114**(6), 1059–67. DOI:10.1093/aob/mcu040.

Bassel, G. W., et al. 2012. "Systems Analysis of Plant Functional, Transcriptional, Physical Interaction, and Metabolic Networks," *Plant Cell* **24**(10), 3859–75. DOI:10.1105/tpc.112.100776.

Bechtel, H. A., et al. 2014. "Ultrabroadband Infrared Nanospectroscopic Imaging," *Proceedings of the National Academy of Sciences of the USA* **111**(20), 7191–96. DOI:10.1073/pnas.1400502111.

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.

Berg, P., et al. 1975. "Summary Statement of the Asilomar Conference on Recombinant DNA Molecules," *Proceedings of the National Academy of Sciences of the USA* **72**(6), 1981–84.

Berlow, E. L., et al. 2009. "Simple Prediction of Interaction Strengths in Complex Food Webs," *Proceedings of the National Academy of Sciences of the USA* **106**(1), 187–91. DOI:10.1073/pnas.0806823106.

Berman, H. M., et al. 2000. "The Protein Data Bank," *Nucleic Acids Research* **28**(1), 235–42. DOI:10.1093/nar/28.1.235.

Bernardi, R. C., et al. 2015. "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems," *Biochimica et Biophysica Acta* **1850**(5), 872–77. DOI:10.1016/j.bbagen.2014.10.019.

Betzig, E., et al. 2006. "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution," *Science* **313**(5793), 1642–45. DOI:10.1126/science.1127344.

Bhabha, G., et al. 2015. "Keep on Moving: Discovering and Perturbing the Conformational Dynamics of Enzymes," *Accounts of Chemical Research* **48**(2), 423–30. DOI:10.1021/ar5003158.

Binzer, A., et al. 2012. "The Dynamics of Food Chains under Climate Change and Nutrient Enrichment," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367**(1605), 2935–44. DOI:10.1098/rstb.2012.0230.

Blakeley, M. P., et al. 2008. "Neutron Crystallography: Opportunities, Challenges, and Limitations," *Current Opinion in Structural Biology* **18**(5), 593–600. DOI:10.1016/j.sbi.2008.06.009.

Blin, K., et al. 2013. "antiSMASH 2.0––A Versatile Platform for Genome Mining of Secondary Metabolite Producers," *Nucleic Acids Research* **41**(W1), W204–12. DOI:10.1093/nar/gkt449.

Boeke, J. D., et al. 2016. "The Genome Project-Write," *Science*: *Genome Engineering* **353**(6295), 126–27. DOI:10.1126/science.aaf6850.

Boit, A., et al. 2012. "Mechanistic Theory and Modelling of Complex Food-Web Dynamics in Lake Constance," *Ecology Letters* **15**(6), 594–602. DOI:10.1111/j.1461-0248.2012.01777.x.

Boras, B. W., et al. 2015. "Bridging Scales Through Multiscale Modeling: A Case Study on Protein Kinase A," *Frontiers in Physiology* **6**, 250. DOI:10.3389/fphys.2015.00250.

Bottanelli, F., et al. 2016. "Two-Colour Live-Cell Nanoscale Imaging of Intracellular Targets," *Nature Communications* **7**, 10778. DOI:10.1038/ncomms10778.

Bouskill, N. J., et al. 2016. "Belowground Response to Drought in a Tropical Forest Soil. II. Change in Microbial Function Impacts Carbon Composition," *Frontiers in Microbiology* **7**, 323. DOI:10.3389/fmicb.2016.00323.

Boyken, S. E., et al. 2016. "De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network–Mediated Specificity," *Science* **352**(6286), 680–87. DOI:10.1126/science.aad8865.

Briegel, A., et al. 2008. "Location and Architecture of the *Caulobacter Crescentus* Chemoreceptor Array," *Molecular Microbiology* **69**(1), 30–41. DOI:10.1111/j.1365-2958.2008.06219.x.

Brose, U., and S. Scheu. 2014. "Into Darkness: Unravelling the Structure of Soil Food Webs," *Oikos* **123**(10), 1153–56. DOI:10.1111/oik.01768.

Brunette, T. J., et al. 2015. "Exploring the Repeat Protein Universe Through Computational Protein Design," *Nature* **528**(7583), 580–84. DOI:10.1038/nature16162.

Bucksch, A., et al. 2014. "Image-Based High-Throughput Field Phenotyping of Crop Roots," *Plant Physiology* **166**(2), 470–86. DOI:10.1104/pp.114.243519.

Burnley, B. T., et al. 2012. "Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement," *eLife* **1**, e00311. DOI:10.7554/eLife.00311.

Cai, Y., et al. 2015. "Intrinsic Biocontainment: Multiplex Genome Safeguards Combine Transcriptional and Recombinational Control of Essential Yeast Genes," *Proceedings of the National Academy of Sciences of the USA* **112**(6), 1803–08. DOI:10.1073/pnas.1424704112.

Campbell, E., et al. 2016. "The Role of Protein Dynamics in the Evolution of New Enzyme Function," *Nature Chemical Biology* **12**(11), 944–50. DOI:10.1038/nchembio.2175.

Cane, D. E., et al. 1998. "Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations," *Science: Biochemistry* **282**(5386), 63–68. DOI:10.1126/science.282.5386.63.

Canfield, D. E., et al. 1993. "The Anaerobic Degradation of Organic Matter in Danish Coastal Sediments: Iron Reduction, Manganese Reduction, and Sulfate Reduction," *Geochimica et Cosmochimica Acta* **57**(16), 3867–83. DOI:10.1016/0016-7037(93)90340-3.

Caprioli, R. M. 2016. "Imaging Mass Spectrometry: Molecular Microscopy for the New Age of Biology and Medicine," *Proteomics* **16**(11–12), 1607–12. DOI:10.1002/pmic.201600133.

Carlson, R. 2016. "Estimating the Biotech Sector's Contribution to the US Economy," *Nature Biotechnology* **34**(3), 247–55. DOI:10.1038/nbt.3491.

Centrone, A. 2015. "Infrared Imaging and Spectroscopy Beyond the Diffraction Limit," *Annual Review of Analytical Chemistry (Palo Alto Calif)* **8**, 101–26. DOI:10.1146/annurev-anchem-071114-040435.

Chae, L., et al. 2014. "Genomic Signatures of Specialized Metabolism in Plants," *Science* **344**(6183), 510–13. DOI:10.1126/science.1252076.

Chaikeeratisak, V., et al. 2017. "Assembly of a Nucleus-Like Structure During Viral Replication in Bacteria," *Science* **355**(6321), 194–97. DOI:10.1126/science.aal2130.

Chan, C. T., et al. 2016. " 'Deadman' and 'Passcode' Microbial Kill Switches for Bacterial Containment," *Nature Chemical Biology* **12**(2), 82–86. DOI:10.1038/nchembio.1979.

Chapman, H. N. 2010. "Microscopy: A New Phase for X-Ray Imaging," *Nature* **467**(7314), 409–10. DOI:10.1038/467409a.

Chen, G., et al. 2013. "Component Particle Structure in Heterogeneous Disordered Ensembles Extracted from High-Throughput Fluctuation X-Ray Scattering," *Physical Review Letters* **110**(19), 195501. DOI:10.1103/PhysRevLett.110.195501.

Chimungu, J. G., et al. 2014. "Large Root Cortical Cell Size Improves Drought Tolerance in Maize," *Plant Physiology* **166**(4), 2166–78. DOI:10.1104/pp.114.250449.

Chiu, W., et al. 2006. "Structural Biology of Cellular Machines," *Trends Cell Biology* **16**(3), 144–50. DOI:10.1016/j.tcb.2006.01.002.

Choi, Y. S., et al. 2013. "Targeted Human Cerebrospinal Fluid Proteomics for the Validation of Multiple Alzheimer's Disease Biomarker Candidates," *Journal of Chromatography B* **930**, 129–35. DOI:10.1016/j.jchromb.2013.05.003.

Choudhary, S., et al. 2012. "Engineered Protein Nano-Compartments for Targeted Enzyme Localization," *PLOS ONE* **7**(3), e33342. DOI:10.1371/journal.pone.0033342.

Chudyk, E. I., et al. 2014. "QM/MM Simulations as an Assay for Carbapenemase Activity in Class A β-Lactamases," *Chemical Communications* **50**(94), 14736–39. DOI:10.1039/c4cc06495j.

Cohen, M. F., et al. 2015. "Bimodal Effect of Hydrogen Peroxide and Oxidative Events in Nitrite-Induced Rapid Root Abscission by the Water Fern *Azolla pinnata*," *Frontiers in Plant Science* **6**, article 518. DOI:10.3389/fpls.2015.00518.

Cong, L., et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science* **339**(6121), 819–23. DOI:10.1126/science.1231143.

da Silva, R. R., et al. 2015. "Illuminating the Dark Matter in Metabolomics," *Proceedings of the National Academy of Sciences of the USA* **112**(41), 12549–50. DOI:10.1073/pnas.1516878112.

Dai, W., et al. 2014. "Zernike Phase-Contrast Electron Cryotomography Applied to Marine Cyanobacteria Infected with Cyanophages," *Nature Protocols* **9**(11), 2630–42. DOI:10.1038/nprot.2014.176.

