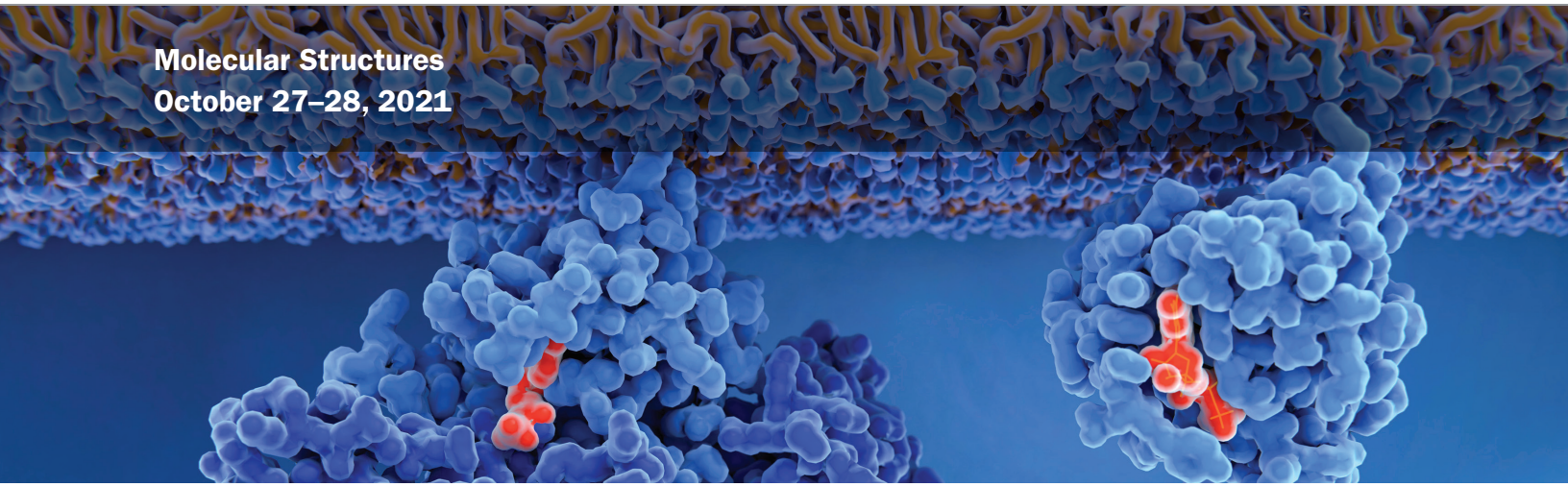
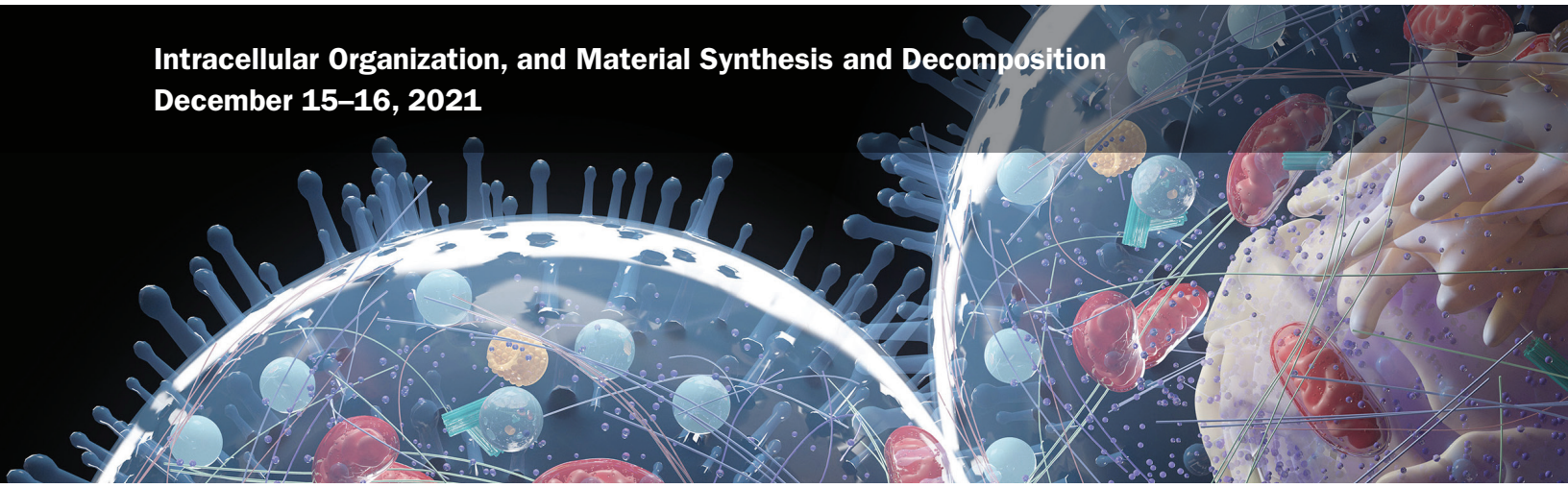
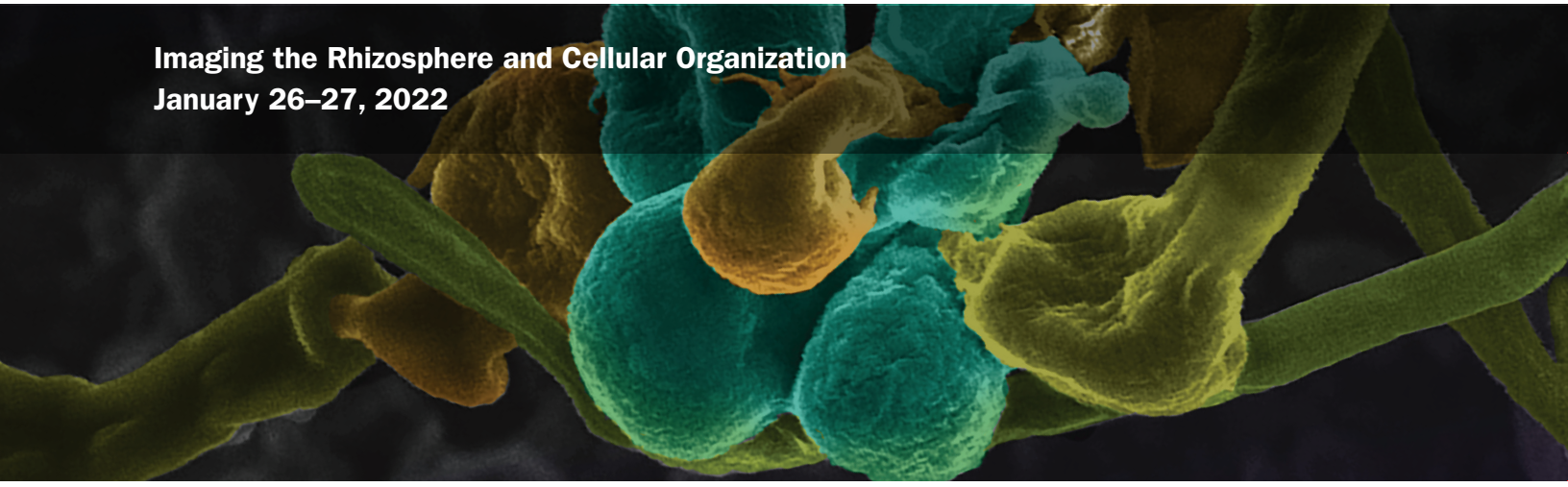
Genomes to Structure and Function

Workshop Report 2022







BERKELEY LAB

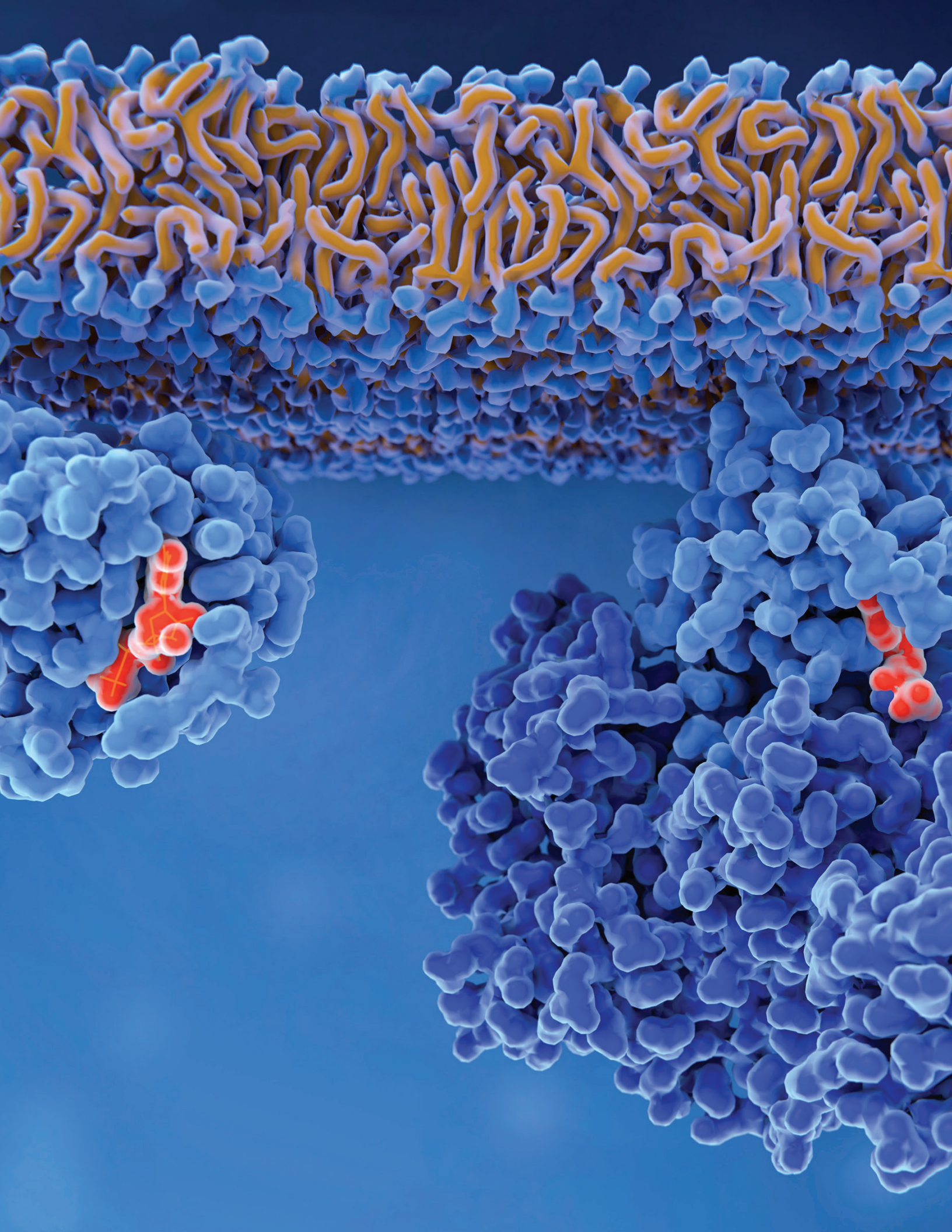
OAK RIDGE National Laboratory

Argonne NATIONAL LABORATORY

Pacific Northwest NATIONAL LABORATORY

Brookhaven National Laboratory

SLAC NATIONAL ACCELERATOR LABORATORY



EXECUTIVE SUMMARY 2

SUMMARY OF OVERARCHING CHALLENGES AND OPPORTUNITIES 3

Science 3

Technology Development 3

DOE Office of Science (SC) User Facility Integration 3

INTRODUCTION 4

Motivation for the Workshop 4

Successful Delivery of Several Pilot Projects 5

Structure-Informed Discovery and Design Cycle 7

Goals of the Workshop 7

OVERARCHING CHALLENGES AND OPPORTUNITIES 8

1. Science 8

2. Technology Development 13

3. User Facility Integration 14

CONCLUSIONS 18

APPENDICES 19

Appendix 1. Organizing Committee Members 19

Appendix 2. Speakers 20

Appendix 3. Participant Count 23

Appendix 4. Workshop Agendas 24

Appendix 5. Summary of Presentations 27

Session 1. Molecular Structures 27

Session 2. Intracellular Organization, and Material Synthesis and Decomposition 29

Session 3. Imaging the Rhizosphere and Cellular Organization 33

EXECUTIVE SUMMARY

The goal of the U.S. Department of Energy (DOE) Biological and Environmental Research (BER) Program is to achieve a predictive understanding of complex biological, earth, and environmental systems with the aim of advancing the nation's energy and infrastructure security. (https://www.energy.gov/science/ber/biological-and-environmental-research). To pursue this goal, collaborations among experts in diverse research areas that lead to multidisciplinary projects are indispensable. The roles of DOE's User Facilities, which offer unique and powerful resources for such research projects, are evolving, and expectations for the facilities are increasing. To respond to Users' needs, the Joint Genome Institute (JGI) and Environmental Molecular Sciences Laboratory (EMSL) initiated the Facilities Integrating Collaborations for User Science (FICUS) program in 2014. This collaboration has grown into a popular and successful program, advancing more than 100 multidisciplinary projects to date. Similarly, the new inter-Facility collaborations among the JGI, EMSL, and User resources for BER structural biology and imaging at the Basic Energy Science (BES) Program's synchrotron and neutron facilities are becoming essential for cutting-edge transdisciplinary science.

To further explore the need for the BER research community to combine genomic, functional, and structural approaches to advance their research, an organizing committee was formed to develop and jointly host a 3-part workshop. The committee's members represented seven DOE National Laboratory User Facilities (**Appendix 1** lists the members). The "Genomes to Structure and Function" virtual workshop (see **Appendices 2–5**) was composed of three sessions. The first session, titled "Molecular Structures" (October 27–28, 2021), highlighted diverse integrative experimental and computational approaches correlating structural data with sequencing and functional information, as well as predicting protein structures to model complex biological systems. The second session, "Intracellular Organization, and Material Synthesis and Decomposition" (December 15–16, 2021), covered imaging methods for observing, quantifying, and manipulating biosystems. The third session, "Imaging the Rhizosphere and Cellular Organization" (January 26–27, 2022) emphasized advanced and non-invasive imaging techniques applied to plant root-microbe-soil interactions.

On average, about 140 BER-supported researchers at diverse career levels (Principal Investigators (PI), Postdoctoral associates, and graduate students) registered for each session, and about 100 people participated in each (**Appendix 3**), demonstrating keen interest and enthusiasm from the BER research community for the science that could be enabled by the expanded inter-Facility collaborations.