Dai, W., et al. 2013. "Visualizing Virus Assembly Intermediates inside Marine Cyanobacteria," *Nature* **502**(7473), 707–10. DOI:10.1038/nature12604.

Daly, K. R., et al. 2015. "Assessing the Influence of the Rhizosphere on Soil Hydraulic Properties Using X-Ray Computed Tomography and Numerical Modelling," *Journal of Experimental Botany* **66**(8), 2305–14. DOI:10.1093/jxb/eru509.

Danev, R., and K. Nagayama. 2001. "Transmission Electron Microscopy with Zernike Phase Plate," *Ultramicroscopy* **88**(4), 243–52. DOI:10.1016/S0304-3991(01)00088-2.

Danev, R., et al. 2014. "Volta Potential Phase Plate for In-Focus Phase Contrast Transmission Electron Microscopy," *Proceedings of the National Academy of Sciences of the USA* **111**(44), 15635–40. DOI:10.1073/pnas.1418377111.

Das, A., et al. 2015. "Digital Imaging of Root Traits (DIRT): A High-Throughput Computing and Collaboration Platform for Field-Based Root Phenomics," *Plant Methods* **11**(1), 51. DOI:10.1186/s13007-015-0093-3.

Dathe, A., et al. 2016. "Impact of Axial Root Growth Angles on Nitrogen Acquisition in Maize Depends on Environmental Conditions," *Annals of Botany* **118**(3), 401–14. DOI:10.1093/aob/mcw112.

Davies, K. M., et al. 2012. "Structure of the Yeast F_1F_o-ATP Synthase Dimer and Its Role in Shaping the Mitochondrial Cristae," *Proceedings of the National Academy of Sciences of the USA* **109**(34), 13602–07. DOI:10.1073/pnas.1204593109.

De Michele, R., et al. 2013. "Fluorescent Sensors Reporting the Activity of Ammonium Transceptors in Live Cells," *eLife* **2**, e00800. DOI:10.7554/eLife.00800.

Dekas, A. E., and V. J. Orphan. 2011. "Chapter Twelve – Identification of Diazotrophic Microorganisms in Marine Sediment via Fluorescence *In Situ* Hybridization Coupled to Nanoscale Secondary Ion Mass Spectrometry (FISH-NanoSIMS)," *Methods in Enzymology* **486**, 281–305. DOI:10.1016/B978-0-12-381294-0.00012-2.

del Río, J. C., et al. 2016. "Lignin–Carbohydrate Complexes from Sisal (*Agave sisalana*) and Abaca (*Musa textilis*): Chemical Composition and Structural Modifications During the Isolation Process," *Planta* **243**(5), 1143–58. DOI:10.1007/s00425-016-2470-1.

Deng, L., et al. 2017. "Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS Using Structures for Lossless Ion Manipulations," *Analytical Chemistry* **89**(8), 4628–34. DOI:10.1021/acs.analchem.7b00185.

Denk, W., and H. Horstmann. 2004. "Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure," *PLOS Biology* **2**(11), e329. DOI:10.1371/journal.pbio.0020329.

Dick-Perez, M., et al. 2011. "Structure and Interactions of Plant Cell-Wall Polysaccharides by Two- and Three-Dimensional Magic-Angle-Spinning Solid-State NMR," *Biochemistry* **50**(6), 989–1000. DOI:10.1021/bi101795q.

Digel, C., et al. 2014. "Unravelling the Complex Structure of Forest Soil Food Webs: Higher Omnivory and More Trophic Levels," *Oikos* **123**(10), 1157–72. DOI:10.1111/oik.00865.

DiMaio, F., and W. Chiu. 2016. "Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps," *Methods in Enzymology* **579**, 255–76. DOI:10.1016/bs.mie.2016.06.003.

Ding, S. Y., and M. E. Himmel. 2006. "The Maize Primary Cell Wall Microfibril: A New Model Derived from Direct Visualization," *Journal of Agricultural and Food Chemistry* **54**(3), 597–606. DOI:10.1021/jf051851z.

Ding, S. Y., et al. 2012. "How Does Plant Cell Wall Nanoscale Architecture Correlate with Enzymatic Digestibility?" *Science* **338**(6110), 1055–60. DOI:10.1126/science.1227491.

Do, M., et al. 2015. "Imaging and Characterizing Cells Using Tomography," *Archives Biochemistry Biophysics* **581**, 111–21. DOI:10.1016/j.abb.2015.01.011.

Donatelli, J. J., et al. 2015. "Iterative Phasing for Fluctuation X-Ray Scattering," *Proceedings of the National Academy of Sciences of the USA* **112**(33), 10286–91. DOI:10.1073/pnas.1513738112.

Donatelli, J. J., et al. 2017. "Reconstruction from Limited Single-Particle Diffraction Data via Simultaneous Determination of State, Orientation, Intensity, and Phase," *Proceedings of the National Academy of Sciences of the USA* **114**(28), 7222–27. DOI:10.1073/pnas.1708217114.

Dong, Y., et al. 2016. "More Than Pictures: When MS Imaging Meets Histology," *Trends in Plant Science* **21**(8), 686–98. DOI:10.1016/j.tplants.2016.04.007.

Downie, H., et al. 2012. "Transparent Soil for Imaging the Rhizosphere," *PLOS ONE* **7**(9), e44276. DOI:10.1371/journal.pone.0044276.

Downie, H. F., et al. 2014. "Transparent Soil Microcosms Allow 3D Spatial Quantification of Soil Microbiological Processes *In Vivo*," *Plant Signaling & Behavior* **9**(10), e970421. DOI:10.4161/15592316.2014.970421.

Duhrkop, K., et al. 2015. "Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID," *Proceedings of the National Academy of Sciences of the USA* **112**(41), 12580–85. DOI:10.1073/pnas.1509788112.

Durbin, K. R., et al. 2016. "Quantitation and Identification of Thousands of Human Proteoforms Below 30 kDa," *Journal of Proteome Research* **15**(3), 976–82. DOI:10.1021/acs.jproteome.5b00997.

Durrant, J. D., and R. E. Amaro. 2014. "LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry," *PLOS Computational Biology* **10**(7), e1003720. DOI:10.1371/journal.pcbi.1003720.

Eichorst, S. A., et al. 2015. "Advancements in the Application of NanoSIMS and Raman Microspectroscopy to Investigate the Activity of Microbial Cells in Soils," *FEMS Microbiology Ecology* **91**(10). DOI:10.1093/femsec/fiv106.

Evans, C. L., and X. S. Xie. 2008. "Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine," *Annual Review of Analytical Chemistry* **1**, 883–909. DOI:10.1146/annurev.anchem.1.031207.112754.

Evans, J. E., and N. D. Browning. 2013. "Enabling Direct Nanoscale Observations of Biological Reactions with Dynamic TEM," *Microscopy* **62**(1), 147–56. DOI:10.1093/jmicro/dfs081.

Evert, R. F. 2006. *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*. Third Ed. John Wiley & Sons, Inc. Hoboken, New Jersey.

Faradjian, A. K., and R. Elber 2004. "Computing Time Scales from Reaction Coordinates by Milestoning," *Journal of Chemical Physics* **120**(23), 10880–89. DOI:10.1063/1.1738640.

Fiore, G., et al. 2016. "*In-Silico* Analysis and Implementation of a Multicellular Feedback Control Strategy in a Synthetic Bacterial Consortium," *ACS Synthetic Biology* **6**(3), 507–17. DOI:10.1021/acssynbio.6b00220.

Flannery, B. P., et al. 1987. "Three-Dimensional X-Ray Microtomography," *Science* **237**(4821), 1439–44. DOI:10.1126/science.237.4821.1439.

Frank, M., et al. 2014. "Femtosecond X-Ray Diffraction from Two-Dimensional Protein Crystals," *International Union of Crystallography* **1**(Pt 2), 95–100. Ed. H. Chapman. DOI:10.1107/S2052252514001444.

Frederick, K. K., et al. 2015. "Sensitivity-Enhanced NMR Reveals Alterations in Protein Structure by Cellular Milieus," *Cell* **163**(3), 620–28. DOI:10.1016/j.cell.2015.09.024.

Freudiger, C. W., et al. 2008. "Label-Free Biomedical Imaging with High Sensitivity by Stimulated Raman Scattering Microscopy," *Science* **322**(5909), 1857–61. DOI:10.1126/science.1165758.

Freund, D. M., and A. D. Hegeman. 2017. "Recent Advances in Stable Isotope-Enabled Mass Spectrometry-Based Plant Metabolomics," *Current Opinion in Biotechnology* **43**, 41–48. DOI:10.1016/j.copbio.2016.08.002.

Fromme, R., et al. 2015. "Serial Femtosecond Crystallography of Soluble Proteins in Lipidic Cubic Phase," *International Union of Crystallography* **2**(Pt 5), 545–51. DOI:10.1107/S2052252515013160.

Fukushima, A., et al. 2009. "Integrated Omics Approaches in Plant Systems Biology," *Current Opinion in Chemical Biology* **13**(5–6), 532–38. DOI:10.1016/j.cbpa.2009.09.022.

Fuller, F. D., et al. 2017. "Drop-on-Demand Sample Delivery for Studying Biocatalysts in Action at X-Ray Free-Electron Lasers," *Nature Methods* **14**(4), 443–49. DOI:10.1038/nmeth.4195.

Galanie, S., et al. 2015. "Complete Biosynthesis of Opioids in Yeast," *Science* **349**(6252), 1095–100. DOI:10.1126/science.aac9373.

Galkin, V. E., et al. 2014. "Near-Atomic Resolution for One State of F-Actin," *Structure* **23**(1), 173–82. DOI:10.1016/j.str.2014.11.006.

Gallagher, R. R., et al. 2015. "Multilayered Genetic Safeguards Limit Growth of Microorganisms to Defined Environments," *Nucleic Acids Research* **43**(3), 1945–54. DOI:10.1093/nar/gku1378.

Gebreselassie, N., and M. R. Antoniewicz. 2014. "^{13}C Metabolic Flux Analysis of Co-Culture Systems: A Novel Approach." In proceedings of *Metabolic Engineering X*, June 15–19, 2014. Vancouver, Canada. American Institute of Chemical Engineers, New York, NY. Printed from e-media with permission by Curran Associates, Inc., Red Hook, NY.

Geddes, B. A., et al. 2015. "Use of Plant Colonizing Bacteria as Chassis for Transfer of N_2-Fixation to Cereals," *Current Opinion in Biotechnology* **32**, 216–22. DOI:10.1016/j.copbio.2015.01.004.

Ghosh, A., et al. 2014. "A Peptide-Based Method for ^{13}C Metabolic Flux Analysis in Microbial Communities," *PLOS Computational Biology* **10**(9), e1003827. DOI:10.1371/journal.pcbi.1003827.

Gibson, D. G., et al. 2010. "Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome," *Science* **329**(5987), 52–56. DOI:10.1126/science.1190719.

Ginn, H. M., et al. 2016. "On the Release of *cppxfel* for Processing X-Ray Free-Electron Laser Images," *Journal of Applied Crystallography* **49**(Pt 3), 1065–72. DOI:10.1107/S1600576716006981.

Glaeser, R. M. 2016. "How Good Can Cryo-EM Become?," *Nature Methods* **13**(1), 28–32. DOI:10.1038/nmeth.3695.

Glukhova, V. A., et al. 2013. "Rapid Assessment of RNAi-Mediated Protein Depletion by Selected Reaction Monitoring Mass Spectrometry," *Journal of Proteome Research* **12**(7), 3246–54. DOI:10.1021/pr400067k.

Guerrero-Ferreira, R. C., and E. R. Wright. 2014. "Zernike Phase Contrast Cryo-Electron Tomography of Whole Bacterial Cells," *Journal of Structural Biology* **185**(1), 129–33. DOI:10.1016/j.jsb.2013.09.011.

Guerrero-Ferreira, R. C., et al. 2011. "Alternative Mechanism for Bacteriophage Adsorption to the Motile Bacterium *Caulobacter crescentus*," *Proceedings of the National Academy of Sciences of the USA* **108**(24), 9963–68. DOI:10.1073/pnas.1012388108.

Ha, M. A., et al. 1998. "Fine Structure in Cellulose Microfibrils: NMR Evidence from Onion and Quince," *The Plant Journal* **16**(2), 183–90. DOI:10.1046/j.1365-313x.1998.00291.x.

Hackett, S. R., et al. 2016. "Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic Flux," *Science* **354**(6311), aaf2786. DOI:10.1126/science.aaf2786.

Hagerty, S. B., et al. 2014. "Accelerated Microbial Turnover but Constant Growth Efficiency with Warming in Soil," *Nature Climate Change* **4**(10), 903–06. DOI:10.1038/Nclimate2361.

Haichar, F. e. Z., et al. 2012. "Stable Isotope Probing of Bacterial Community Structure and Gene Expression in the Rhizosphere of *Arabidopsis thaliana*," *FEMS Microbiology Ecology* **81**(2), 291–302. DOI:10.1111/j.1574-6941.2012.01345.x.

Haimovich, A. D., et al. 2015. "Genomes by Design," *Nature Reviews Genetics* **16**(9), 501–16. DOI:10.1038/nrg3956.

Hammel, M., et al. 2016. "HU Multimerization Shift Controls Nucleoid Compaction," *Science Advances* **2**(7), e1600650. DOI:10.1126/sciadv.1600650.

Hampton, C. M., et al. 2017. "Correlated Fluorescence Microscopy and Cryo-Electron Tomography of Virus-Infected or Transfected Mammalian Cells," *Nature Protocols* **12**(1), 150–67. DOI:10.1038/nprot.2016.168.

Hazen, T. C., et al. 2010. "Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria," *Science* **330**(6001), 204–08. DOI:10.1126/science.1195979.

Henderson, G. P., et al. 2007. "3-D Ultrastructure of *O. tauri*: Electron Cryotomography of an Entire Eukaryotic Cell," *PLOS ONE* **2**(8), e749. DOI:10.1371/journal.pone.0000749.

Henry, C. S., et al. 2016. "Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction," *Journal of Cellular Physiology* **231**(11), 2339–45. DOI:10.1002/jcp.25428.

Heredia-Guerrero, J. A., et al. 2014. "Infrared and Raman Spectroscopic Features of Plant Cuticles: A Review," *Frontiers in Plant Science* **5**, 305. DOI:10.3389/fpls.2014.00305.

Hernandez, M., et al. 2015. "Different Bacterial Populations Associated with the Roots and Rhizosphere of Rice Incorporate Plant-Derived Carbon," *Applied Environmental Microbiology* **81**(6), 2244–53. DOI:10.1128/AEM.03209–14.

Heymann, J. A., et al. 2006. "Site-Specific 3D Imaging of Cells and Tissues with a Dual Beam Microscope," *Journal of Structural Biology* **155**(1), 63–73. DOI:10.1016/j.jsb.2006.03.006.

Hicks, N., et al. 2017. "Using Prokaryotes for Carbon Capture Storage," *Trends in Biotechnology* **35**(1), 22–32. DOI:10.1016/j.tibtech.2016.06.011.

Holman, H.-Y. N., et al. 1999. "Real-Time Characterization of Biogeochemical Reduction of Cr(VI) on Basalt Surfaces by SR-FTIR Imaging," *Geomicrobiology Journal* **16**(4), 307–24. DOI:10.1080/014904599270569.

Holman, H. Y., et al. 2010. "Synchrotron IR Spectromicroscopy: Chemistry of Living Cells," *Analytical Chemistry* **82**(21), 8757–65. DOI:10.1021/ac100991d.

Holman, H. Y. N., et al. 2009a. "Real-Time Chemical Imaging of Bacterial Activity in Biofilms Using Open-Channel Microfluidics and Synchrotron FTIR Spectromicroscopy," *Analytical Chemistry* **81**(20), 8564–70. DOI:10.1021/ac9015424.

Holman, H. Y., et al. 2009b. "Real-Time Molecular Monitoring of Chemical Environment in Obligate Anaerobes During Oxygen Adaptive Response," *Proceedings of the National Academy of Sciences of the USA* **106**(31), 12599–604. DOI:10.1073/pnas.0902070106.

Holman, H. Y., et al. 2002a. "Synchrotron Infrared Spectromicroscopy as a Novel Bioanalytical Microprobe for Individual Living Cells: Cytotoxicity Considerations," *Journal of Biomedical Optics* **7**(3), 417–24. DOI:10.1117/1.1485299.

Holman, H. Y. N., et al. 2002b. "Catalysis of PAH Biodegradation by Humic Acid Shown in Synchrotron Infrared Studies," *Environmental Science & Technology* **36**(6), 1276–80. DOI:10.1021/es0157200.

Hom, E. F. Y., et al. "A Chemical Perspective on Microalgal-Microbial Interactions," *Trends in Plant Science* **20**(11), 689–93. DOI:10.1016/j.tplants.2015.09.004.

Hryc, C. F., et al. 2017. "Accurate Model Annotation of a Near-Atomic Resolution Cryo-EM Map," *Proceedings of the National Academy of Sciences of the USA* **114**(12), 3103–08. DOI:10.1073/pnas.1621152114.

Hu, B., et al. 2013. "The Bacteriophage T7 Virion Undergoes Extensive Structural Remodeling During Infection," *Science* **339**(6119), 576–9. DOI:10.1126/science.1231887.

Hu, P., et al. 2016. "Integrated Bioprocess for Conversion of Gaseous Substrates to Liquids," *Proceedings of the National Academy of Sciences of the USA* **113**(14), 3773–78. DOI:10.1073/pnas.1516867113.

Huang, F., et al. 2016. "Ultra-High Resolution 3D Imaging of Whole Cells," *Cell* **166**(4), 1028–40. DOI:10.1016/j.cell.2016.06.016.

Hungate, B. A., et al. 2015. "Quantitative Microbial Ecology through Stable Isotope Probing," *Applied Environmental Microbiology* **81**(21), 7570–81. DOI:10.1128/AEM.02280–15.

Hutchison, C. A., et al. 2016. "Design and Synthesis of a Minimal Bacterial Genome," *Science* **351**(6280), aad6253. DOI:10.1126/science.aad6253.