At each workshop session, invited speakers first presented the current state-of-the-art science and examples of research projects (**Appendix 5**). Two organizing committee members moderated each day. Several breakout discussions were subsequently held to discuss how the capabilities discussed can be used by BER research communities. Each breakout group focused on one of 5 main topics: grand challenges, expanded inter-Facility collaborations, multimodal and correlative imaging, high-throughput technologies, and biosystems design. The first workshop session included two additional topics: high resolution structure prediction and complex and heterogenous systems. After each breakout discussion, the moderators reported on key discussion points to the entire group. These findings are summarized in three major overarching challenges and opportunities: science, technology development, and inter-Facility integration challenges.

https://doi.org/10.2172/1959294

SUMMARY OF OVERARCHING CHALLENGES AND OPPORTUNITIES

The participants recognized the value of expanded inter-Facility collaborations for their research. However, challenges exist in three main areas to realize such collaborations. Addressing these challenges will open up more opportunities for users to combine genomic, functional, and structural approaches to address important biological questions and to advance biosystems design capabilities for BER science.

Science

- **Discovery and predictive design**: The ability to discover and predictively design biological macromolecules with desired functions, which will help advance understanding of biological systems and provide the basis for biosystems design, needs to be improved and can be facilitated through increased inter-Facility collaborations.
- **Genotype-phenotype relationships**: To understand underlying mechanisms that result in particular biological phenotypes, interactive and holistic cell models need to be developed at multiple scales. These developments, led through inter-Facility collaborations, would also allow for integrative analysis of omics datasets and development of new predictive tools for biosystems design.
- **Cellular responses to environmental conditions**: Improving knowledge of microbial responses to changing environments is essential for designing robust biosystems with novel functions. Understanding the roles and interactions of vital genes and gene products, metabolites, regulatory systems, and machinery involved in cellular responses will help inform molecular mechanisms that guide adaptation of microbes to new conditions and environments.
- **Chemical interactions between microbes and plants**: Imaging and characterizing chemical exchanges between microbes and plants in diverse environmental conditions is critical for understanding and leveraging their interactions for more sustainable production of energy crops and for effective carbon storage in soil. Inter-Facility collaborations can play a major role in establishing and making available such capabilities.

Technology Development

- **Experimental scale gaps**: Predicting and modeling biological responses and mechanisms requires integration of data and methods across different experimental scales in all domains of BER science.
- **Sample complexity**: Novel separation, preparation, and delivery technologies are necessary for studying more complex samples.
- **High throughput data generation**: Automation and miniaturization of all processes should be pursued to allow development or utilization of high-throughput platforms. This would help researchers survey a larger range of conditions, dissect complex or heterogeneous data, and screen for suitable samples.

DOE Office of Science (SC) User Facility Integration

- **Administration**: Approaches should be developed to make it easier for research teams to submit proposals for inter-Facility collaborations. Possible improvements include expanding inter-Facility solicitations and adopting universal proposal applications.
- **Scientific expertise**: To increase interactions between the BER community and the user facilities, better integration of facility scientists with BER research projects would help bridge knowledge gaps in the community. Furthermore, BER scientists unfamiliar with User Facility capabilities could benefit if the facilities offered more educational opportunities and training programs on available capabilities.
- **Data integration and access**: Data could be integrated better if the User Facilities could find ways to standardize data formats (including metadata) across the many analytical techniques and facilitate sharing of data across the facilities. Standard access policies across User Facilities will be required for the FAIR principles of data management.
- **Experimental workflows**: Developing standard protocols for handling biological samples including sample preparation, image registration would help integrate techniques across the User Facilities.

INTRODUCTION

Motivation for the Workshop

BER 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 producing fuels and chemicals from renewable energy crops that could be grown compatibly with food or animal feed crops while not competing with other societal needs. Additionally, BER supports research leading to understanding 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.

BER researchers have made notable advances in genomics and molecular and systems biology research that have enabled them to apply their knowledge to designing biosystems. BER User Facilities, such as the Joint Genome Institute (JGI) and Environmental and Molecular Science Laboratory (EMSL), contribute much to these efforts. To support this work further, BER supports structural biology and imaging user resources (https://berstructuralbioportal.org/) at six other User Facilities operated by DOE Basic Energy Sciences (BES): the Advanced Light Source (ALS), the Advanced Photon Source (APS), the National Synchrotron Light Source II (NSLS-II), the High Flux Isotope Reactor (HFIR), the Spallation Neutron Source (SNS), and the Stanford Synchrotron Radiation Lightsource (SSRL). These synchrotron, neutron, and cryo-EM resources offer unique and powerful capabilities to provide the highest-resolution information at all levels of biological investigation, from atomic and molecular levels, through organelle and cellular levels and beyond, to systems and community levels that define interactions in microbiomes and in the rhizosphere.

Structural biology and imaging results can be combined with genomics and molecular and systems biology approaches to understand diverse genome-encoded properties of plants and microbes and to harness them for designing more efficient biosystems. For example, high-resolution structures are important for understanding the function and dynamics of biological macromolecules and their interactions. Structural information is also critical for genetic engineering efforts to enhance and alter protein function in engineered systems. Additionally, organelle and cell images can provide important information on how substrates and products are transported, accumulated, and stored. These images can also help researchers redesign cellular processes for synthesizing and decomposing more complex materials. Imaging plants and their associated microbes helps researchers understand their spatial distributions, dynamics, and interactions. This information is critical for microbiome engineering to develop more sustainable ways to produce energy crops.

There is an opportunity for the BER research community to increase the impact of their research by combining large-scale genomics technologies available at the JGI and high-throughput functional genomics capabilities at the EMSL, with the powerful and unique capabilities at the structural biology and imaging resources. JGI and EMSL have established an effective approach to encourage collaborative use of the complementary capabilities at both facilities that will advance a BER research project further than either one could do alone. This approach, "Facilities Integrating Collaborations for User Science" (FICUS) offers joint proposal calls that solicit BER projects requesting capabilities from each of the two facilities (https://jgi.doe.gov/user-programs/program-info/ficus-overview/emsl/, https://www.emsl.pnnl.gov/basic/ficus-program/1872). Recent expanded FICUS proposal calls have included additional federal facilities or user resources with distinct capabilities. For example, Oak Ridge National Laboratory's (ORNL)'s Center for Structural Molecular Biology (CSMB), which offers biological small angle neutron scattering (Bio-SANS) and the NSF-funded National Ecological Observatory Network (NEON), which provides access to its repository of representative samples from their long-term field research sites, have participated in recent calls. Further expansion of FICUS, or development of other types of inter-Facility opportunities for users involving cooperative arrangements, would be ideal for offering strategic combinations of User Resources with unique capabilities.

To better enable access to inter-Facility capabilities to advance BER research, some capability gaps need to be filled, operational differences between User Facilities must be addressed, and sample and data sharing must be improved and standardized. Furthermore, new technologies must be developed, such as computational methods for assembling realistic models of complex biological systems by integrating multiple sources of

experimental data. To better understand the needs of the BER researchers, the organizing committee therefore hosted a workshop for the BER research community titled, "Genomes to Structure and Function" to explore these opportunities.

Successful Delivery of Several Pilot Projects

The JGI and EMSL have offered BER researchers their capabilities through the FICUS program since 2014 and have worked on over 100 joint user projects. Many of these projects have been completed and their results have been published, including a number in high-impact journals such as *Science* (10.1126/science.aad1431). In line with these efforts, the JGI, EMSL, and scientists at DOE BES's synchrotron and neutron User Facilities together selected several pilot projects which led to impactful outcomes. For example:

- The SSRL and the JGI collaborated on a project with Dr. Tobias Erb at the Max Planck Institute (MPI) for Terrestrial Microbiology. In this collaboration, the SSRL scientists elucidated the structure of enoyl-CoA carboxylase reductase (ECR), the world's fastest CO_2 fixation enzyme. The ECR structure allowed the researchers to identify residues involved in CO_2 fixation (*PNAS*: 10.1073/pnas.1901471116), as well as oligomeric structures and their dynamics important for accelerating the speed and efficiency of CO_2 fixation (*ACS Central Science*: 10.1021/acscentsci.2c00057) (**Figure 1**).
- In another pilot project, scientists at the ALS and the JGI collaborated with Dr. Steven Withers at the University of British Columbia. In this collaboration, the researchers discovered a glycosyl hydrolase family 94 enzyme (a phosphorylase) that could catalyze concatenation of *N*-acetylglucosamine

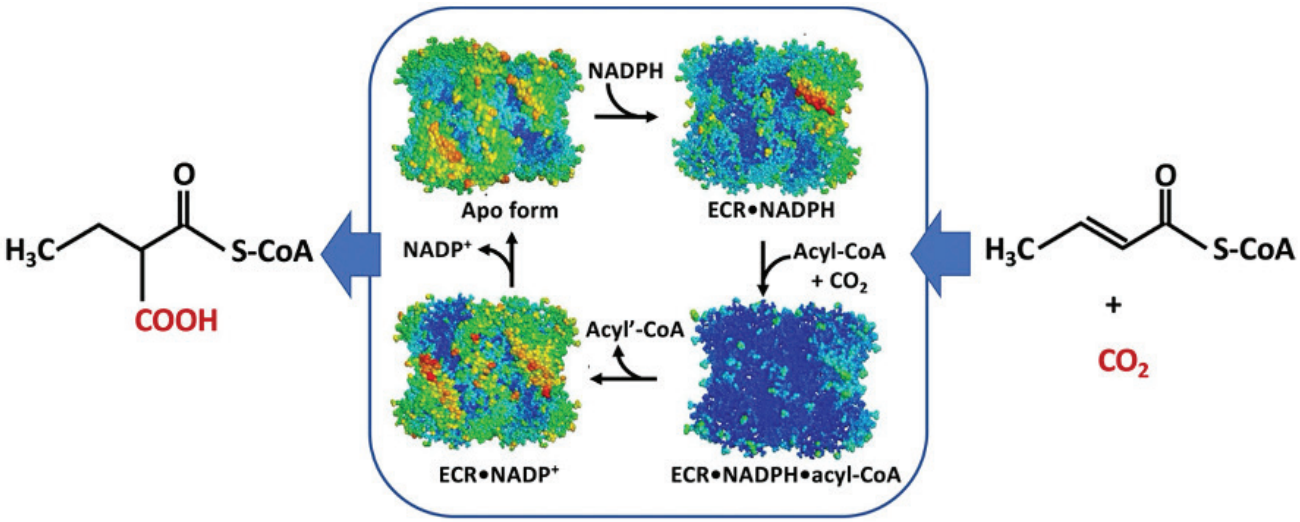


Figure 1. Molecular-basis understanding of ECR, the world fastest CO_2 fixation enzyme

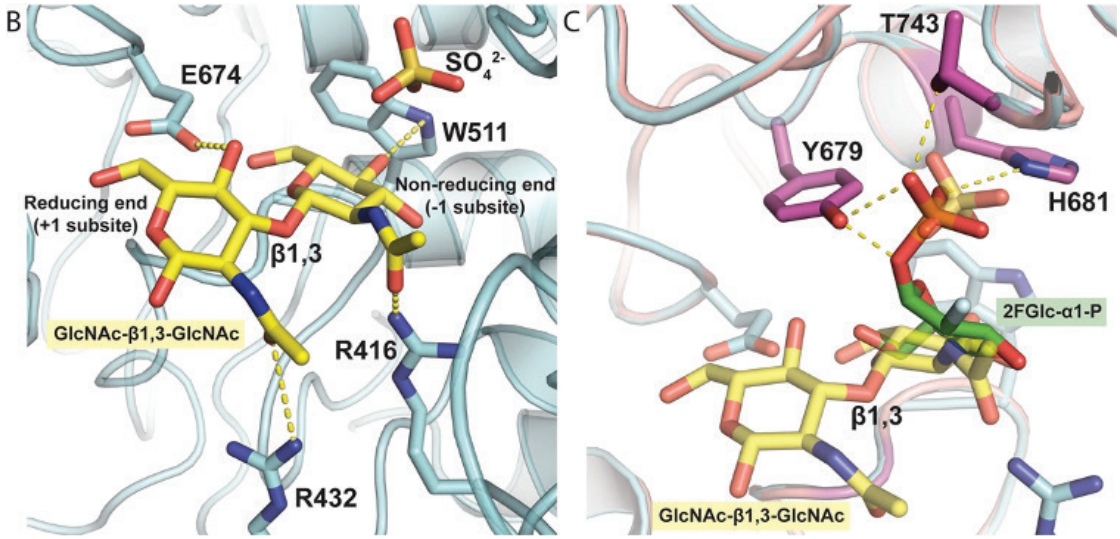


Figure 2. Molecular-basis understanding novel GH94 family enzyme, acholetin phosphorylase.

(GlcNac) to yield *ß-1,3-N*-acetylglucosaminide, which Dr. Withers named "acholetin" after a chitin-like polysaccharide produced by *Acholeplasma*. The LBNL scientists then elucidated the structure of this acholetin phosphorylase, providing molecular-basis mechanisms for synthesis of acholetin catalyzed by acholetin phosphorylase (*ACS Central Science*: 10.1021/acscentsci.1c01570) (**Figure 2**).

Other pilot projects are ongoing and several more papers are currently underway. These results show that inter-Facility collaborations among the JGI, EMSL, and groups at BES synchrotron and neutron User Facilities can work well and advance BER community science and its impact.

These collaborative efforts identified several capability gaps due to differences in operations, procedures for sample and data sharing, and standards for implementation of new technologies among the Facilities and User Resources. For example, unlike EMSL and the JGI, the beamline facilities typically only offer users access to particular instruments and have limited resources for sample preparation. To approach the broader BER community, these facilities must be able to help produce and purify proteins and other biomacromolecules at a large-scale, prepare samples for protein crystallization, cryo-EM, etc. and offer characterization capabilities.

BER community's interest in genome and metagenome analyses of environmental samples has continued to grow. These studies have led to the discovery of novel enzymes and enzyme complexes which play important roles in ecosystems, carbon cycling, and biogeochemistry. JGI's and EMSL's investment in high through put -omics technologies and associated imaging, microscopy and spectroscopy techniques have played a critical role in enabling advances in this research area. However, some limitations in capabilities have been identified that need to be addressed. The lack of capabilities for biochemical and functional characterization and the limited capacity to integrate data across scales and combine information from different analytical techniques have been identified as significant gaps. In addition, differences in operations, procedures for sample and data sharing, and standards for implementation of new technologies among the Facilities and User Resources also need to be addressed. Uniform access requirements to beamline facilities and resources for producing proteins and other biomacromolecules at a large-scale for protein crystallization and cryo-EM are also needed. Additionally, computational analyses to support systems biology, biosystems design, structural biology, and bioimaging is highly demanded. These gaps hamper multimodal correlative investigations across the User facilities that ultimately limits in-depth understanding of these systems and will need to be filled to realize more extensive inter-Facility collaborations to service the BER research community.