Isaacs, F. J., et al. 2011. "Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement," *Science* **333**(6040), 348–53. DOI:10.1126/science.1205822.

Jakobsen, I., and L. Rosendahl. 1990. "Carbon Flow into Soil and External Hyphae from Roots of Mycorrhizal Cucumber Plants," *New Phytologist* **115**(1), 77–83. DOI:DOI 10.1111/j.1469-8137.1990.tb00924.x.

Jastrow, J. D., and R . M. Miller. 1998. "Soil Aggregate Stabilization and Carbon Sequestration: Feedbacks Through Organomineral Associations." *In Soil Processes and the Carbon Cycle*. 207–23. Eds. R . Lal, R . F. Follett, and B. A. Stewart. CRC Press LLC, Boca Raton, Florida.

Jain, A., et al. 2012. "Single-Molecule Pull-Down for Studying Protein Interactions," *Nature Protocols* **7**(3), 445–52. DOI:10.1038/nprot.2011.452.

Jayawickreme, D. H., et al. 2014. "Geophysical Subsurface Imaging for Ecological Applications," *New Phytologist* **201**(4), 1170–75. DOI:10.1111/nph.12619.

Jehmlich, N., et al. 2010. "Protein-Based Stable Isotope Probing," *Nature Protocols* **5**(12), 1957–66. DOI:10.1038/nprot.2010.166.

Jensen, L. B., et al. 1993. "A Substrate-Dependent Biological Containment System for *Pseudomonas putida* Based on the *Escherichia coli gef* Gene," *Applied Environmental Microbiology* **59**(11), 3713–17.

Jinek, M., et al. 2013. "RNA-Programmed Genome Editing in Human Cells," *eLife* **2**, e00471. DOI:10.7554/eLife.00471.

Jinek, M., et al. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science* **337**(6096), 816–21. DOI:10.1126/science.1225829.

Johnson, G. T., et al. 2015. "Cellpack: A Virtual Mesoscope to Model and Visualize Structural Systems Biology," *Nature Methods* **12**(1), 85–91. DOI:10.1038/nmeth.3204.

Johnson, S. N., and U. N. Nielsen. 2012. "Foraging in the Dark – Chemically Mediated Host Plant Location by Belowground Insect Herbivores," *Journal of Chemical Ecology* **38**(6), 604–14. DOI:10.1007/s10886-012-0106-x.

Jones, A. M., et al. 2014. "Abscisic Acid Dynamics in Roots Detected with Genetically Encoded FRET Sensors," *eLife* **3**, e01741. DOI:10.7554/eLife.01741.

Jones, A. M., et al. 2013. "*In Vivo* Biochemistry: Applications for Small Molecule Biosensors in Plant Biology," *Current Opinion in Plant Biology* **16**(3), 389–95. DOI:10.1016/j.pbi.2013.02.010.

Kaiser, C., et al. 2015. "Exploring the Transfer of Recent Plant Photosynthates to Soil Microbes: Mycorrhizal Pathway vs Direct Root Exudation," *New Phytologist* **205**(4), 1537–51. DOI:10.1111/nph.13138.

Kalluri, U. C., et al. 2014. "Systems and Synthetic Biology Approaches to Alter Plant Cell Walls and Reduce Biomass Recalcitrance," *Plant Biotechnology Journal* **12**(9), 1207–16. DOI:10.1111/pbi.12283.

Karl, D. M., et al. 2012. "Predictable and Efficient Carbon Sequestration in the North Pacific Ocean Supported by Symbiotic Nitrogen Fixation," *Proceedings of the National Academy of Sciences of the USA* **109**(6), 1842–9. DOI:10.1073/pnas.1120312109.

Karunatilaka, K. S., et al. 2014. "Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts," *MBio* **5**(6), e02172. DOI:10.1128/mBio.02172–14.

Keedy, D. A., et al. 2015a. "Exposing Hidden Alternative Backbone Conformations in X-Ray Crystallography Using qFit," *PLOS Computational Biology* **11**(10), e1004507. DOI:10.1371/journal.pcbi.1004507.

Keedy, D. A., et al. 2015b. "Mapping the Conformational Landscape of a Dynamic Enzyme by Multitemperature and XFEL Crystallography," *eLife* **4**, e07574. DOI:10.7554/eLife.07574.

Keiluweit, M., et al. 2015. "Mineral Protection of Soil Carbon Counteracted by Root Exudates," *Nature Climate Change* **5**(6), 588–95. DOI:10.1038/nclimate2580.

Keiluweit, M., et al. 2012. "Nano-Scale Investigation of the Association of Microbial Nitrogen Residues with Iron (Hydr)oxides in a Forest Soil O-Horizon," *Geochimica et Cosmochimica Acta* **95**, 213–26. DOI:10.1016/j.gca.2012.07.001.

Kelley, S. S., et al. 2002. "Use of NIR and Pyrolysis–MBMS Coupled with Multivariate Analysis for Detecting the Chemical Changes Associated with Brown-Rot Biodegradation of Spruce Wood," *FEMS Microbiology Letters* **209**(1), 107–11. DOI:10.1111/j.1574-6968.2002.tb11117.x.

Kemner, K. M., et al. 1998. "Using X-Ray Microprobes for Environmental Research," *Proceedings of the Society of Photo Optical Instrumentation Engineers* **3449**, 45–54. DOI:10.1117/12.330354.

Kern, J., et al. 2013. "Simultaneous Femtosecond X-Ray Spectroscopy and Diffraction of Photosystem II at Room Temperature," *Science* **340**(6131), 491–5. DOI:10.1126/science.1234273.

Klar, T. A., and S. W. Hell. 1999. "Subdiffraction Resolution in Far-Field Fluorescence Microscopy," *Optics Letters* **24**(14), 954–56. DOI:10.1364/OL.24.000954.

Kockenberger, W., et al. 2004. "High Resolution NMR Microscopy of Plants and Fungi," *Journal of Microscopy* **214**(Pt 2), 182–89. DOI:10.1111/j.0022–2720.2004.01351.x.

Kong, W., et al. 2008. "Regulated Programmed Lysis of Recombinant *Salmonella* in Host Tissues to Release Protective Antigens and Confer Biological Containment," *Proceedings of the National Academy of Sciences of the USA* **105**(27), 9361–66. DOI:10.1073/pnas.0803801105.

Kopf, S. H., et al. 2016. "Trace Incorporation of Heavy Water Reveals Slow and Heterogeneous Pathogen Growth Rates in Cystic Fibrosis Sputum," *Proceedings of the National Academy of Sciences of the USA* **113**(2), E110–16. DOI:10.1073/pnas.1512057112.

Kroon-Batenburg, L. M., et al. 2017. "Raw Diffraction Data Preservation and Reuse: Overview, Update on Practicalities and Metadata Requirements," *International Union of Crystallography* **4**(Pt 1), 87–99. DOI:10.1107/S2052252516018315.

Kupitz, C., et al. 2017. "Structural Enzymology Using X-Ray Free Electron Lasers," *Structural Dynamics* **4**(4), 044003. DOI:10.1063/1.4972069.

Kusano, M., et al. 2015. "Using Metabolomic Approaches to Explore Chemical Diversity in Rice," *Molecular Plant* **8**(1), 58–67. DOI:10.1016/j.molp.2014.11.010.

Langan, P., et al. 2014. "Common Processes Drive the Thermochemical Pretreatment of Lignocellulosic Biomass," *Green Chemistry* **16**(1), 63–68. DOI:10.1039/C3GC41962B.

Lajoie, M. J., et al. 2013. "Genomically Recoded Organisms Expand Biological Functions," *Science* **342**(6156), 357–60. DOI:10.1126/science.1241459.

Lang, B., et al. 2014. "Effects of Environmental Warming and Drought on Size-Structured Soil Food Webs," *Oikos* **123**(10), 1224–33. DOI:10.1111/j.1600-0706.2013.00894.x.

Lang, B., et al. 2012. "Warming Effects on Consumption and Intraspecific Interference Competition Depend on Predator Metabolism," *Journal of Animal Ecology* **81**(3), 516–23. DOI:10.1111/j.1365-2656.2011.01931.x.

Lang, P. T., et al. 2014. "Protein Structural Ensembles Are Revealed by Redefining X-Ray Electron Density Noise," *Proceedings of the National Academy of Sciences of the USA* **111**(1), 237–42. DOI:10.1073/pnas.1302823110.

Larabell, C. A., and K. A. Nugent. 2010. "Imaging Cellular Architecture with X-Rays," *Current Opinion in Structural Biology* **20**(5), 623–31. DOI:10.1016/j.sbi.2010.08.008.

Lau, W., and E. S. Sattely. 2015. "Six Enzymes from Mayapple That Complete the Biosynthetic Pathway to the Etoposide Aglycone," *Science* **349**(6253), 1224–28. DOI:10.1126/science.aac7202.

Lau, W., et al. 2014. "Key Applications of Plant Metabolic Engineering," *PLOS Biology* **12**(6), e1001879. DOI:10.1371/journal.pbio.1001879.

Lawton, J. H., et al. 1993. "The Ecotron: A Controlled Environmental Facility for the Investigation of Population and Ecosystem Processes," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **341**(1296), 181–94. DOI: 10.1098/rstb.1993.0102.