Structure-Informed Discovery and Design Cycle Guided the Discussion among Facility Scientists and BER Researchers

To bring these pilot inter-Facility collaborations to the next level, the workshop further explored the need for BER researchers to combine genomic, functional, and structural data to advance their research. The workshop also explored opportunities, such as how inter-Facility collaborations can support efforts to bridge and/or combine different types of data. To facilitate the discussion, the organizing committee created the "structure-informed discovery and design cycle" outlined in **Figure 3.** This cycle guided workshop participants and leaders to see how inter-Facility collaborations could help advance their research and to identify how their research projects could be better integrated and support development of the broader BER research portfolio.

This cycle has four different phases: 1. genomics/phenomics, 2. molecular function and structure, 3. design and evolution, and 4. biosystems function and structure. Data processing and analysis and predictive artificial intelligence/machine learning (AI/ML) capabilities support all aspects of this cycle. This cycle was also designed to broadly describe the project portfolios that are supported by BER. Examples include microbial and plant genome science, environmental microbiomes, Bioenergy Research Centers, biosystems design, secure biosystems design, and systems and computational biology. In this way, the prospective Users for the inter-Facility collaborations could identify how their projects could potentially be integrated into broader research portfolios of BER and impact other research projects within the community to maximize the overall scientific impacts and outputs. Another major goal for the workshop was to have inter-Facility collaborations to be a hub for accelerating collaborations among Users.

Goals of the Workshop

Goals for this workshop were to explore community interests and to learn more about the needs of BER-funded researchers to combine genomics, functional, and structural approaches in their research. The goals also included identifying gaps and challenges that needed to be overcome to enhance inter-Facility collaborations. To accomplish these goals, the organizing committee focused on the following essential questions:

- How can researchers best exploit investigations into, and knowledge of, structures and functions of biomolecules, cellular systems, and rhizosphere communities to improve understanding of biological systems and their interactions?
- How can this knowledge be best applied to the challenge of creating a more sustainable bioeconomy?
- How can the barriers due to operational differences between Facilities be lowered so prospective Users can access unique and powerful capabilities offered at different Facilities?
- What access models and capabilities are needed to maximize User Facility impact to BER science?
- How can data be standardized and shared across multiple data types and different Facilities?
- What new technologies should be developed?

The presentations and discussion sessions encompassed a broad spectrum of BER research, including biosystems design, bioenergy and bioproducts, systems biology, biomaterials synthesis, environmental microbiology,

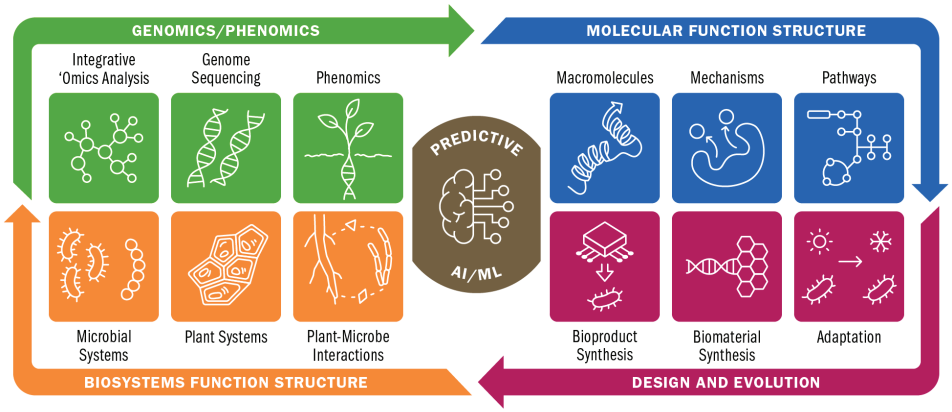


Figure 3. Structure-informed discovery and design cycle

element cycling, secure biosystems design, climate change, and plant-microbe interactions. The potential for leveraging BER-supported databases and computational resources was also explored; these databases are the Protein Data Bank (PDB), the DOE Systems Biology Knowledge Base (KBase), and the National Microbiome Data Collaborative (NMDC). Finally, the potential impact of expanded inter-Facility collaborations focusing on genomics, structural biology, biosystem design and creating interactive and realistic models at multiple scales was discussed.

OVERARCHING CHALLENGES AND OPPORTUNITIES

On each session day, invited speakers first presented the current state-of-the-art science and examples of research projects (see **v** for summary of science talks). Several breakout discussions were subsequently held to discuss how the highlighted capabilities were being used in BER research communities. Each breakout group focused on one of five main topics: grand challenges, inter-Facility collaborations, multimodal and correlative imaging, high-throughput technologies, and biosystems design. The first workshop session included two additional topics: high resolution structure prediction and complex and heterogeneous systems.

After each breakout discussion, the moderators reported on key discussion points from the breakout discussions. Among these breakout discussions, three key overarching challenges and opportunities were identified: 1. Science, 2. Technology development, and 3. User Facility integration. These are discussed below.

1. Science

The BER researchers who participated in the workshop were enthusiastic about potential inter-Facility collaborations. The researchers expect these collaborations to serve as hubs for diverse integrative research programs requiring world-class multi-disciplinary capabilities to generate and analyze large-scale high-quality datasets and exploiting capabilities uniquely available through DOE User Facilities. More specifically, the breakout sessions hosted during the workshop identified four areas of scientific challenges that could be addressed through such collaborations:

A. Discovering and Engineering Proteins with Novel and/or Desired Function

The process of designing and building new biosystems for biofuel, biochemical, and biomaterial synthesis relies on understanding the functions of the individual macromolecular components, which are primarily proteins. **Therefore, improving the ability to link a protein's function with its sequence and genomic context, structure and dynamics, metabolic pathways, and interactions with other biomolecules in the context of cellular organelles and cells is critical.** To achieve this goal, breakout discussion participants determined that DOE User Facilities might aim to develop "structural proteomics workflows" to funnel target proteins most efficiently for structural characterization as structural characterization is usually the rate-limiting step. The overall process would be to develop and combine technologies to:
1. mine genome databases and identify target proteins;
2. express proteins in a high throughput manner (*in vivo* and

Box 1. Recent Advances in Prediction of Protein Structure

The remarkable recent advances in protein structure prediction, available in AlphaFold and RoseTTAFold, are game-changing. Although several challenges remain, the vastly improved ability to use sequence information to predict atomic models with high accuracy, at least in local domain structure, presents a huge opportunity for biology. Workshop participants saw great potential for predicted models to be used widely in BER research to identify binding pockets, ligand and cofactor interactions, protein, RNA, and DNA interactions, and post-translational modifications. They also identified opportunities to explore and understand the evolutionary relationships between protein sequences at the structural level. BER researchers will be able to develop hypotheses about enzyme mechanisms and specificities that can be tested biochemically, and ultimately validated using structural and molecular biology methods, to obtain accurate details not currently provided by prediction methods. This wider adoption of structural studies to understand function will inevitably lead to an increased need for experimental structure determination, in which inter-Facility collaborations can play an essential role.

cell-free systems); 3. characterize biochemical functions; 4. predict structures (**Box 1**); 5. study dynamics and modifications; 6. predict, design, and tailor functions; and/or 7. link diverse omics results with the functions of target proteins. **Figure 4** summarizes the structural proteomics workflows and was generated based on the scientific talks (**Appendix 5**) describing this integrative approach.

B. Building Holistic Cell Models for DOE-Mission-Relevant Microorganisms

The ability to design and build commercially viable biosystems to produce biofuels, bioproducts, and biomaterials from renewable resources has relied on limited information about cellular systems. To improve these abilities, the community must increase knowledge of the biological, physical, and chemical processes of biomacromolecules at the systems level. Knowledge of nano- to mesoscale organization of cellular assemblies is critical for understanding interactions and processes in cellular metabolism that will lead to the development of new products such as biofuels, bioproducts, and advanced bio-based materials. The workshop presentations related to this topic highlighted different approaches linking genetic methods with various analytical tools such as spectroscopy and microscopy techniques, to study environment-relevant genes in the context of their cellular environment. These studies will lead to new knowledge of the functions of microbial organisms and also of biomolecules in their native biosystems that can be used to accelerate biosystems design applications.

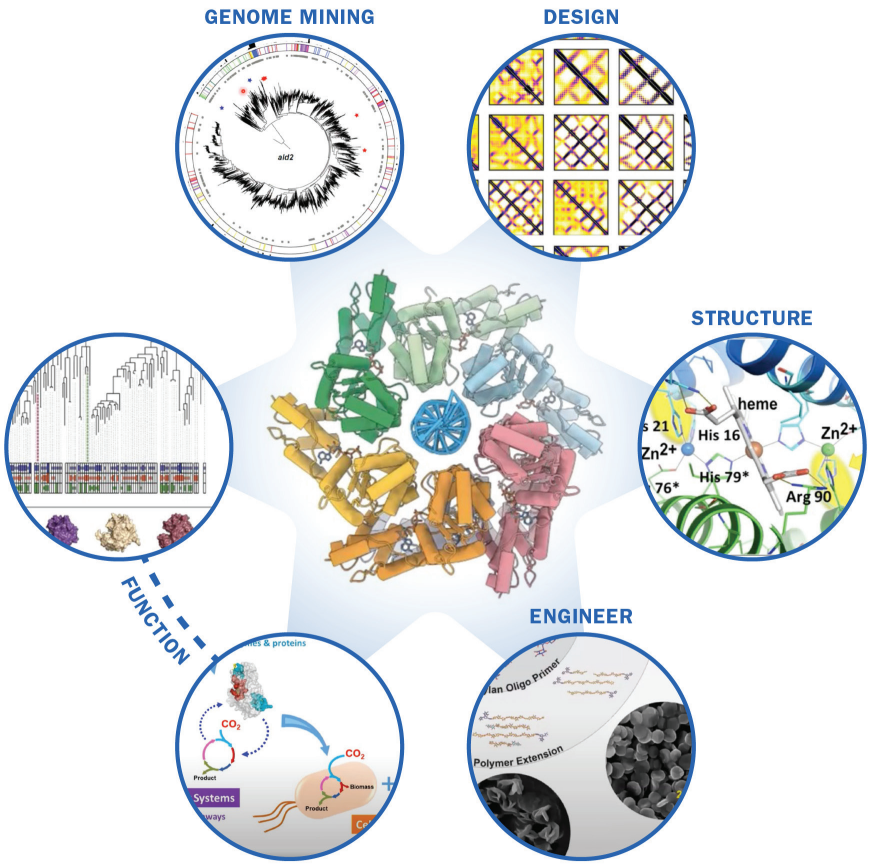
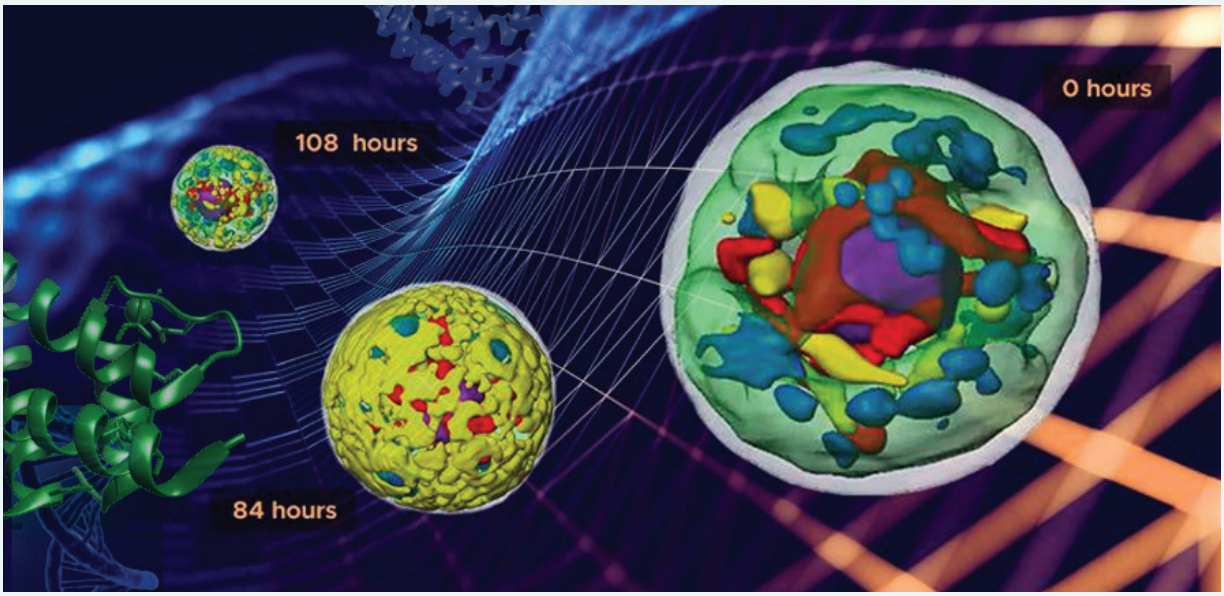


Figure 4. Structural proteomics workflows. Discovering and engineering proteins with novel or desired functions is one opportunity and challenge identified during the breakout sessions. Breakout participants also recommended that the inter-Facility collaboration consider developing structural proteomics workflows such as genome mining, HTP expression and purification, function prediction, characterization, design, and engineering applications. Because structural characterization is usually the liminting step, such workflows could select target proteins quickly through the workflow. Images courtesy of the workshop speakers Dr. Tobias Erb, Dr. Liz Kellogg, Dr. Crysten Blaby, Dr. Brian Fox, Dr. Breeanna Urbanowicz, and Dr. Ivan Anishchanka.

The breakout discussions identified that community science efforts led by inter-Facility collaborations can facilitate creation of **"interactive and holistic cell models at multiple scales integrating multiple sources of experimental data."** (**Box 2**) In these models, the locations, coordination, quantities, and dynamics of proteins and metabolites can be mapped at both organelle and single-cell levels. These models will be important resources for next generation biosystems design. For example, Bioenergy Research Centers have been developing yeast platforms for lipid biosynthesis. Their strategies currently focus mainly on systems and synthetic biology approaches, understanding and manipulating expressions of key enzymes involved in this process, and modulating redox states of the cells. However, understanding how fatty acid synthase (FAS) multi-enzyme complex works, where it is located in the cell, how it is trafficked there, how many complexes optimally fit in the space, what other proteins it is interacting with, and how it takes the substrates and releases the products are all equally important. Relatedly, lipids are accumulated in an organelle called a lipid body. Increasing understanding of the machinery that is involved in lipid body formation, maturation, and maintenance (determining their size, number, and location) in the cells as well as pathways for lipid transportation, accumulation, and degradation is highly desired.