Le Gros, M. A., et al. 2012. Visualizing Sub-Cellular Organization Using Soft X-Ray Tomography. In *Comprehensive Biophysics,* First Ed. 90–110. Ed. H. E. Edward. Elsevier, Amsterdam.

Le, R. K., et al. 2014. "Analysis of the Solution Structure of *Thermosynechococcus elongatus* Photosystem I in *n*-dodecyl-β-D-Maltoside Using Small-Angle Neutron Scattering and Molecular Dynamics Simulation," *Archives Biochemistry Biophysics* **550–551**, 50–57. DOI:10.1016/j.abb.2014.04.005.

Leis, A., et al. 2009. "Visualizing Cells at the Nanoscale," *Trends in Biochemical Sciences* **34**(2), 60–70. DOI:10.1016/j.tibs.2008.10.011.

Levantino, M., et al. 2015. "Using Synchrotrons and XFELS for Time-Resolved X-Ray Crystallography and Solution Scattering Experiments on Biomolecules," *Current Opinion in Structural Biology* **35**, 41–48. DOI:10.1016/j.sbi.2015.07.017.

Li, X., et al. 2013. "Electron Counting and Beam-Induced Motion Correction Enable Near-Atomic-Resolution Single-Particle Cryo-EM," *Nature Methods* **10**(6), 584–90. DOI:10.1038/nmeth.2472.

Li, Z., and G. J. Jensen. 2009. "Electron Cryotomography: A New View into Microbial Ultrastructure," *Current Opinion in Microbiology* **12**(3), 333–40. DOI:10.1016/j.mib.2009.03.007.

Liao, M., et al. 2013. "Structure of the TRPV1 Ion Channel Determined by Electron Cryo-Microscopy," *Nature* **504**(7478), 107–12. DOI:10.1038/nature12822.

Liao, Y., et al. 2016. "Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells," *Biophysical Journal* **111**(12), 2562–69. DOI:10.1016/j.bpj.2016.11.006.

Lomnitz, J. G., and M. A. Savageau. 2016. "Design Space Toolbox V2: Automated Software Enabling a Novel Phenotype-Centric Modeling Strategy for Natural and Synthetic Biological Systems," *Frontiers in Genetics* **7**, article 118. DOI:10.3389/fgene.2016.00118.

Loutherback, K., et al. 2016. "Microfluidic Approaches to Synchrotron Radiation-Based Fourier Transform Infrared (SR-FTIR) Spectral Microscopy of Living Biosystems," *Protein and Peptide Letters* **23**(3), 273–82. DOI:10.2174/0929866523666160106154035.

Lovley, D. R. 2000. "Chapter 1: Fe(III) and Mn(IV) Reduction." In *Environmental Microbe-Metal Interactions*, 3–30. Ed. D. R. Lovley. American Society of Microbiology, Washington, D.C. Book DOI:10.1128/9781555818098; Chapter DOI:10.1128/9781555818098.ch1.

Lu, P., et al. 2014. "Three-Dimensional Structure of Human Gamma-Secretase," *Nature* **512**(7513), 166–70. DOI:10.1038/nature13567.

Ludtke, S. J. 2016. "Single-Particle Refinement and Variability Analysis in Eman2.1," *Methods in Enzymology* **579**, 159–89. DOI:10.1016/bs.mie.2016.05.001.

Lyubimov, A. Y., et al. 2016. "IOTA: Integration Optimization, Triage and Analysis Tool for the Processing of XFEL Diffraction Images," *Journal of Applied Crystallography* **49**(Pt 3), 1057–64. DOI:10.1107/S1600576716006683.

Lyumkis, D., et al. 2013. "Optimod – An Automated Approach for Constructing and Optimizing Initial Models for Single-Particle Electron Microscopy," *Journal of Structural Biology* **184**(3), 417–26. DOI:10.1016/j.jsb.2013.10.009.

Ma, N. J., and F. J. Isaacs. 2016. "Genomic Recoding Broadly Obstructs the Propagation of Horizontally Transferred Genetic Elements," *Cell Systems* **3**(2), 199–207. DOI:10.1016/j.cels.2016.06.009.

MacLean, B., et al. 2010. "Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments," *Bioinformatics* **26**(7), 966–8. DOI:10.1093/bioinformatics/btq054.

Maeght, J.-L., et al. 2013. "How to Study Deep Roots—and Why It Matters," *Frontiers in Plant Science* **4**, article 299. DOI:10.3389/fpls.2013.00299.

Mahamid, J., et al. 2015. "A Focused Ion Beam Milling and Lift-Out Approach for Site-Specific Preparation of Frozen-Hydrated Lamellas from Multicellular Organisms," *Journal of Structural Biology* **192**(2), 262–9. DOI:10.1016/j.jsb.2015.07.012.

Mahamid, J., et al. 2016. "Visualizing the Molecular Sociology at the HeLa Cell Nuclear Periphery," *Science* **351**(6276), 969–72. DOI:10.1126/science.aad8857.

Maia, F. R., and J. Hajdu. 2016. "The Trickle before the Torrent-Diffraction Data from X-Ray Lasers," *Scientific Data* **3**, 160059. DOI:10.1038/sdata.2016.59.

Mairhofer, S., et al. 2015. "Extracting Multiple Interacting Root Systems Using X-Ray Microcomputed Tomography," *The Plant Journal* **84**(5), 1034–43. DOI:10.1111/tpj.13047.

Mairhofer, S., et al. 2013. "Recovering Complete Plant Root System Architectures from Soil via X-Ray μ-Computed Tomography," *Plant Methods* **9**(1), article 8, 1–7. DOI:10.1186/1746-4811-9-8.

Mali, P., et al. 2013. "RNA-Guided Human Genome Engineering via Cas9," *Science* **339**(6121), 823–26. DOI:10.1126/science.1232033.

Malmstrom, R. D., et al. 2015. "Allostery Through the Computational Microscope: Camp Activation of a Canonical Signalling Domain," *Nature Communications* **6**, article 7588. DOI:10.1038/ncomms8588.

Mandell, D. J., et al. 2015. "Biocontainment of Genetically Modified Organisms by Synthetic Protein Design," *Nature* **518**(7537), 55–60. DOI:10.1038/nature14121.

Mandy, D. E., et al. 2014. "Metabolic Flux Analysis Using ^{13}C Peptide Label Measurements," *The Plant Journal* **77**(3), 476–86. DOI:10.1111/tpj.12390.

Marchetti, R., et al. 2016. "NMR Analysis of the Binding Mode of Two Fungal *endo*-β-1,4-Mannanases from GH5 and GH26 Families," *Organic & Biomolecular Chemistry* **14**(1), 314–22. DOI:10.1039/C5OB01851J.

Marko, M., et al. 2007. "Focused-Ion-Beam Thinning of Frozen-Hydrated Biological Specimens for Cryo-Electron Microscopy," *Nature Methods* **4**(3), 215–17. DOI:10.1038/nmeth1014.

Martin, M. C., et al. 2013. "3D Spectral Imaging with Synchrotron Fourier Transform Infrared Spectro-Microtomography," *Nature Methods* **10**(9), 861–64. DOI:10.1038/nmeth.2596.

Mason, O. U., et al. 2012. "Metagenome, Metatranscriptome and Single-Cell Sequencing Reveal Microbial Response to Deepwater Horizon Oil Spill," *The ISME Journal* **6**(9), 1715–27. DOI:10.1038/ismej.2012.59.

Mattei, S., et al. 2016. "The Structure and Flexibility of Conical HIV-1 Capsids Determined Within Intact Virions," *Science* **354**(6318), 1434–37. DOI:10.1126/science.aah4972.

Mayali, X., et al. 2012. "High-Throughput Isotopic Analysis of RNA Microarrays to Quantify Microbial Resource Use," *The ISME Journal* **6**(6), 1210–21. DOI:10.1038/ismej.2011.175.

McDermott, G., et al. 2012. "Visualizing and Quantifying Cell Phenotype Using Soft X-Ray Tomography," *Bioessays* **34**(4), 320–7. DOI:10.1002/bies.201100125.

McFarlane, H. E., et al. 2014. "The Cell Biology of Cellulose Synthesis," *Annual Review of Plant Biology* **65**, 69–94. DOI:10.1146/annurev-arplant-050213-040240.

Meng, Y., et al. 2016. "Transition Path Theory Analysis of c-Src Kinase Activation," *Proceedings of the National Academy of Sciences of the USA* **113**(33), 9193–98. DOI:10.1073/pnas.1602790113.

Meyer, P. A., et al. 2016. "Data Publication with the Structural Biology Data Grid Supports Live Analysis," *Nature Communications* **7**, article 10882. DOI:10.1038/ncomms10882.

Mih, N., et al. 2016. "A Multi-Scale Computational Platform to Mechanistically Assess the Effect of Genetic Variation on Drug Responses in Human Erythrocyte Metabolism," *PLOS Computational Biology* **12**(7), e1005039. DOI:10.1371/journal.pcbi.1005039.

Minty, J. J., et al. 2013. "Design and Characterization of Synthetic Fungal-Bacterial Consortia for Direct Production of Isobutanol from Cellulosic Biomass," *Proceedings of the National Academy of Sciences of the USA* **110**(36), 14592–97. DOI:10.1073/pnas.1218447110.