The same principle can be applied to understanding other DOE mission-relevant biochemical processes such as CO_2 fixation, NH_3 oxidation, lignocellulosic degradation, and composite material synthesis. The concept for building interactive and holistic cell models is shown in **Figure 5.** This figure was generated based on the scientific talks describing this integrative approach.



Box 2. Futuristic Cell Models—Modeling Entire Cell in Four Dimensions

A long-term goal for DOE-funded bioimaging research is to model an entire cell in four dimensions. Such a model will enable connecting atomic 3D structures of macromolecules (PDB databank [x-ray and cryo-EM structures] or AlphaFold/RoseTTAfold models) to the intracellular contents. Development of such a model also requires information about abundance and localization of target proteins with different time points, which can be obtained using fluorescent light microscopy and omics and systems biology. The model will help us improve the ability to design biosystems useful for efficient production of biofuels and bioproducts. Which organisms that might be selected as an example to build this futuristic cell model is still an important question for the BER research community.

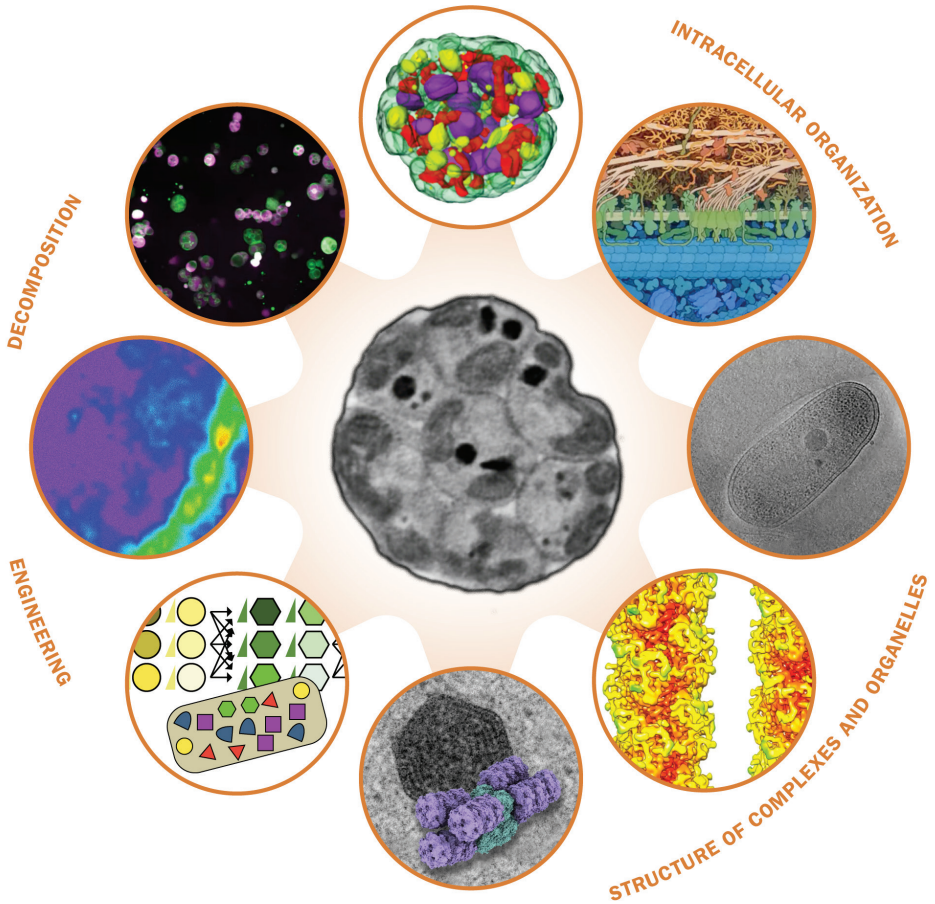


Figure 5. Building interactive and holistic cell models. Holistic cell models were identified as one challenge and opportunity during the breakout sessions. Building such models would enable researchers to connect the structures of macromolecules with organelles and cells, and to study how proteins involved in important biological processes operate in the context of organelles and cells. Moreover, combining this model with other imaging methods, such as fluorescence, IR, and Raman microscopy; integrative omics; and systems biology, would allow real-time and system-wide monitoring of protein behaviors in cells. These models would also provide critical information on strategies for engineering biosystems important for production of fuels and chemicals using other state-of-the-art technologies. Images courtesy of the workshop speakers Dr. Lisa Miller, Dr. Michael Jewett, Dr. Carolyn Larabell, Dr. Stephen Burley, Dr. Cheryl Kerfeld, Dr. Bonnie Bartel, Dr. Nikhil Malvankar, and Dr. James Evans.

C. Understanding Microbial Cellular Responses under Stress Conditions

Improving knowledge of microbial cellular responses under stress conditions is essential for designing robust biosystems to convert lignocellulosic feedstock into biofuels, bioproducts, and biomaterials. To achieve this goal, the breakout discussions determined that community science efforts led by inter-Facility collaborations to **"visualize morphological and compositional responses associated with tolerating chemical inhibitors and physical stresses"** will be essential. However, toxicity caused by lignocellulosic inhibitors, substrates, and products, as well as physical stresses during fermentation prohibit advances in the realization of these processes. The same principle of community science efforts led by inter-Facility collaborations discussed in section B. can be applied to further understand key genes and gene products, metabolites, and the machinery involved in cellular responses to these stress conditions. These community efforts can also include studying underlying genomic and molecular mechanisms for key biological processes such as proliferation, differentiation, aging, and adaptation of microbes to new environments.

D. Imaging Chemical Exchanges in Plant Rhizospheres (beyond Mass-Spec Imaging)

Root-associated microbes play important roles in promoting host plant growth, development, and health. Plants exude nutrients through their roots thereby recruiting these beneficial microbes. In turn, microbes help provide macronutrients and metabolites that

function as biofertilization, biostimulation, and biocontrol agents. **"Imaging chemical exchanges between plants and microbes within the rhizosphere"** is therefore crucial for understanding these plant-microbe interactions and for leveraging these interactions to help increase CO_2 fixation, the yield of energy crops, and carbon storage as soil organic carbon (SOC). As an example, when studying interactions essential for sustainable cultivation of energy crops, multimodal imaging of dynamic rhizosphere processes (e.g., biofilm formation, recruitment of bacteria to the rhizosphere and endosphere) will be important to identify individual microbes and their chemical environments (e.g., root exudate, degradation processes) as well as dynamics. For multimodal imaging, there is a need to develop ways to combine information from non-destructive but relatively low-resolution high throughput methods (e.g., fluorescent-based imaging, radioisotope imaging, positron emission tomography [PET]) that can identify and analyze hot spots and timings for plant-microbe interactions and provide a map for more in-depth but destructive approaches with high-resolution (e.g., cryo-ET, fluorescence *in situ* hybridization [FISH]). Multi-omics-based systems biology studies provide powerful insights, but new approaches are needed to correlate omics and bioimaging datasets. Model ecosystems, such as EcoFAB, will significantly contribute to characterizing rhizosphere processes. Such model ecosystems also allow measurement of abundance and dynamics of metabolites in plant exudates and are used to correlate aboveground phenotypes with the belowground physiology. Currently, there is no direct method for measuring and/or visualizing chemical exchanges, and such new technology must be developed. It is also conceivable that synthetic biology approaches (e.g., biosensors) and new chemical probes may be used in developing direct methods to visualize chemical exchanges between plants and microbes. The concept for imaging chemical exchange between plants and microbes, based on the scientific talks describing this integrative approach, is shown in **Figure 6.**

Figure 6. Imaging chemical exchange between plants and microbes. Interactions between plants and microbes are often mediated by chemical exchanges. These chemicals include nutrients, secondary metabolites, macronutrients, and trace elements. Water activity plays an important role in secretion, diffusion, and intake of these chemicals. Researchers develop devices to facilitate monitoring of plant-microbe interactions using multimodal imaging. Images courtesy of the workshop speakers Dr. Nikolay Kardjilov, Dr. Robert Egbert, Dr. Ryan Tappero, Dr. Jennifer Brophy, and Dr. Trent Northen.

2. Technology Development

There are no mature technologies or workflows that can help connect the experimental findings collected in a lab-designed setting (for example: molecular structure, cellular structure, EcoFab ecosystem) to "real life" observations to reliably and accurately predict and model biological responses and mechanisms. These much-needed workflows should be bidirectional such that the data themselves should provide feedback and serve further as a guide for smarter experimental design. The Design-Build-Test-Learn (DBTL) cycles used in synthetic biology and metabolic engineering enterprise are good examples of such workflows but similar approaches are warranted for studying all other processes such as protein functional annotation and microbe- and plant-microbe interactions. To establish such workflows, there is a need to focus on critical technological aspects that limit their deployment. Current major caveats include **A. insufficient data and limited throughput, and B. limited ability to study complex samples, whether whole or dissected**.

A. Insufficient Data and Limited Throughput

To address limited throughput, miniaturization of all processes should be pursued to minimize the quantity of material required and to accelerate the generation of new high-throughput (HTP) discovery and engineering platforms. All aspects of current technologies (mass spectrometry, x-ray crystallography, cryo-EM, Raman and IR spectroscopy, microbe culturing, and plant growth) from sample preparation to data acquisition and analysis require further automation or improvements in workflows. Increasing throughput is not only critically needed for sampling of a wide range of conditions, but also to increase the number of replicates and thus reduce the impact of sample variability, which is particularly crucial for complex samples. HTP sample screening technologies are also critically important and will be very beneficial for all imaging technologies to downselect the best quality samples.

At the molecular and cellular levels, HTP workflows for structural and functional characterization of proteins along with key resources, such as chemical compound/ligand libraries, computational genome mining, and whole gene deletion libraries will promote scientific discovery for poorly characterized proteins or proteins of unknown function. These workflows (for example, multiplexed protein functional screening, HTP protein synthesis and purification, HTP phenotyping) remain largely underdeveloped. At the community level, there should be a specific focus on developing automation and HTP tools to study the structure of whole microbial communities, which include identifying and quantifying microbes in their natural environments (for example, using spectral signatures and mass spectrometry) along with HTP methods for their isolation. Such HTP tools can include imaging to identify specific or novel microbes, which themselves could use HTP culturing and genomics for their identification and analysis.

For biosystems design efforts, development of HTP DBTL workflows should ideally be universal and compatible with a wide range of functional assays. In terms of data acquisition, it is critical to develop on-the-fly data analysis technologies that allow operator-free data collection and adjustment of conditions along with data quality reporting ideally in real time (self-driving DBTL cycle). The latter will need investments in advanced AI/ML methods and edge computing, but those methods will also require new experimental datasets driven by robotics and automation or curation of current datasets that can be used as training data. Acquiring more in-depth data on a wide range of experimental conditions and perturbations in both spatial and temporal resolutions is necessary to capture dynamics. Investment in multifactorial experimental design to generate and analyze combinatorial effects is also needed.

B. Limited Ability to Produce or Study Complex Samples, whether Whole or Dissected.

To handle biological complexity, the breakout discussion identified several capabilities that inter-Facility collaboration can offer to facilitate improvements in the production of more complex *in vitro* samples or improvements in the characterization of natural heterogeneous sample mixtures.

In terms of *in vitro* production of samples of increased complexity at the macromolecular level, cell-free protein synthesis was highlighted as a powerful tool for generating combinatorial assemblies of biological machinery (e.g., pathways). Although further exploration is required, this capability can offer an efficient way to identify optimal stoichiometry of each protein component and conditions for a subsequent functional and structural characterization. This pipeline is being widely used to accelerate DBTL cycles in biosystems design but should be explored for other HTP workflows.

The majority of protein studies are conducted on proteins lacking any post-translational modifications

(PTMs), although many proteins require appropriate PTMs to enable their true biological and physiological activities. Only a limited number of *in vitro* tools that express proteins in their physiologically active state are available to the User community. Specifically, only a few exist for select post-translational modification of bacterial proteins, and none are available for proteins of eukaryotic origin. Therefore, capabilities for *in vitro* synthesis of post-translational modification of proteins require further development.

As biological samples often comprise complex mixtures of diverse macromolecules and metabolites, improving sample and metadata collections, characterization, and processing and data generation practices and strategies for natural samples of high heterogeneity is critical. Meanwhile, to characterize targets isolated from the mixtures, separation capabilities (e.g., chromatography, filtration, desalting, microfluidics) are important to deconvolute those samples, and such capabilities should be available at User Facilities. Ideally, these should be scalable technologies either automated or amenable to automation in the future. For biological samples consisting of microbial communities, development of novel biosensors and chemicals for fluorescent labeling are needed to dissect complex spatio-temporal behaviors of these microbes *in situ*. To learn about the chemical exchanges between microbes and plants, combining model ecosystems such as EcoFABs with beamline data collection will be useful (**Box 3**). Such a strategy would standardize sample collection practices and allow measurements of samples to be taken in controllable environmental conditions, with the ability to vary those conditions on the fly. The data obtained through these systems will: 1. be more representative of experimental conditions/systems from which samples are being taken from and 2. allow for resulting data to be more interoperable and reusable across datasets.

The inter-Facility collaboration scientists should proactively seek development of devices that are compatible with sample environments at synchrotron and user beamlines should be prioritized.

3. User Facility Integration Challenges

The inter-Facility integration of capabilities at BER Facilities and BER structural biology and bioimaging resources (at the BES Facilities) has the potential to significantly enhance the impact of research conducted by the BER research community. The opportunities that arise from this integration are numerous considering the breadth of technical capabilities each Facility has to offer. Increasing integration of the facilities, therefore, should be examined holistically rather than in a pair-wise manner so that a common set of approaches and principles is established. But there are numerous issues that need to be addressed that impede the integration of capabilities across DOE Facilities in pursuit of BER science. These problems can be broadly classified into three distinct areas: **A. Administration, B. Inter-Facility support and C. Data integration** (**Figure 7**). Below, some of the challenges and opportunities in each of these three main areas are highlighted.

Box 3. Model Systems vs. Non-Model Systems

Model systems played a significant role in understanding biological systems. Genomic data expanded our knowledge to a huge number of organisms, many novel, evolutionary and functionally very distant from currently available model systems. Bridging the gap between model organisms and non-model organisms is a significant challenge. The research community is still interested in the use of model organisms, particularly in the development of models and tools for heterogeneous systems due to a significant accumulated knowledge of these organisms. Such models and tools cannot be developed quickly mainly because unknown/uncharacterized organisms may be very difficult to cultivate.