Morris, B. E., et al. 2013. "Microbial Syntrophy: Interaction for the Common Good," *FEMS Microbiology Reviews* **37**(3), 384–406. DOI:10.1111/1574-6976.12019.

Moult, J., et al. 2014. "Critical Assessment of Methods of Protein Structure Prediction (CASP)––Round X," *Proteins* **82**(Suppl 2), 1–6. DOI:10.1002/prot.24452.

Moussaieff, A., et al. 2013. "High-Resolution Metabolic Mapping of Cell Types in Plant Roots," *Proceedings of the National Academy of Sciences of the USA* **110**(13), E1232–41. DOI:10.1073/pnas.1302019110.

Moyes, A. B., et al. 2016. "Evidence for Foliar Endophytic Nitrogen Fixation in a Widely Distributed Subalpine Conifer," *New Phytologist* **210**(2), 657–68. DOI:10.1111/nph.13850.

Mueller, L. A., et al. 2003. "AraCyc: A Biochemical Pathway Database for Arabidopsis," *Plant Physiology* **132**(2), 453–60. DOI:10.1104/pp.102.017236.

Muller, E. A., et al. 2015. "Infrared Chemical NanoImaging: Accessing Structure, Coupling, and Dynamics on Molecular Length Scales," *Journal of Physical Chemistry Letters* **6**(7), 1275–84. DOI:10.1021/acs.jpclett.5b00108.

Musat, N., et al. 2016. "Tracking Microbial Interactions with NanoSIMS," *Current Opinion in Biotechnology* **41**, 114–21. DOI:10.1016/j.copbio.2016.06.007.

Nargund, S., et al. 2014. "Flux and Reflux: Metabolite Reflux in Plant Suspension Cells and Its Implications on Isotope-Assisted Metabolic Flux Analysis," *Molecular BioSystems* **10**, 1496–508. DOI:10.1039/C3MB70348G.

Nealson, K. H., and D. Saffarini. 1994. "Iron and Manganese in Anaerobic Respiration: Environmental Significance, Physiology, and Regulation," *Annual Review of Microbiology* **48**, 311–43. DOI:10.1146/annurev.mi.48.100194.001523.

Neufeld, J. D., et al. 2007. "Who Eats What, Where and When? Isotope-Labelling Experiments Are Coming of Age," *The ISME Journal* **1**(2), 103–10. DOI:10.1038/ismej.2007.30.

Nevin, K. P., et al. 2011. "Electrosynthesis of Organic Compounds from Carbon Dioxide Is Catalyzed by a Diversity of Acetogenic Microorganisms," *Applied and Environmental Microbiology* **77**(9), 2882–86. DOI:10.1128/AEM.02642–10.

Nguyen, T. H., et al. 2016. "Cryo-EM Structure of the Yeast U4/U6.U5 Tri–snRNP at 3.7 Å Resolution," *Nature* **530**(7590), 298–302. DOI:10.1038/nature16940.

Northen, T. R., et al. 2008. "A Nanostructure-Initiator Mass Spectrometry-Based Enzyme Activity Assay," *Proceedings of the National Academy of Sciences of the USA* **105**(10), 3678–83. DOI:10.1073/pnas.0712332105.

Nuccio, E. E., et al. 2013. "An Arbuscular Mycorrhizal Fungus Significantly Modifies the Soil Bacterial Community and Nitrogen Cycling During Litter Decomposition," *Environmental Microbiology* **15**(6), 1870–81. DOI:10.1111/1462-2920.12081.

Nunan, N., et al. 2006. "Investigating Microbial Micro-Habitat Structure Using X-Ray Computed Tomography," *Geoderma* **133**(3–4), 398–407. DOI:10.1016/j.geoderma.2005.08.004.

O'Brien, J. T., et al. 2015. "Ambient Infrared Laser Ablation Mass Spectrometry (AIRLAB-MS) of Live Plant Tissue with Plume Capture by Continuous Flow Solvent Probe," *Analytical Chemistry* **87**(5), 2631–38. DOI:10.1021/ac503383p.

O'Connor, S. E. 2015. "Engineering of Secondary Metabolism," *Annual Review of Genetics* **49**, 71–94. DOI:10.1146/annurev-genet-120213-092053.

O'Dell, W. B., et al. 2017. "Oxygen Activation at the Active Site of a Fungal Lytic Polysaccharide Monooxygenase," *Angewandte Chemie International Edition in English* **56**(3), 767–70. DOI:10.1002/anie.201610502.

Oberortner, E., and D. Densmore. 2015. "Web-Based Software Tool for Constraint-Based Design Specification of Synthetic Biological Systems," *ACS Synthetic Biology* **4**(6), 757–60. DOI:10.1021/sb500352b.

Oburger, E., and H. Schmidt. "New Methods to Unravel Rhizosphere Processes," *Trends in Plant Science* **21**(3), 243–55. DOI:10.1016/j.tplants.2015.12.005.

Oghbaey, S., et al. 2016. "Fixed Target Combined with Spectral Mapping: Approaching 100% Hit Rates for Serial Crystallography," *Acta Crystallographica Section D: Structural Biology* **72**(Pt 8), 944–55. DOI:10.1107/S2059798316010834.

Ohad, I., and D. Danon. 1964. "On the Dimensions of Cellulose Microfibrils," *Journal of Cell Biology* **22**(1), 302–05.

Ohad, I., et al. 1962. "Synthesis of Cellulose by *Acetobacter xylinum*: V. Ultrastructure of Polymer," *Journal of Cell Biology* **12**(1), 31–46.

Okumoto, S., et al. 2012. "Quantitative Imaging with Fluorescent Biosensors," *Annual Review of Plant Biology* **63**, 663–706. DOI:10.1146/annurev-arplant-042110-103745.

Orth, J. D., et al. 2010. "What Is Flux Balance Analysis?" *Nature Biotechnology* **28**(3), 245–48. DOI:10.1038/nbt.1614.

Ostrov, N., et al. 2016. "Design, Synthesis, and Testing Toward a 57-Codon Genome," *Science* **353**(6301), 819–22. DOI:10.1126/science.aaf3639.

Paddon, C. J., and J. D. Keasling. 2014. "Semi-Synthetic Artemisinin: A Model for the Use of Synthetic Biology in Pharmaceutical Development," *Nature Reviews Microbiology* **12**(5), 355–67. DOI:10.1038/nrmicro3240.

Paddon, C. J., et al. 2013. "High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin," *Nature* **496**(7446), 528–32. DOI:10.1038/nature12051.

Palma, A. S., et al. 2015. "Unravelling Glucan Recognition Systems by Glycome Microarrays Using the Designer Approach and Mass Spectrometry," *Molecular & Cellular Proteomics* **14**(4), 974–88. DOI:10.1074/mcp.M115.048272.

Park, S. Y., et al. 2009. "Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins," *Science* **324**(5930), 1068–71. DOI:10.1126/science.1173041.

Parkinson, D. Y., et al. 2013. "Nanoimaging Cells Using Soft X-Ray Tomography," *Methods in Molecular Biology* **950**, 457–81. DOI:10.1007/978-1-62703-137-0_25.

Petersen, B. O., et al. 2014. "Probing the Structural Details of Xylan Degradation by Real-Time NMR Spectroscopy," *Carbohydrate Polymers* **112**, 587–94. DOI:10.1016/j.carbpol.2014.06.049.

Pett-Ridge, J., and P. K. Weber. 2012. "NanoSIP: NanoSIMS Applications for Microbial Biology," *Methods in Molecular Biology* **881**, 375–408. DOI:10.1007/978–1–61779–827–6_13.

Pett-Ridge, J., and M. K. Firestone. 2017. "Using Stable Isotopes to Explore Root--Metal-Mineral Interactions in Soil," *Rhizosphere* **3**(Pt 2), 244–53. DOI:10.1016/j.rhisph.2017.04.016.

Petzold, C. J., et al. 2015. "Analytics for Metabolic Engineering," *Frontiers in Bioengineering and Biotechnology* **3**, article 135. DOI:10.3389/fbioe.2015.00135.

Phillips, J. C., et al. 2005. "Scalable Molecular Dynamics with NAMD," *Journal of Computational Chemistry* **26**(16), 1781–802. DOI:10.1002/jcc.20289.

Pieper, D. H., and W. Reineke. 2000. "Engineering Bacteria for Bioremediation," *Current Opinion in Biotechnology* **11**(3), 262–70.

Pingali, S. V., et al. 2014. "Morphological Changes in the Cellulose and Lignin Components of Biomass Occur at Different Stages During Steam Pretreatment," *Cellulose* **21**(2), 873–78. DOI:10.1007/s10570–013–0162–6.

Plancot, B., et al. 2014. "Structural Characterization of Arabinoxylans from Two African Plant Species *Eragrostis nindensis* and *Eragrostis tef* Using Various Mass Spectrometric Methods," *Rapid Communications in Mass Spectrometry* **28**(8), 908–16. DOI:10.1002/rcm.6859.

Plattner, N., and F. Noe. 2015. "Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models," *Nature Communications* **6**, article 7653. DOI:10.1038/ncomms8653.

Postma, J. A., and J. P. Lynch. 2011. "Root Cortical Aerenchyma Enhances the Growth of Maize on Soils with Suboptimal Availability of Nitrogen, Phosphorus, and Potassium," *Plant Physiology* **156**(3), 1190–201. DOI:10.1104/pp.111.175489.

Pound, M. P., et al. 2013. "RootNav: Navigating Images of Complex Root Architectures," *Plant Physiology* **162**(4), 1802–14. DOI:10.1104/pp.113.221531.

Probst, A. J., et al. 2013. "Tackling the Minority: Sulfate-Reducing Bacteria in an Archaea-Dominated Subsurface Biofilm," *The ISME Journal* **7**(3), 635–51. DOI:10.1038/ismej.2012.133.

Qi, Lei S., et al. 2013. "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," *Cell* **152**(5), 1173–83. DOI:10.1016/j.cell.2013.02.022.

Quemener, B., et al. 2015. "Negative Electrospray Ionization Mass Spectrometry: A Method for Sequencing and Determining Linkage Position in Oligosaccharides from Branched Hemicelluloses," *Journal of Mass Spectrometry* **50**(1), 247–64. DOI:10.1002/jms.3528.

Rabaey, K., and R. A. Rozendal. 2010. "Microbial Electrosynthesis — Revisiting the Electrical Route for Microbial Production," *Nature Reviews Microbiology* **8**(10), 706–16. DOI:10.1038/nrmicro2422.

Radajewski, S., et al. 2000. "Stable-Isotope Probing as a Tool in Microbial Ecology," *Nature* **403**(6770), 646–49. DOI:10.1038/35001054.

Ranganathan, S., et al. 2010. "OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions," *PLOS Computational Biology* **6**(4), e1000744. DOI:10.1371/journal.pcbi.1000744.

Ranganathan, S., et al. 2012. "An Integrated Computational and Experimental Study for Overproducing Fatty Acids in *Escherichia coli*," *Metabolic Engineering* **14**(6), 687–704. DOI:10.1016/j.ymben.2012.08.008.

Rellán-Álvarez, R., et al. 2016. "Environmental Control of Root System Biology," *Annual Review of Plant Biology* **67**, 619–42. DOI:10.1146/annurev-arplant-043015-111848.

Riede, J. O., et al. 2010. "Scaling of Food-Web Properties with Diversity and Complexity Across Ecosystems." In *Advances in Ecological Research; Ecological Networks, Vol.42*. 139–70. Ed. G. Woodward. Elsevier Ltd. Academic Press, London and Amsterdam.

Roden, E. E., and R. G. Wetzel. 1996. "Organic Carbon Oxidation and Suppression of Methane Production by Microbial Fe(III) Oxide Reduction in Vegetated and Unvegetated Freshwater Wetland Sediments," *Limnology and Oceanography* **41**(8), 1733–48. DOI:10.4319/lo.1996.41.8.1733.

Rodriguez-Escamilla, Z., et al. 2016. "Epigenetics Knocks on Synthetic Biology's Door," *FEMS Microbiology Letters* **363**(17). DOI:10.1093/femsle/fnw191.

Rogers, E. D., et al. 2012. "Cell Type-Specific Transcriptional Profiling: Implications for Metabolite Profiling," *Plant Journal* **70**(1), 5–17. DOI:10.1111/j.1365-313X.2012.04888.x.

Rogers, E. D., et al. 2016. "X-Ray Computed Tomography Reveals the Response of Root System Architecture to Soil Texture," *Plant Physiology* **171**(3), 2028–40. DOI:10.1104/pp.16.00397.

Ronchel, M. C., and J. L. Ramos. 2001. "Dual System to Reinforce Biological Containment of Recombinant Bacteria Designed for Rhizoremediation," *Applied Environmental Microbiology* **67**(6), 2649–56. DOI:10.1128/AEM.67.6.2649-2656.2001.

Rovner, A. J., et al. 2015. "Recoded Organisms Engineered to Depend on Synthetic Amino Acids," *Nature* **518**(7537), 89–93. DOI:10.1038/nature14095.

Russel, D., et al. 2012. "Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies," *PLOS Biology* **10**(1), e1001244. DOI:10.1371/journal.pbio.1001244.

Rust, M. J., et al. 2006. "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)," *Nature Methods* **3**(10), 793–95. DOI:10.1038/nmeth929.

Saldin, D. K., et al. 2011. "Reconstructing an Icosahedral Virus from Single-Particle Diffraction Experiments," *Optics Express* **19**(18), 17318–35. DOI:10.1364/OE.19.017318.

Sali, A., et al. 2015. "Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop," *Structure* **23**(7), 1156–67. DOI:10.1016/j.str.2015.05.013.

Sargent, F., et al. 2013. "A Synthetic System for Expression of Components of a Bacterial Microcompartment," *Microbiology* **159**(11), 2427–36. DOI:10.1099/mic.0.069922-0.

Scheres, S. H. 2016. "Chapter Six – Processing of Structurally Heterogeneous Cryo-EM Data in RELION," *Methods in Enzymology* **579**, 125–57. DOI:10.1016/bs.mie.2016.04.012.

Schmidt, M., et al. 2013. "Protein Energy Landscapes Determined by Five-Dimensional Crystallography," *Acta Crystallographica Section D Biological Crystallography* **D69**(Pt 12), 2534–42. DOI:10.1107/S0907444913025997.

Schreiber, J., et al. 2016. "Model-Guided Combinatorial Optimization of Complex Synthetic Gene Networks," *Molecular Systems Biology* **12**(12), article 899. DOI:10.15252/msb.20167265.

Sene, C. F. B., et al. 1994. "Fourier-Transform Raman and Fourier-Transform Infrared-Spectroscopy – An Investigation of Five Higher-Plant Cell-Walls and Their Components," *Plant Physiology* **106**(4), 1623–31. DOI:10.1104/pp.106.4.1623.

Shi, S., et al. 2015. "Successional Trajectories of Rhizosphere Bacterial Communities over Consecutive Seasons," *MBio* **6**(4), e00746-15. DOI:10.1128/mBio.00746-15.

Sluiter, A., et al., 2013. *Chapter 8: Methods for Biomass Compositional Analysis*. National Renewable Energy Laboratory (NREL). Ed. A. Sluiter et al. Golden, Colo.

Smanski, M. J., et al. 2014. "Functional Optimization of Gene Clusters by Combinatorial Design and Assembly," *Nature Biotechnology* **32**(12), 1241–49. DOI:10.1038/nbt.3063.

Smith, E. A., et al. 2014. "Quantitatively Imaging Chromosomes by Correlated Cryo-Fluorescence and Soft X-Ray Tomographies," *Biophysical Journal* **107**(8), 1988–96. DOI:10.1016/j.bpj.2014.09.011.

Solomon, K., et al. 2014. "Extracting Data from the Muck: Deriving Biological Insight from Complex Microbial Communities and Non-Model Organisms with Next Generation Sequencing," *Current Opinion in Biotechnology* **28**, 103–10. DOI: 10.1016/j.copbio.2014.01.007.

Solomon, K. V., et al. 2016. "Early-Branching Gut Fungi Possess a Large, Comprehensive Array of Biomass-Degrading Enzymes," *Science* **351**(6278), 1192–95. DOI:10.1126/science.aad1431.

Song, C., et al. 2015. "Molecular and Chemical Dialogues in Bacteria-Protozoa Interactions," *Scientific Reports* **5**, article 12837. DOI:10.1038/srep12837.

Spence, J. C., and P. W. Hawkes. 2008. "Diffract-and-Destroy: Can X-Ray Lasers 'Solve' the Radiation Damage Problem?" *Ultramicroscopy* **108**(12), 1502–03. DOI:10.1016/j.ultramic.2008.05.003.

Steidler, L. 2003. "Genetically Engineered Probiotics," *Best Practice & Research Clinical Gastroenterology* **17**(5), 861–76. DOI:10.1016/S1521-6918(03)00072-6.

Steidler, L., et al. 2003. "Biological Containment of Genetically Modified *Lactococcus lactis* for Intestinal Delivery of Human Interleukin 10," *Nature Biotechnology* **21**(7), 785–89. DOI:10.1038/nbt840.

Steinhauser, M. L., and C. P. Lechene. 2013. "Quantitative Imaging of Subcellular Metabolism with Stable Isotopes and Multi-Isotope Imaging Mass Spectrometry," *Seminars in Cell & Developmental Biology* **24**(8–9), 661–67. DOI:10.1016/j.semcdb.2013.05.001.

Stewart, B. J., et al. 2010. "Yeast Dynamic Metabolic Flux Measurement in Nutrient-Rich Media by HPLC and Accelerator Mass Spectrometry," *Analytical Chemistry* **82**(23), 9812–17. DOI:10.1021/ac102065f.

Sturcova, A., et al. 2004. "Structural Details of Crystalline Cellulose from Higher Plants," *Biomacromolecules* **5**(4), 1333–39. DOI:10.1021/bm034517p.

Sugiyama, M., et al. 1998. "Small-Angle Neutron Scattering Observation of Aqueous Suspension of Microcrystalline Cellulose," *Japanese Journal of Applied Physics Part 2–Letters* **37**(4a), L404–05. DOI:10.1143/Jjap.37.L404.