Figure 7. Inter-Facity workflow conceptual design

A. Administration

The ability to access many of the beamlines is limited for many researchers. Therefore, it often takes many months to years to obtain useful results. Inter-Facility integration could be facilitated through implementation of mechanisms for quicker access to the relevant capabilities.

a. Universal Proposal System and an Easy-Access Mechanism

Proposals should be viewed from a holistic standpoint across multiple Facilities, and this requires better communication among the different Facilities during the review. However, access to Facilities is often complicated by the lack of a unified application system that would make the process more efficient. The adoption of a universal proposal system or an easy-access mechanism, with more consistent formats and requirements, would increase submissions to multiple facilities, aiding faster and efficient access and successful multimodal experiments.

b. Internal Technical Experts

While Facilities provide extensive information on the technical specifications and points of contact for individual capabilities, some BER researchers still have difficulty in grasping the possibilities of the kinds of scientific investigations that can be performed with these capabilities. Regular inter-Facility meetings among technical experts could accelerate learning and exchanges across various measurement platforms. In addition to explaining technical aspects, Facilities should provide examples of the scientific questions that could be addressed with specific measurements as well as information on the workflows needed for these measurements.

B. Inter-Facility Support

a. Scientific and Technical Knowledge

The BER community does not have the expertise needed to take full advantage of the Office of Science (SC) User Facilities' capabilities broadly. Therefore, there is a need for a more extensive teaching campaign, which should also include a future workforce development component. While many SC User Facilities offer training programs and educational materials to educate BER PIs and junior scientists on the existing capabilities, workshop participants found these to be inadequate or not sufficiently advertised. Complicating these deficiencies, numerous SC User Facility staff do not always have adequate backgrounds in biological sciences as found within the BER-focused User Facilities, and therefore have difficulty communicating with the BER community. Increasing the BER knowledge base of staff at these facilities would facilitate utilization by the BER community. This may be accomplished by having the User Facilities hire staff with BER-relevant domain knowledge or providing educational opportunities for existing staff, such as attendance at BER PI meetings, BER User Facility User meetings, and BER- relevant webinars. Discussions of specific projects with individual BER Users to prepare for the beamline experiments will also naturally have an educational component.

b. Experimental Workflows

A common theme that arose in the presentations and discussions was the need for faster measurement and analysis throughput to accelerate the DBTL cycle. For instance, cell-free protein expression was presented as an important step to accelerate

experiments through combinatorial approaches. Other techniques would also benefit greatly from increased throughput by incorporating automation into workflows. In addition, it was recognized that improved computational capabilities and the application of AI/ML, could have a significant impact on experiment design and throughput, as well as data processing and analysis. This broadly applies to the topics covered in all three workshop sessions.

To address the differences in the length and time scales interrogated by different techniques, many currently independent methods will have to be "normalized" or unified. For instance, sample preparation protocol standardization would help with multimodal data integration. However, various techniques typically require different conditions and often the samples are in different forms (i.e., concentration, phase, pH), complicating data integration. This is a challenging problem, but adoption of standard cell lines, organisms, and experimental setups could play an important role in standardization. Additionally, sample holders and sample environments should be versatile so they can be used on different instruments. Ways to incorporate fiducial markers so that researchers are confident that identical regions-of-interest are interrogated with the different techniques will also need to be developed. There is also an opportunity to employ AI-guided experimental planning to aid integration by identifying gaps and planning workflows for subsequent measurements. Last but not least, attention must be paid to the extent of sample damage, which accumulates when using the same sample in different instruments and which may compromise data quality.

One approach to advance the integration of multimodal techniques (i.e., to validate different operating conditions for the different analysis techniques) is for each User Facility to proactively support efforts that accomplish this integration. This can be achieved by identifying appropriate science drivers that demonstrate how destructive and non-destructive multimodal measurements can be combined in a thoughtful way to provide insights that cannot be obtained by individual techniques.

Integrating structural biology information and imaging output (data derived from the image), as well as predictions drawn from the integrated data, into whole-cell, all-molecule models will become very important. A sufficient number of images or samples is needed to make statistical inferences, as cell heterogeneity could be significant. Careful calibration of quantification for each of the methods is also needed to permit statistical measurements.

c. Sample Shipping and Receiving

Shipping challenges remain an issue and can have an impact on sample integrity and the ability to perform cross-correlative measurements. Improving and streamlining sample shipping/receiving to minimize delays and mitigate improper environmental controls will facilitate this integration.

C. Data Integration

The BER research community employs many different approaches that link genetic methods with various analytical tools such as spectroscopy and microscopy techniques, to study environmental-relevant genes in the context of their cellular environment. Although each approach is powerful by itself and offers tremendous opportunity to the BER research community for advancing science for more efficient design of biosystems, it was clear from the presentations and the ensuing discussions that integration of information from different measurements remains a challenge. New ways to combine information from different techniques in a multimodal correlative manner would provide insights not possible from applying techniques individually and would result in a deeper understanding of the relevant processes occurring at the cellular scale.

Different measurement techniques generate different data types and information content, which makes multimodal integration complicated. These differences among different structural and imaging data are particularly noticeable when working across different spatial and temporal scales. Additionally, omics datasets are numerical and quantitative. These datasets also need to be co-registered with the structural and imaging datasets. Although the new methods should be developed to interpret these data altogether, consistent data formats and standardization are critical, and their successful implementation requires buy-in and adoption from the User community, software developers, and instrument manufacturers (**Box 4**).

Integration of sample metadata into the data structure is also crucial, but this process is often manual, so automated methods to facilitate this integration are needed. In addition, technical and social barriers impede the adoption of metadata into the data structure and need to be addressed. The establishment of an integrated

database with broadly accepted standard metadata formats will be necessary to support the development of AI/ML-driven analysis. Furthermore, data accessibility and re-use policies need to be standardized across the various institutions to facilitate the FAIR principles of data management. Led by BER, progress has been made in standardizing and integrating microbiome-related data through the National Microbiome Data Collaborative (NMDC), but similar initiatives are required for the broader range of experimental data types as discussed during the workshop. For many of the scientific questions being pursued by BER researchers, new data management and computational methods will be essential in finding meaning from the large volumes of disparate data types.

Correlating structural information with functional processes is important for linking genomic information to phenotypic responses. The potential for 3D atomic structures (PDB, EMDB, AlphaFold models) to contribute to functional insights is increased when combined with data from NMDC, super-resolution microscopy and post-translational modification data, ion- and ligand-binding data, biophysical characterization data (pH, UV/Vis, IR), electronic structures and fluorescence properties, as well as nano-computed tomography (NanoCT), which is used as a screening tool for macromolecular structure localization. Neutron and x-ray scattering can be integrated with data from other methods to study environmental perturbation conditions that drive the structural changes. For many of the scientific questions being pursued by BER researchers, new data management and computational methods will be essential to find meaning from the large volumes of disparate data types that will be generated from these methods.

Box 4. Best Practices for Data Integration

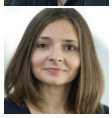
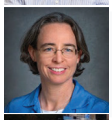
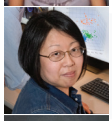
Data sharing among the various facilities needs to improve. Much of the existing data sharing is performed manually but could be facilitated through the adoption of common platforms (e.g., KBase, PDB, etc.) to understand the inner workings of intracellular machinery to the level of the atomic detail, and for recording changes over time during the lifecycle or in response to stimuli. Major gaps remain in the availability of resources and tools to connect atomic 3D structures of individual macromolecules (either experimentally determined or computed) to its cellular environment. Emerging efforts are focused on characterizing and reconstructing significantly larger molecular assemblies (> 10 proteins) which then can serve further as building bricks to reconstruct an entire cell in the future. Emerging integrative methods, also known as hybrid methods, have proven to be key in addressing those challenges and require combining an array of experimental or computational approaches. Examples of large macromolecular complexes have been demonstrated using a combination of cryo-electron tomography data with PDB data (or computed structure models) and proteomics. However, integrating all available data would provide even more context, detail and confidence. Data types may include PDB data augmented with AI/ML computational structure modeling, x-ray/electron tomography, optical microscopy and proteomics, and more, and technologies for data aggregation. Progress has been made in standardizing data and sharing protocols.

CONCLUSIONS

The organizing committee hosted three Genomes to Structure and Function workshop sessions to explore the need for BER researchers to combine genomics, functional, and structural approaches to advance their projects, recognize how the inter-Facility collaboration programs support these efforts, and understand how these programs could be expanded or improved to better meet the needs of the community. In general, the BER researchers attending the workshops showed keen interest, and each session had approximately 100 participants. The breakout discussions during the workshop indicated that inter-Facility collaborations would be a highly valuable mode of accessing an appropriate constellation of resources best suited to address a particular research question. Within the BER research community, such collaborations should be driven by three overarching challenges: science, technology development, and User Facility integration. To make this happen, the following will be needed: 1. For addressing science challenges, the expanded inter-Facility collaborations could drive and coordinate interdisciplinary research projects such as discovering and engineering proteins with novel functions, building holistic cell models to build efficient biosystems and to understand cellular responses to stress conditions, and visualizing chemical exchanges between plants and microbes. 2. To address technology development challenges, capability gaps that result in insufficient data output and limited throughput need to be filled, and new abilities to handle biological complexity require further development. 3. Finally, to address User Facility integration challenges, three components need to be improved and standardized: administration processes, inter-facility sharing of expertise and experimental workflows, and methods for integrating and sharing data. All these challenges represent areas of opportunity that will further BER's goal of understanding of fundamental genome-encoded properties of plants and microbes, their complex interactions, and important biogeochemical environmental processes.

APPENDICES

Appendix 1. Organizing Committee Members

	Paul Adams	Biosciences Area and Advanced Light Source, Lawrence Berkeley National Laboratory (LBNL)	pdadams@lbl.gov
	Hassina Bilheux	Spallation Neutron Source and High Flux Isotope Reactor, Oak Ridge National Laboratory (ORNL)	bilheuxhn@ornl.gov
	Andrzej Joachimiak	Advanced Photon Source (APS), Argonne National Laboratory (ANL)	andrzejj@anl.gov
	Scott Lea	Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory (PNNL)	scott.lea@pnnl.gov
	Sean McSweeney	National Synchrotron Light Source-II (NSLS-II), Brookhaven National Laboratory (BNL)	smcsweeney@bnl.gov
	Karolina Michalska	Advanced Photon Source (APS), ANL	kmichalska@anl.gov
	Irina Novikova	Environmental Molecular Sciences Laboratory (EMSL), PNNL	irina.novikova@pnnl.gov
	Hugh O'Neill	Center for Structural Molecular Biology, ORNL	oneillhm@ornl.gov
	Corie Ralston	The Molecular Foundry (TMF), LBNL	cyralston@lbl.gov
	Ritimukya Sarangi	Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory (SLAC)	ritis@slac.stanford.edu
	Wuxian Shi	National Synchrotron Light Source-II (NSLS-II), BNL	wushi@bnl.gov
	Susannah Tringe	US Department of Energy, Joint Genome Institute (JGI), LBNL	sgtringe@lbl.gov
	Soichi Wakatsuki	Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory (SLAC)	soichi.wakatsuki@stanford.edu
	Yang Yang	National Synchrotron Light Source-II (NSLS-II), BNL	yyang@bnl.gov
	Yasuo Yoshikuni	US Department of Energy, Joint Genome Institute (JGI), LBNL	yyoshikuni@lbl.gov

Appendix 2. Speakers

Philipp Zerbe	Associate Professor, University of California, Davis	*Genomes to Structure and Function in Switchgrass Terpenoid Stress Responses*	pzerbe@ucdavis.edu
Flora Meilleur	Neutron Scattering Scientist, IMAGINE and MaNDi diffractometers; Associate Professor, NCSU Biochemistry, North Carolina State University	*Cellulose-Degrading Oxidative Enzymes: Functional Insights through Neutron Diffraction and Scattering*	fmeille@ncsu.edu
Ivan Anishchanka	(David Baker's Group), Researcher, University of Washington	*Deep Neural Networks for Protein-Protein Complex Prediction and Design*	aivan@uw.edu
Breeanna Urbanowicz	Assistant Professor, University of Georgia	*(Re)engineering Plant Enzymes to Alter Polysaccharide Structure*	breeanna@uga.edu
Justin North	Research Assistant Professor, Ohio State University	*Bacterial Organic Sulfur Metabolism: Mining Novels Genes and Enzymes for Bioproducts*	north.62@osu.edu
Brian Fox	Professor, University of Wisconsin, Madison	*KBase Tools for Integrating Multi-Omics and Structure Data to Discover Protein Function*	bgfox@biochem.wisc.edu
Crysten Blaby-Haas	Scientist, BNL	*Plant Biosynthesis by Design: Understanding Functional Space*	cblaby@bnl.gov
Christopher Henry	Computational Biologist, ANL	*KBase Tools for Integrating Multi-Omics and Structure Data to Discover Protein Function*	chrisshenry@gmail.com
Elizabeth Kellogg	Assistant Professor, Cornell University	*Structural Basis of Target-Site Selection in RNA-Guided DNA Transposition Systems*	lizkellogg@gmail.com
Justin Siegel	Associate Professor, UC Davis	*Discovery and Design of Novel Enzymes for Alkane Production*	jbsiegel@ucdavis.edu
Tobias Erb	Director of Department, Max Planck Institute (MPI)	*Re-Designing CO_2 fixation: New-to-Nature Enzymes, Synthetic Pathways and Artificial Cells for Sustainable Carbon Capture*	toerb@mpi-marburg.mpg.de

Appendix 2. Speakers (continued)

Stephen K. Burley	Director, Rutgers School of Arts and Sciences	*Tools and Resources for Understanding Intracellular Organization in Three Dimensions at the Atomic Level*	stephen.burley@rcsb.org
Carolyne Larabelle	Department Head Cellular and Tissue Imaging, LBNL	*Quantitative 3D Imaging of Biological Specimens Using Soft X-ray Tomography*	CALarabell@lbl.gov
Christopher Francis	Professor, Stanford University	*Molecular Ecology and Structural Biology of Key C- and N-Cycling Enzymes of Ammonia-Oxidizing Thaumarchaeota*	caf@stanford.edu
Cheryl Kerfeld	Department Head Cellular and Tissue Imaging, LBNL	*Bacterial Microcompartments: From Metabolic Organelles to Biomaterials*	ckerfeld@lbl.gov
Bonnie Bartel	Ralph and Dorothy Looney Professor Department of BioSciences, Rice University	*Peroxisome Imaging in Arabidopsis*	bartel@rice.edu
James Evans	Staff Scientist — Biochemist, PNNL	*Visualizing Cellular Ultrastructure with Cryo Electron Microscopy*	james.evans@pnnl.gov
Nikhil Malvankar	Assistant Professor of Molecular Biophysics and Biochemistry, Yale School Of Medicine	*Seeing is Believing: Novel Imaging Methods help Identify Structure and Function of Microbial Nanowires*	nikhil.malvankar@yale.edu
Lisa Miller	Biophysical Chemist, National Synchrotron Light Source II, BNL	*Chemical Imaging of Composite Systems: From Bones to Biofuels*	lmiller@bnl.gov
Jennifer Morrell-Falvey	Senior Scientist and Leader of the Molecular and Cellular Imaging Group, ORNL	*Investigating Plant-Microbe Interactions Using Optical and Chemical Imaging*	morrelljl1@ornl.gov
Michael Jewett	Professor for Chemical and Biological Engineering, Northwestern University	*Establishing Cell-Free Systems to Accelerate Design of Sustainable Biosynthesis Pathways.*	m-jewett@northwestern.edu

Appendix 2. Speakers (cont.)