Symonova, O., et al. 2015. "Dynamicroots: A Software Platform for the Reconstruction and Analysis of Growing Plant Roots," *PLOS ONE* **10**(6), e0127657. DOI:10.1371/journal.pone.0127657.

Szafranski, P., et al. 1997. "A New Approach for Containment of Microorganisms: Dual Control of Streptavidin Expression by Antisense RNA and the T7 Transcription System," *Proceedings of the National Academy of Sciences of the USA* **94**(4), 1059–63.

Terwilliger, T. C., et al. 2007. "Interpretation of Ensembles Created by Multiple Iterative Rebuilding of Macromolecular Models," *Acta Crystallographica Section D Biological Crystallography* **63**(Pt 5), 597–610. DOI:10.1107/S0907444907009791.

Thamdrup, B. 2000. "Bacterial Manganese and Iron Reduction in Aquatic Sediments," In *Advances in Microbial Ecology, Vol.16*. 41–84. Ed. B. Schink. Springer, U.S. DOI:10.1007/978-1-4615-4187-5.

Thodey, K., et al. 2014. "A Microbial Biomanufacturing Platform for Natural and Semisynthetic Opioids," *Nature Chemical Biology* **10**(10), 837–44. DOI:10.1038/nchembio.1613.

Trajano, H. L., et al. 2013. "The Fate of Lignin During Hydrothermal Pretreatment," *Biotechnology for Biofuels* **6**(1), article 110. DOI:10.1186/1754-6834-6-110.

Tringe, S. G., et al. 2005. "Comparative Metagenomics of Microbial Communities," *Science* **308**(5721), 554–57. DOI:10.1126/science.1107851.

Van Benschoten, A. H., et al. 2016. "Measuring and Modeling Diffuse Scattering in Protein X-Ray Crystallography," *Proceedings of the National Academy of Sciences of the USA* **113**(15), 4069–74. DOI:10.1073/pnas.1524048113.

van den Bedem, H., and J. S. Fraser. 2015. "Integrative, Dynamic Structural Biology at Atomic Resolution—It's About Time," *Nature Methods* **12**(4), 307–18. DOI:10.1038/nmeth.3324.

Veličković, D., et al. 2014. "New Insights into the Structural and Spatial Variability of Cell-Wall Polysaccharides During Wheat Grain Development, as Revealed through Maldi Mass Spectrometry Imaging," *Journal of Experimental Botany* **65**(8), 2079–91. DOI:10.1093/jxb/eru065.

Veličković, D., et al. 2016. "Mass Spectrometric Imaging of Wheat (*Triticum* Spp.) and Barley (*Hordeum vulgare* L.) Cultivars: Distribution of Major Cell Wall Polysaccharides According to Their Main Structural Features," *Journal of Agricultural and Food Chemistry* **64**(32), 6249–56. DOI:10.1021/acs.jafc.6b02047.

Venter, J. C., et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* **304**(5667), 66–74. DOI:10.1126/science.1093857.

Votapka, L. W., and R. E. Amaro. 2015. "Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning," *PLOS Computational Biology* **11**(10), e1004381. DOI:10.1371/journal.pcbi.1004381.

Voytas, D. F., and C. Gao. 2014. "Precision Genome Engineering and Agriculture: Opportunities and Regulatory Challenges," *PLOS Biology* **12**(6), e1001877. DOI:10.1371/journal.pbio.1001877.

Walk, T. C., et al. 2004. "Modelling Applicability of Fractal Analysis to Efficiency of Soil Exploration by Roots," *Annals of Botany* **94**(1), 119–28. DOI:10.1093/aob/mch116.

Wall, M. E., et al. 2014. "Diffuse X-Ray Scattering to Model Protein Motions," *Structure* **22**(2), 182–84. DOI:10.1016/j.str.2014.01.002.

Wang, H. H., et al. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated Evolution," *Nature* **460**(7257), 894–98. DOI:10.1038/nature08187.

Wang, T., and M. Hong. 2016. "Solid-State NMR Investigations of Cellulose Structure and Interactions with Matrix Polysaccharides in Plant Primary Cell Walls," *Journal of Experimental Botany* **67**(2), 503–14. DOI:10.1093/jxb/erv416.

Wang, Z., et al. 2014. "An Atomic Model of Brome Mosaic Virus Using Direct Electron Detection and Real-Space Optimization," *Nature Communications* **5**, article 4808. DOI:10.1038/ncomms5808.

Ward, A. B., et al. 2013. "Biochemistry. Integrative Structural Biology," *Science* **339**(6122), 913–15. DOI:10.1126/science.1228565.

Warren, J. M., et al. 2013. "Neutron Imaging Reveals Internal Plant Water Dynamics," *Plant and Soil* **366**(1), 683–93. DOI:10.1007/s11104-012-1579-7.

Way, J. C., et al. 2014. "Integrating Biological Redesign: Where Synthetic Biology Came from and Where It Needs to Go," *Cell* **157**(1), 151–61. DOI:10.1016/j.cell.2014.02.039.

Weston, L. A., et al. 2012. "Mechanisms for Cellular Transport and Release of Allelochemicals from Plant Roots into the Rhizosphere," *Journal of Experimental Botany* **63**(9), 3445–54. DOI:10.1093/jxb/ers054.

Wheeler, H. L., et al. 2015. "Comprehensive Multiphase NMR: A Promising Technology to Study Plants in Their Native State," *Magnetic Resonance in Chemistry* **53**(9), 735–44. DOI:10.1002/mrc.4230.

White, W. E., et al. 2015. "The Linac Coherent Light Source," *Journal of Synchrotron Radiation* **22**(3), 472–76. DOI:10.1107/S1600577515005196.

Wieder, W. R., et al. 2013. "Global Soil Carbon Projections Are Improved by Modelling Microbial Processes," *Nature Climate Change* **3**(10), 909–12. DOI:10.1038/Nclimate1951.

Woldeyes, R. A., et al. 2014. "*E Pluribus Unum*, No More: From One Crystal, Many Conformations," *Current Opinion in Structural Biology* **28**, 56–62. DOI:10.1016/j.sbi.2014.07.005.

Wolf, S. G., et al. 2014. "Cryo-Scanning Transmission Electron Tomography of Vitrified Cells," *Nature Methods* **11**(4), 423–8. DOI:10.1038/nmeth.2842.

York, L. M., and J. P. Lynch. 2015. "Intensive Field Phenotyping of Maize (*Zea mays* L.) Root Crowns Identifies Phenes and Phene Integration Associated with Plant Growth and Nitrogen Acquisition," *Journal of Experimental Botany* **66**(18), 5493–505. DOI:10.1093/jxb/erv241.

York, L. M., et al. 2016. "The Holistic Rhizosphere: Integrating Zones, Processes, and Semantics in the Soil Influenced by Roots," *Journal of Experimental Botany* **67**(12), 3629–43. DOI:10.1093/jxb/erw108.

Young, I. D., et al. 2016. "Structure of Photosystem II and Substrate Binding at Room Temperature," *Nature* **540**(7633), 453–57. DOI:10.1038/nature20161.

Young, I. M., and J. W. Crawford. 2004. "Interactions and Self-Organization in the Soil-Metal Complex," *Science* **304**(5677), 1634–37. DOI:10.1126/science.1097394.

Yu, I., et al. 2016. "Biomolecular Interactions Modulate Macromolecular Structure and Dynamics in Atomistic Model of a Bacterial Cytoplasm," *eLife* **5**. DOI:10.7554/eLife.19274.

Yu, J. S., and N. Bagheri. 2016. "Multi-Class and Multi-Scale Models of Complex Biological Phenomena," *Current Opinion in Biotechnology* **39**, 167–73. DOI:10.1016/j.copbio.2016.04.002.

Zanacchi, F. C., et al. 2011. "Live-Cell 3D Super-Resolution Imaging in Thick Biological Samples," *Nature Methods* **8**(12), 1047–49. DOI:10.1038/nmeth.1744.

Zeng, Y., et al. 2014. "Lignin Plays a Negative Role in the Biochemical Process for Producing Lignocellulosic Biofuels," *Current Opinion in Biotechnology* **27**, 38–45. DOI:10.1016/j.copbio.2013.09.008.

Zhang, R., et al. 2014. "High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA," *Plant Cell* **26**(4), 1398–409. DOI:10.1105/tpc.114.124099.

Zhang, X., et al. 2013. "Cryo-EM Structure of the Mature Dengue Virus at 3.5-Å Resolution," *Nature Structural & Molecular Biology* **20**(1), 105–10. DOI:10.1038/nsmb.2463.

Zhao, G., et al. 2013. "Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics," *Nature* **497**(7451), 643–46. DOI:10.1038/nature12162.

Zhu, X., et al. 2016. "Measuring Spectroscopy and Magnetism of Extracted and Intracellular Magnetosomes Using Soft X-Ray Ptychography," *Proceedings of the National Academy of Sciences of the USA* **113**(51), E8219–27. DOI:10.1073/pnas.1610260114.

Zhu, X. G., et al. 2015. "Plants *In Silico*: Why, Why Now and What?—An Integrative Platform for Plant Systems Biology Research," *Plant, Cell & Environment* **39**(5), 1049–57. DOI:10.1111/pce.12673.