Nikolay Kardjilov	Helmholtz-Zentrum Berlin für Materialien und Energie	*Advanced Imaging of Root-Soil Interaction*	kardjilov@helmholtz-berlin.de
Si Chen	ANL	*Cellular Trace Element Analysis with the Bionanoprobe*	sichen@anl.gov
Richard Ferrieri	University of Missouri	*Carbon Radioisotope Imaging in the Visualization and Measurement of Plant Resource Allocation to Root-Associating Bacteria*	ferrierir@missouri.edu
Jennifer Brophy	Stanford Bioengineering Schools of Engineering and Medicine	*Synthetic Biology Tools for Engineering Both Sides of Plant-Microbe Interactions*	jennbrophy@gmail.com
Trent Northen	DOE Joint Genome Institute at LBNL	*Exploring Microbial Interactions in the Rhizosphere Using Fabricated Ecosystems*	trnorthen@lbl.gov
Tiina Roose	University of Southampton	*Imaging and Modeling of Rhizosphere Processes*	t.roose@soton.ac.uk
Elizabeth Shank	University of Massachusetts Chan Medical School	*Imaging Microbes and Their Activities Using Transparent Soil Microcosms*	Elizabeth.Shank@umassmed.edu
Ryan Tappero	BNL	*Imaging Trace Elements in the Rhizosphere with X-Ray Fluorescence Microscopy*	rtappero@bnl.gov
Rachel Hestrin	Lawrence Livermore National Laboratory (LLNL)	*Imaging Microbes and Nutrient Flows in the Rhizosphere*	hestrin1@llnl.gov
Kirsten Hofmockel	PNNL	*Multi-Omic and Imaging Analysis of Rhizosphere Interactions*	kirsten.ofmockel@pnnl.gov
Robert Egbert	PNNL	*Rhizosphere Biocontainment through Persistence Control Engineering*	robert.egbert@pnnl.gov

Appendix 3. Participant Count

Molecular Structures Session

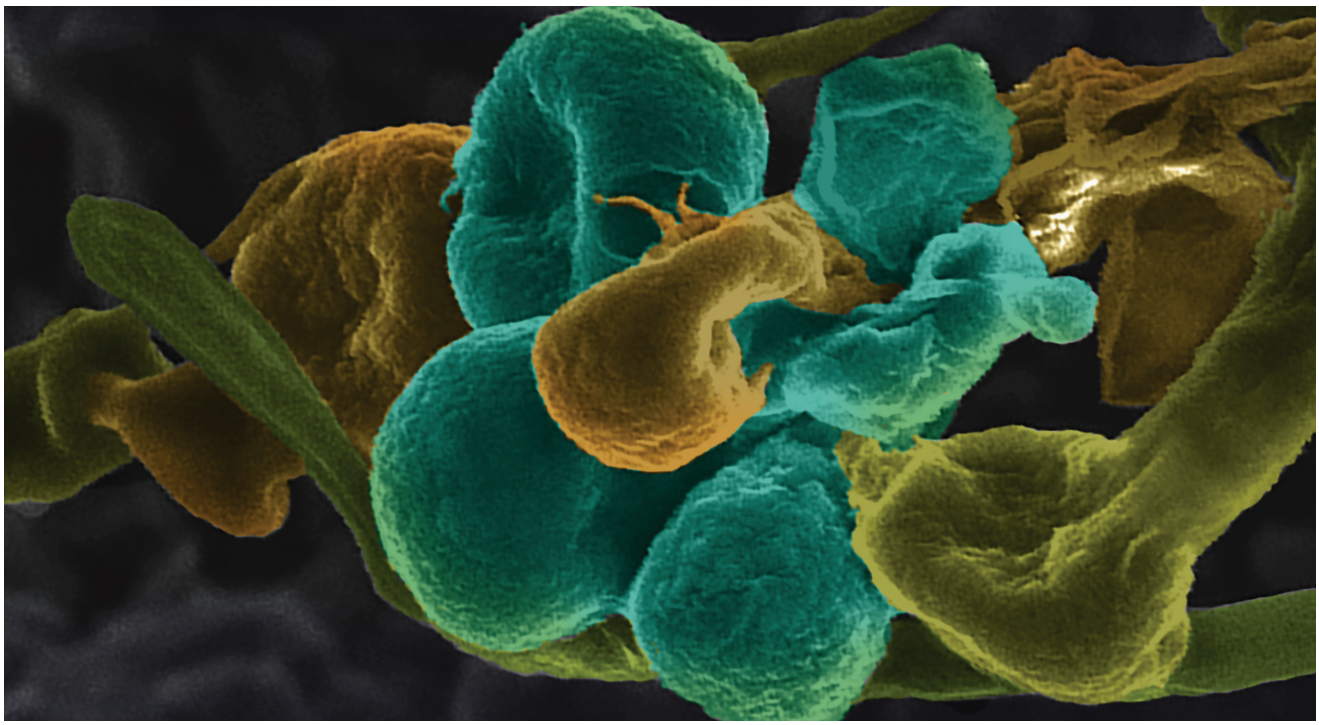
Day 1	115
Day 2	99

Intracellular Organization, and Material Synthesis and Decomposition

Day 1	91
Day 2	78

Imaging the Rhizosphere and Cellular Organization Session

Day 1	101
Day 2	100



Appendix 4. Workshop Agendas

Session 1. Molecular Structures

October 27, 2021, 10:00 am–1:00 pm PT

10:00–10:15 am PT	**Yasuo Yoshikuni**, LBNL, Introduction to the first day of the *Genomes to Structure and Function* workshop Session Chair: **Sean McSweeney**, BNL, **Irina Novikova**, PNNL
10:15–10:30 am PT	**Philipp Zerbe**, UC Davis
10:30–10:45 am PT	**Flora Meilleur**, North Carolina State University
10:45–11:00 am PT	**Ivan Anishchanka**, University of Washington
11:00–11:15 am PT	**Breeanna Urbanowicz**, University of Georgia
11:15–11:30 am PT	**Justin North**, Ohio State University
11:30–11:40 am PT	*Break*
11:40 am–12:40 pm PT	**Breakout Session 1**
12:40–13:00 pm PT	*Report Out*

October 28, 2021, 10:00 am–1:00 pm PT

10:00–10:05 am PT	**Yasuo Yoshikuni**, LBNL, Introduction to the second day of the *Genomes to Structure and Function* workshop Session Chair: **Karolina Michalska**, ANL, **Soichi Wakatsuki**, SLAC
10:05–10:20 am PT	**Brian Fox**, University of Wisconsin, Madison
10:20–10:35 am PT	**Crysten Haas Blaby**, BNL
10:35–10:50 am PT	**Christopher Henry**, ANL
10:50–11:05 am PT	**Elizabeth Kellogg**, Cornell University
11:05–11:20 am PT	**Justin Siegel**, UC Davis
11:20–11:35 am PT	**Tobias Erb**, MPI
11:35–11:45 am PT	*Break*
11:45 am–12:35 pm PT	**Breakout Session 2**
12:35–12:55 pm PT	*Report Out*
12:55–13:00 pm PT	**Closeout**

Session 2. Intracellular Organization, and Material Synthesis and Decomposition

December 15, 2021, 10:00 am–1:00 pm PT

10:00–10:15 am PT	**Yasuo Yoshikuni**, LBNL, Introduction to the first day of the *Genomes to Structure and Function* workshop Session Chair: **Andrzej Joachimiak**, ANL, **Hugh O'Neill**, SLAC
10:15–10:30 am PT	**Stephen Burley**, Director, Rutgers School of Arts and Sciences
10:30–10:45 am PT	**Carolyn Larabell**, LBNL
10:45–11:00 am PT	**Christopher Francis**, Stanford University
11:00–11:15 am PT	**Cheryl Kerfeld**, Michigan State University
11:15–11:30 am PT	**Bonnie Bartel**, Rice University
11:30–11:40 am PT	*Break*
11:40 am–12:40 pm PT	**Breakout Session 1**
12:40–12:55 pm PT	*Report Out*
12:55–13:00 pm PT	**Closeout**

December 16, 2021, 10:00 am–1:00 pm PT

10:00–10:05 am PT	**Yasuo Yoshikuni**, LBNL, Introduction to the second day of the *Genomes to Structure and Function* workshop Session Chair: **Scott Lea**, PNNL, **Yang Yang**, BNL
10:05–10:20 am PT	**James Evans**, PNNL
10:20–10:35 am PT	**Nikhil Malvankar,** Yale School Of Medicine
10:35–10:50 am PT	**Lisa Miller**, BNL
10:50–11:05 am PT	**Michael Jewett**, Northwestern University
11:05–11:20 am PT	**Jennifer Morrell-Falvey**, Scientist, ORNL
11:20–11:30 am PT	***Break***
11:30 am–12:40 pm PT	**Breakout Session 2**
12:40–12:55 pm PT	***Report Out***
12:55–13:00 pm PT	**Closeout**

Session 3. Imaging the Rhizosphere and Cellular Organization

January 26, 2022, 10:00 am–1:00 pm PT

10:00–10:05 am PT	**Yasuo Yoshikuni**, LBNL Session Chair: **Wuxian Shi**, BNL, **Ritimukta Sarangi**, SLAC
10:15–10:30 am PT	**Nikolay Kardjilov**, Helmholtz-Zentrum Berlin für Materialien und Energie
10:30–10:45 am PT	**Si Chen**, ANL
10:45–11:00 am PT	**Richard Ferrieri,** University of Missouri
11:00–11:15 am PT	**Jennifer Brophy**, Stanford University
11:15–11:30 am PT	**Trent Northen**, LBNL
11:30–11:40 am PT	*Break*
11:40 am–12:40 am PT	**Breakout Session 1**
12:40–12:55 am PT	*Report Out*
12:55–13:00 am PT	**Closeout**

January 27, 2022, 10:00 am–1:00 pm PT

10:00–10:05 am PT	**Yasuo Yoshikuni**, LBNL Session Chair: **Susannah Tringe**, LBNL, **Corie Ralston**, LBNL
10:05–10:20 am PT	**Tiina Roose**, University of Southampton
10:20–10:35 am PT	**Elizabeth Shank**, University of Massachusetts
10:35–10:50 am PT	**Ryan Tappero**, BNL
10:50–11:05 am PT	**Rachel Hestrin**, LLNL
11:05–11:20 am PT	**Kirsten Hofmockel**, PNNL
11:20–11:35 am PT	**Robert Egbert**, PNNL
11:35–11:45 am PT	*Break*
11:45 am–12:40 pm PT	**Breakout Session 2**
12:40–12:55 pm PT	*Report Out*
12:55–13:00 pm PT	**Closeout**

Appendix 5. Summary of Presentations

Session 1. Molecular Structures

Understanding the biophysical mechanisms and functional properties of macromolecules and metabolites is a main goal behind molecular structure determination. This 2-day session on Molecular Structure was designed to show the applications and the breadth of structural biology technologies on the scientific topics important to a DOE BER community. The selected topics were genome mining to uncover structural/functional diversity of metabolites and enzymes, high-resolution bioimaging to characterize critical enzymes and/or protein complexes, advances in artificial intelligence/machine learning (AI/ML) prediction of protein structure and protein-protein interactions, and de novo protein design for synthetic biology applications. Although these presentations mainly focused on advances in science projects led by each of the speakers, they also provided critical information to our discussion during the breakout sessions, which ultimately helped us identify key science, technology development, and Facility integration challenges. These are discussed in the overarching challenges and opportunities section.

Philipp Zerbe, University of California, Davis

Genomes to Structure and Function in Switchgrass Terpenoid Stress Responses

Dr. Zerbe's talk was focused on the study of plant natural products—terpenoids. These compounds make up the largest and most diverse metabolite group among plants. Only a few have been functionally characterized, showing roles in plant-environment interactions and stress resilience. Dr. Zerbe has shown that switchgrass has the largest gene family of terpene synthases among all analyzed monocot crops. Integrating multi-omics, biochemical, and genetic approaches to study the switchgrass gene family, he characterized the biodiversity and functional roles of obtained products. He found both terpenoid scaffolds that were common across species and those that were unique to switchgrass. His studies determined that diterpenoid compounds contribute to drought response in switchgrass roots.

Flora Meilleur, Carolina State University

Cellulose-Degrading Oxidative Enzymes: Functional Insights through Neutron Diffraction and Scattering

Neutron crystallography has a rich history of revealing details of enzymatic function by precisely localizing protons in functionally important positions, rather than only in the enzyme's active site. In small-angle neutron scattering, contrast matching allows the contribution of different components in the solution exposed to the neutron beam to be switched on and off. To highlight the unique features of neutron methods, Dr. Meilleur focused her presentation on x-ray and neutron studies of lytic polysaccharide monooxygenase (LPMO) and on a neutron scattering study of the LPMO cellobiose dehydrogenase (CDH) complex. LPMO studies revealed details of Cu^{2+} coordination in the active site. Small-angle neutron scattering data allowed creation of a model of LPMO with its redox partner, CDH.

Ivan Anishchanka (David Baker's group), University of Washington

Deep Neural Networks for Protein-Protein Complex Prediction and Design

The recent breakthrough in prediction of three-dimensional structure was led by groups from the DeepMind AlphaFold consortium and the group led by the Baker laboratory through RoseTTAFold. Dr. Anischanka's presentation emphasized the importance of elucidating the protein structure to enable deep understanding of function. RoseTTAFold's ML approach to de novo structure prediction was described in some detail, and the opportunity for developing models for higher-order complexes was discussed.

Breeanna Urbanowicz, University of Georgia

(Re)engineering Plant Enzymes to Alter Polysaccharide Structure

Dr. Urbanowicz's main research focus has been further development of improved feedstocks for biomass valorization, specifically, the biomass structure needed and the biochemical pathways involved. Dr. Urbanowicz re-engineered xylans, the second-most-abundant natural biopolymers (after cellulose), to alter their properties. She used Design-Build-Test (DBT) workflows that involved an array of tools from structural biology, computational biology, protein expression, and systems biology. One case included a plant enzyme, xylan acetyl

transferase (AtXOAT1). The atomic structure of this enzyme was first solved by x-ray crystallography and then used for computational design of enzyme variants. The variants were expressed and screened by *in vitro* platforms to identify optimal enzymes and generate a variety of xylan derivatives that had altered acetylation profiles and, therefore, altered architectural properties. These studies show the great potential of DBT workflows, but high-throughput pipelines are warranted.

Justin North, Ohio State University

Bacterial Organic Sulfur Metabolism: Mining Novels Genes and Enzymes for Bioproducts

In sulfate-limited environments such as lakes and rivers, bacteria use organic sulfur to synthesize methionine and cysteine. Using omics approaches, Dr. North discovered a novel pathway in this metabolism, described as a dihydroxyacetone phosphate (DHAP) shunt. In this pathway, a novel aldolase, ald2, cleaves methylthioribulose (MTRu-1-P) into DHAP and a methionine precursor, 2-(methylthio) acetaldehyde. One of the several distinct pathways to methionine biosynthesis that occur after that cleavage was found to yield an ethylene gas as a carbon byproduct. As ethylene is a widely used material in the plastics industry, Dr. North explored further to learn more about how bacteria synthesize it. A new cluster of multiple antibiotic resistance (mar) genes, previously misannotated as nitrogenases, was found to be responsible for generation of ethylene gas. This research was the first demonstration of ethylene biosynthesis under anaerobic conditions in bacteria and plants. Further research efforts focused on finding better-performing enzymes by mining metagenomes in combination with high-throughput screening. These efforts were successful, as more catalytically efficient aldolases were found.

Brian Fox, University of Wisconsin, Madison

Genes to Structure, Function and Application

Dr. Fox summarized the work his group has been doing in collaboration with the Great Lakes Bioenergy Research Center (GLBRC) and the JGI, focusing on the structure-function relationship of enzymes involved in plant biomass formation and degradation. Systematic functional and structural analysis of protein families allowed the group to create a consensus model of the active site associated with a specific enzymatic function, as has been created for the glycohydrolase clade GH5. Similar study of BAHD transferases, the enzymes involved in incorporating an ester bond into the lignin molecule, identified an enzyme that, when introduced into plants, increases hydrolysable p-hydroxybenzoate content in lignin by 40%. However, many challenges to further study of these enzymes remain, as the current workflows do not necessarily match the requirements of the studied enzymes and function cannot always be readily analyzed.

Crysten Blaby-Haas, Brookhaven National Laboratory (BNL)

Plant Biosynthesis by Design: Understanding Functional Space

Dr. Blaby-Haas introduced her group's work on systems biology grounded in molecular-level understanding of plant genes and their products. The work stems from realizing the scarcity of plant protein structures deposited in the Protein Data Bank (PDB) (only 3% of the structures in PDB are from plants), and the lack of system-level knowledge needed to engineer phenotype. An intriguing example of plant gene products is a heme protein of unknown function that is converted from a non-heme, zinc-binding protein. When a heme is added, the enzyme miraculously coordinates it by reorganizing zinc-binding site residues. This discovery has led the group to perform *in vitro* and *in vivo* engineering in cyanobacteria.

Chris Henry, Argonne National Laboratory (ANL)

KBase Tools for Integrating Multi-omics and Structure Data to Discover Protein Function

Dr. Henry provided an overview of the current capabilities of the DOE Systems Biology Knowledgebase (KBase) for integration of multiomics and structural data. He explained the concepts behind the mechanistic knowledge cycle, in which the knowledge map is completed by alternating between gathering experimental data and building models. The completed map can be used for predictive biology. Dr. Henry also discussed tools for incorporation of protein-structure-derived information, ongoing integration with the PDB, and the need to include both experimental and theoretical structures.

Elizabeth Kellog, Cornell University

Structural Basis of Target-Site Selection in RNA-Guided DNA Transposition Systems

Dr. Kellogg described her group's pioneering work on using DNA transposases as genome-editing tools using molecular biology and cryo-EM. Unlike CRISPR-based genome editing, DNA transposition involves no breaks in the DNA double strand during the cut-and-paste editing process. The group's work has shown T7 transposons precisely select their target sites, and ATP hydrolysis distorts the DNA double helix, leading to filament disassembly. This is likely to be generalizable to other transposition systems.

Justin Siegel, University of California, Davis

Discovery and Design of Novel Enzymes for Alkane Production

Dr. Siegel described a study focused on characterizing a long-sought microbial enzyme involved in alkene biosynthesis, aldehyde deformylating oxygenase. Collaborative work with the JGI that combined functional and structural analysis of a larger ferritin superfamily identified ADO activity not only in the ADO family, but also among representatives of most of the other families. Analyzing the features common among enzymes with ADO activity allowed the researchers to introduce such activity into an enzyme from an inactive family by extensive protein engineering. This work opens new opportunities for further improvements of catalytic efficiency among alkene-producing enzymes.

Tobias Erb, Max Planck Institute (MPI) Marburg

Re-designing CO_2 Fixation: New-to-Nature Enzymes, Synthetic Pathways and Artificial Cells for Sustainable Carbon Capture

Dr. Erb introduced his group's systems and synthetic biology studies on CO_2 fixation, which involvedengineering a novel *in vitro* (cell-free) system comprising enzymes from multiple species and structure-based protein engineering. The group began by studying the genetic, biochemical, and structural characteristics of enoyl-coA reductase/carboxylase (ECR). Based on these studies, they began developing the *in vitro* system using ECR as a starting point. These studies led to their pioneering work on the CETCH cycle, which incorporates 17 different enzymes from nine organisms and three engineered enzymes to realize synthetic CO_2 fixation pathways. The group's future directions include synthetic chloroplasts and transplanting the synthetic CO_2 fixation system into living organisms.

Session 2. Intracellular Organization, and Material Synthesis and Decomposition

Knowledge of the mesoscale to nanoscale organization and interactions of cellular assemblies is needed to understand critical processes in cellular metabolism and for developing approaches for the synthesis of bio-based materials. This session featured ten speakers with expertise linking genetic approaches with spectroscopic and microscopic techniques as a combinatorial approach to study environmental-relevant genes or microcompartments, functions of microbial organisms and biomolecules in their native biosystems, and to accelerate biodesign through cell-free approaches, thus help to understand molecular ecology, intracellular compartments and organization, and structural biology in microbial and plant systems for sustainable production of biofuels, bioproducts, and biomaterials. Although the presentations mainly focused on advances in science projects led by each of the speakers, they also provided critical information to our discussion during the breakout sessions, which ultimately helped us identify key science, technology development, and Facility integration challenges. These are discussed in the overarching challenges and opportunities section.

Stephen Burley, Director, Rutgers School of Arts and Sciences

Tools and Resources for Understanding Intracellular Organization in Three Dimensions at the Atomic Level

Dr. Burley discussed tools and resources for understanding intracellular organization in three dimensions at the atomic level. These include the Protein Data Bank (PDB), the first open-access digital data resource in all of biology. This resource is funded by DOE, NSF, and NIH and serves many millions of researchers, educators, and students annually. Open-access data is a cornerstone for developing AI/ML approaches to predicting protein structures, so PDB now makes 3D models for all proteins and their complexes available to the biology community. AI/ML approaches use PDB data in combination with genomic sequence information to predict protein structures with reasonable accuracy using AlphaFold2 and RoseTTAFold. These

models aid all structural biology approaches including x-ray crystallography, cryo-tomography, and cryo-EM. Dr. Burley provided several examples for important human proteins and complexes. Developing integrative/hybrid methods that combine experimental structural data with computational 3D models and proteomics provides enhanced understanding of large assemblies, such as a human nuclear pore complex, which is composed of over 1000 protein chains. These approaches also enable visualization of molecular machines within operating membranes. It is believed that connecting atomic-level experimental 3D structure data generated using different techniques with AI/ML computational methodologies will allow visualization of complex processes inside cells, as well as of whole cells. The PDB global archive plays an essential role in these developments.

Carolyn Larabell, Lawrence Berkeley National Laboratory (LBNL)

Quantitative 3D Imaging of Biological Specimens Using Soft X-Ray Tomography

Dr. Larabell discussed quantitative 3D imaging using soft x-ray tomography and its application to complex biological systems. This technique uses full-field transmission x-ray microscopy, taking advantage of synchrotron radiation from the Advanced Light Source research facility. High contrast is accomplished by operating x-rays in a "water window." Soft x-ray tomography allows 3D imaging of whole, hydrated cells in their native state. Cellular structures and morphology can be identified, and, when x-ray tomography and fluorescence images are correlated, even molecules can be observed. Data processing allows 3D reconstructions and observation of virtual sections of the internal components of cells. One model organism is *Chromochloris zofingiensis*, a unicellular green microalga that can produce high levels of lipids for biofuels and a valuable carotenoid, astaxanthin. Soft x-ray tomography can be used to rapidly map ultrastructural reorganization and inter-organelle interactions in intact cells, taking advantage of naturally occurring differential x-ray absorption of carbon-rich compounds in each organelle. Rapid and reversible changes have been observed in oxygenic photosynthesis, in the photosynthetic apparatus, in thylakoid ultrastructure, and in energy stores, including lipid bodies and starch. Understanding regulation of photosynthesis and metabolism in algae could enable bioengineering to reroute metabolism toward beneficial bioproducts for energy, food, and pharmaceuticals. These studies highlight the importance of a comprehensive and unbiased mapping at the mesoscale to characterize cell reorganization that would be difficult to detect with other existing methodologies.

Christopher Francis, Stanford University

Molecular Ecology and Structural Biology of Key C- and N-Cycling Enzymes of Ammonia-Oxidizing Thaumarchaeota

Dr. Francis discussed the molecular ecology and structural biology of key C- and N-cycling enzymes in ammonia-oxidizing archaebacterial species. These archaebacteria dominate many marine environments. Genes for ammonia monooxygenases (AMOs), typically associated with bacteria, were discovered in *Crenarchaeota* genomes by environmental sequencing. Of these genes, putative genes *amoA*, *amoB*, and *amoC* were linked to ammonia-oxidizing crenarchaeal (AMA) species. Gene *amoA* was isolated from many environments (marine, sediments, soil) known to recycle nitrogen. Sequencing *amoA* genes showed extensive diversity, with species in marine water columns forming distinct clusters and ecotypes. In soils, ammonia-oxidizing archaeal (AOA) species predominated over their bacterial counterparts. Dr. Francis's lab used an integrated structural, functional, and molecular ecology approach to characterize *amoA* and functionally relevant *nirK* (Cu-containing nitrate reductase) genes. Results linked diversity of enzymes to the enzymes' structure and function, and to environmental properties. Structural modeling allowed comparison of enzymes from different environments. The NirK enzyme is an attractive candidate for biochemical and structural studies. DNA synthesis was used to provide insight and to structurally characterize enzymes from the very efficient hydroxypropionate/ hydroxybutyrate cycle that is involved in CO_2 fixation. Combining advanced metagenomics, DNA synthesis, and structural biology approaches is critical for understanding the biology of nutrient-limited environments.

Cheryl Kerfeld, Michigan State University

Bacterial Microcompartments: From Metabolic Organelles to Biomaterials

The Kerfeld group uses a combination of bioinformatics, structural biology, and synthetic biology to study bacterial microcompartments (BMCs). The carboxysome

was presented as a prototypical example of a BMC. It performs the dark reactions of photosynthesis by sequestering carbonic anhydrase and rubisco in the interior of the microcompartment. Bioinformatics revealed that approximately 25% of bacterial genomes encode BMC proteins that perform a wide range of catabolic functions in cells. A common functional theme for BMCs is to entrap oxygen-sensitive enzymes or enzymes that produce volatile or toxic compounds. BMCs are composed solely of proteins that assemble from hexameric or pentameric building blocks. The symmetry axes of the homooligomers form the pores that allow transport of metabolites, and the particular amino acids that form the pores define the metabolite specificity. Dr. Kerfield pointed out that our understanding of how these proteins assemble to make BMCs has been aided by heterologous expression of empty microcompartments for structural studies. This ability has been further exploited to design BMCs for a variety of applications, such as forming platform materials for catalysis in confinement or enabling communication in synthetic communities. We can now consider a future functional genomics approach for designing BMCs for use as building materials. By starting with bioinformatics and structural characterization and following with synthetic biology, it will be possible to design tailored architectures for biocatalysis, biomaterials, and molecular transport with nano- to millimeter-scale organization for a variety of applications.

Bonnie Bartel, Rice University

Peroxisome Imaging in Arabidopsis

The Bartel group investigates a variety of peroxisome-related processes in Arabidopsis using a combination of genetic tools and fluorescence imaging techniques. Their interests include peroxisome biogenesis, peroxisomal degradation pathways, and interaction of peroxisomes with other organelles such as lipid droplets. Peroxisomes are membrane-bound organelles possessed by most eukaryotes. They are primarily known for fatty acid beta oxidation, but there are other oxidative processes sequestered in these organelles that support the biology of the organisms. In plants, they are important for seedling growth and development and pathogen defense. Dr. Bartel presented one research direction of her group: studying peroxisome membranes during peroxisome biogenesis. By combining different fluorescent tags, researchers can observe the formation of peroxisomes at the endoplasmic reticulum and the accumulation of lumenal proteins. These studies unexpectedly revealed that the peroxisomes also had internal membrane structures, which are called intralumenal vesicles (IVs). Time-resolved imaging showed that the large nascent peroxisomes accumulated more of these IVs over time and became smaller. This suggested that IVs are being formed from the outer membrane, an explanation that was then confirmed by genetic manipulation of the ESCRT (endosomal sorting complex required for transport) machinery. Furthermore, the group found that fatty acid beta-oxidation mutants were defective in formation of these vesicles, suggesting that the beta-oxidation machinery in the peroxisomes is important for stabilizing the IVs. This work highlights how using a combination of genetic tools and imaging can reveal metabolic processes in living cells.

James Evans, Pacific Northwest National Laboratory (PNNL)

Visualizing Cellular Ultrastructure with Cryo Electron Microscopy

Dr. Evans's presentation focused on the physical and technical challenges impacting the ability of cryo-EM to determine the structure of proteins and protein complexes and to visualize cellular organization. To attain near-atomic resolution in structures of interest using subtomogram averaging, tens of thousands of particles are needed, which is prohibitively time-consuming and demonstrates the need to speed acquisition rates. Tomogram montaging is one possible solution, but current computational and storage limitations make this approach impractical. These limitations also highlight the need for artificial intelligence and machine learning to assist in data analysis. To help identify structures observed in tomograms (which rely on electron density for contrast), multimodal approaches using CARS or SRS could be used.

Nikhil Malvankar, Yale School of Medicine

Seeing Is Believing: Novel Imaging Methods Help Identify Structure and Function of Microbial Nanowires

Dr. Malvankar's presentation focused on how microbes survive and communicate in the environment in the absence of oxygen or soluble molecules through nanowire appendages. He described how multimodal cryo-EM and near-field spectroscopic approaches helped to determine the structure and conformation of these

nanowires and how this knowledge transformed the understanding of nanowire-based electron transport, which occurs over micron distances at ultrafast rates. Using cryo-EM, the Malvankar group demonstrated that nanowires are made up of Omc cytochromes with hemes stacked together. Use of cryo-EM also showed that two different forms of pili, pili A-N and pili A-C, join to form low-conductivity structures. These structures are important for the secretion of cytochrome nanowires outside of *Geobacter*.

Lisa Miller, Brookhaven National Laboratory (BNL)

Chemical Imaging of Composite Systems: From Bones to Biofuels

Dr. Miller described how full-spectrum IR spectroscopy can be used to determine the structure and composition of biomaterials and biological systems and to provide a rapid and facile approach for chemical mapping and sample screening. Coupling chemical mapping with structural information is key to understanding metabolic processes in BER relevant biological systems. This analytical technique, therefore, is an important capability to have available to the research community.She highlighted several systems, including the examination of recalcitrance through reduction of acyl esterification in transgenic plants, the effect of light on lipid saturation for biofuel production in algae, the role of fungi in the recruitment of nitrate near root hairs in symbiotic fungal plant systems, rates of bone mineralization in systems in which osteoporosis is predominant, and secondary structure determination and formation of amyloid beta proteins in an Alzheimer model.

Michael Jewett, Northwestern University

Establishing Cell-Free Systems to Accelerate Design of Sustainable Biosynthesis Pathways

Dr. Jewett presented his group's ongoing work on using cell-free biosynthesis systems to rapidly optimize biosynthetic pathways for sustainability. He illustrated a cell-free system for identifying gene candidates that reduce acetone formation and/or result in byproduct formation. Candidate genes for genomic knockouts were identified when added to the acetone pathway in a cell-free protein-synthesis-based prototyping system. Acetone production in cells was increased by iterative knock-out of the identified gene candidates. Cell-free systems were also used to accelerate selection and optimization of biosynthetic pathways. A biosynthetic pathway was applied to isoprenoids using *in vitro* Prototyping and Rapid Optimization of Biosynthetic Enzymes (iPROBE) to build the pathway toward limonene synthesis. Cell-free protein synthesis was used to drive enzyme engineering and accelerate design-build-test-learn cycles. This acceleration can quicken the solution to optimized biosynthetic pathways that lead to the production of sustainable biofuel and bioproducts.

Jennifer Morrell-Falvey, Oak Ridge National Laboratory (ORNL)

Investigating Plant-Microbe Interactions Using Optical and Chemical Imaging

Dr. Morrell-Falvey introduced developments in optical and chemical imaging that increase researchers' ability to observe, quantify, and manipulate intact biosystems in order to characterize the functions of organisms and biomolecules in native systems. Dr. Morrell-Falvey's group designed 3D-printed chambers and microfluidic chambers to image the dynamics of plant-microbe interactions. They implemented chemical mapping using liquid extraction-mass spectrometry to non-destructively map chemical environments in-situ along microfluidic flow cells. This system revealed heterogenous amino acid distributions at specific root locations. After describing this work, Dr. Morell-Falvey then discussed combining optical and chemical imaging to identify interactions between organisms, in particular to identify interaction-induced metabolites. Transcriptional reporters and proteomic profiling are used to investigate the interactions' influence on microbial expression patterns. These combined optical and chemical imaging approaches, when applied *in situ* to the rhizosphere, can provide key insight into communications between microbes and plants which is necessary for understanding symbiotic relationships that maintain health for optimal plant production.

Session 3. Imaging the Rhizosphere and Cellular Organization

Improving our knowledge about microbe and plant activities in the rhizosphere is an important step toward understanding carbon and nitrogen cycling in environments and harnessing the power of soil-microbe-plant interactions to create sustainable production of energy crops. This session featured five talks on Day 1 and six talks on Day 2, highlighting a range of imaging methods to characterize various aspects of the rhizosphere at different length and time scales. Methods included time-resolved neutron imaging of water uptake, x-ray fluorescence microscopy for spatial mapping of trace elements in microbes, plants and soils, radio-carbon imaging, mass spectrometry analysis for following carbon and nitrogen flow, and computational modeling of nutrient uptake, as well as unique sample preparation technologies such as the EcoFAB and the Transparent Soil Model. Although these presentations mainly focused on advances in science projects led by each of the speakers, they also provided critical information to our discussion during the breakout sessions, which ultimately helped us identify key science, technology development, and Facility integration challenges. These are discussed in the overarching challenges and opportunities section.

Nikolay Kardjilov, Helmholtz Zentrum Berlin

Advanced Imaging of Root-Soil Interaction

Dr. Kardjilov presented his group's work on advanced imaging of root-soil interaction. Neutron imaging is a non-invasive technique that can allow researchers to visualize root physiology and track water flow in root and soil. Dr. Kardjilov described a first 3D time-resolved study using fast neutron tomography. With D_2O as a tracer, his group observed water infiltration in the soil and water uptake and transport in the root. Using a combined technique of neutron imaging and x-ray tomography, they got a complete picture of the rhizosphere and water distribution in the soil around the root. The combination of the two techniques in the same instrument to get dynamic imaging capability is planned at NeXT-Grenoble.

Si Chen, Argonne National Laboratory (ANL)

Cellular Trace Element Analysis with the Bionanoprobe

Dr. Chen introduced cellular trace element analysis with the bionanoprobe at the APS at ANL. The technique, a type of hard x-ray fluorescence microscopy, produces a series of 2D maps showing the element distribution in the sample. Dr. Chen presented a few examples, including a study on trace metal storage in a Chlamydomonas cell showing Mn sequestration in small substructures in the cell. The APS upgrade planned in 2 years will increase the brilliance of the facility by 100 times. The bionanoprobe facility will also be upgraded to deliver 10 nm spatial resolution with high-speed scanning capability. This opens new scientific opportunities including 3D imaging of microbes in soil aggregates, imaging of structural elements in wood cell walls at nanoscale, and imaging of element distribution and functions of subcellular organelles.

Rich Ferrieri, University of Missouri

Carbon Radioisotope Imaging in the Visualization and Measurement of Plant Resource Allocation to Root-Associating Bacteria

Dr. Ferrieri discussed the radio-carbon imaging technologies available at the University of Missouri. Carbon-11 is a radioactive isotope of C with a half-life of approximately 20 minutes. This short half-life allows researchers to study plant processes on the same plant routinely (as opposed to techniques that require large samples). The isotope is delivered to the plant by $^{11}CO_2$. The technique can also be applied to various metals (such as Fe) to study metal nutrient cycles. Multimodal imaging at relatively high spatial resolution (20 to 40 μm, compared with older techniques, which had mm-scale resolution) is performed by tissue back contrast filtered autoradiography, which was developed in the Ferrieri lab. The much-improved spatial resolution allowed imaging of root exudation and epifluorescent bacteria on plant roots. Similarly, green fluorescent protein (GFP) labeling revealed that the endophytic bacterium *Herbaspirillum seropedicae* preferentially colonizes the internal root endodermal ring of maize. *H. seropedicae* increases carbon input to the plant through CO_2 fixation. PET imaging shows that this carbon makes it through the plant to the root system. Dr. Ferrieri's group's method, combining fluorescence and radio-imaging, has shown that *H. seropedicae* has an increased need for plant carbon under low nitrogen conditions.

Jennifer Brophy, Stanford University

Synthetic Biology Tools for Engineering Both Sides of Plant-Microbe Interactions

Dr. Brophy discussed her group's research on how plant structure (form) can variously impact different plants' ability to tolerate environmental stressors. Specifically, Dr. Brophy uses biological engineering to change root structure development in *Arabidopsis*, which ultimately affects how roots acquire nutrients and water in the soil. This type of biological engineering requires precise spatial control over tissue-specific gene expression. Using synthetic genetic circuits and a series of logic gates, Dr. Brophy can control and select for specific gene expressions that will affect root development in *Arabidopsis*. Using green fluorescent protein (GFP) and a red fluorescent protein (mCherry), logic gate activity can be observed in different tissue layers of a 5-day-old *Arabidopsis* root tip by comparing outputs predicted by a model and images produced by confocal microscopy. Furthermore, this application can be used to engineer soil bacteria, since plants are considered holobionts and the entire microbiome impacts plant health and growth. Dr. Brophy will continue this research with bacteria and other plant species relevant to agriculture and bioenergy, then implement field tests.

Trent Northen, Lawrence Berkeley National Laboratory (LBNL)

Exploring Microbial Interactions in the Rhizosphere Using Fabricated Ecosystems.

Dr. Northen described EcoFABs, standardized rhizosphere ecosystems that are small microfluidic devices enabling reproducible studies of plants and plant-microbe interactions. EcoFABs provide a platform for improving plant microbiome research using reproducible habitats, defined microbiota, defined protocols, spatiotemporal measurements, standardized dissemination, and machine learning. Dr. Northen said new high-resolution technologies are needed to measure spatiotemporal dynamics of plant microbiomes. The focus of Dr. Northen's group thus far has been on the model grass *Brachypodium*. An interesting observation from their research using the EcoFABs is that when microbes are present, plants produce serotonin, which is important to new root development. Several researchers are working on various imaging techniques compatible with EcoFABs. The Northen Lab has also developed the EcoBOT, a robot that can host a large number of EcoFABs to enable automated, high throughput and standardized sampling and imaging.

Tiina Roose, University of Southampton

Multiscale Image-Based Modeling of Plant-Soil Interaction

Dr. Roose described recent progress in mathematical models to describe nutrient uptake in the rhizosphere. This work points to methods for providing soil amendments that will reduce the need for fertilizer. One challenge is that rhizosphere modeling must be multiscale, providing images at the scale of soil particles (using methods such as micro x-ray CT, NMR, x-ray fluorescence, and extended x-ray absorption fine structure) and information at crop scale (measured by crop yield, fertilizer runoff, and sensors), as well as every scale in between. Dr. Roose described a process for predicting uptake and release of nutrients in soil and roots involving imaging of root hairs using x-ray CT (10 microns in diameter and 100 to 1000 microns long), microdialysis, correlation with XRF elemental mapping of specific elements such as phosphorus or sulfur, and mathematical modeling.

Elizabeth Shank, University of Massachusetts

Imaging Microbes and Their Activities Using Transparent Soil Microcosms

Dr. Shank described development of the Transparent Soil Model, which she is using to build models that capture the main components of microscale processes in complex soil environments. The models are designed to characterize how microbes interact with each other chemically and physically, and how carbon is degraded and transferred throughout an ecosystem. The Transparent Soil Model allows non-destructive spatiotemporal characterization at the microbial spatial scale, and is compatible with the methods of fluorescence microscopy to identify microbial cells, Raman microspectroscopy to identify cells incorporating carbon, and MS to spatially localize metabolites. Dr. Shank has used this model to grow soil microcosms and characterize plant root growth and bacterial metabolic activity in the absence or presence of fungal hyphae, and throughout wet/dry cycles.

Ryan Tappero, Brookhaven National Laboratory (BNL)

Imaging Trace Elements in the Rhizosphere with X-Ray Fluorescence Microscopy

Dr. Tappero described a method of x-ray fluorescence imaging performed at the NSLS-II at BNL, specifically, the x-ray fluorescence microprobe (XFM) beamline. Scanning x-ray fluorescence microscopes can be used for in-situ nanometer-resolution x-ray microanalysis of complex, heterogeneous materials including soils and plants. Dr. Tappero emphasized that the program at the NSLS-II supports multimodal measurements, and contains a family of microscopes specialized in different wavelength regions and spatial scales. XFM, in particular, allows imaging and determination of trace element composition. XFM also allows use of x-ray absorption methods such as EXAFS and XANES, which can reveal oxidation states and speciation of absorbing elements such as Fe, Zn, and Ca at micron-level spatial resolution. Dr. Tappero described imaging of mycorrhiza, as well as chemical analysis of iron and trace elements bound to iron through symbiotic interactions between plants and fungi. Imaging and spectroscopy performed by Dr. Tappero's group at the plant-microbe interface revealed that host-specific ectomycorrhizal fungi affect breakdown of iron-coated grains, accelerating the "weathering" of iron minerals in soil as part of the symbiotic relationship.

Rachel Hestrin, Lawrence Livermore National Laboratory (LLNL)

Imaging Microbes and Nutrient Flows in the Rhizosphere

Dr. Hestrin described the nanoSIMS (nanoscale secondary ion MS) method combined with isotope tracking for characterizing and understanding carbon and nitrogen flow in the rhizosphere. The nanoSIMS method uses a primary ion source to eject secondary ions from a sample. These secondary ions go into a high-sensitivity MS. Rastering the primary ion beam across the sample creates a high-spatial-resolution map of elemental constituents. ^{13}C labeling then allows carbon to be tracked as plants photosynthesize and distribute carbon to roots and microbes, and ^{15}N labeling allows tracking of nitrogen from organic matter into the rhizosphere. Using isotope labeling with nanoSIMS, Dr. Hestrin's group found evidence that fungi help transfer plant carbon to soil. While the nanoSIMS method is high resolution and sensitive, it is also destructive and low-throughput, so the group's longer-term goal is to integrate nanoSIMS with other microscopy methods, such as 2-photon fluorescence and resonance Raman methods, together with single-cell sequencing in order to link genomes to structure and function in the rhizosphere and hyphosphere.

Kirsten Hofmockel, Pacific Northwest National Laboratory (PNNL)

Multi-Omic and Imaging Analysis of Rhizosphere Interactions

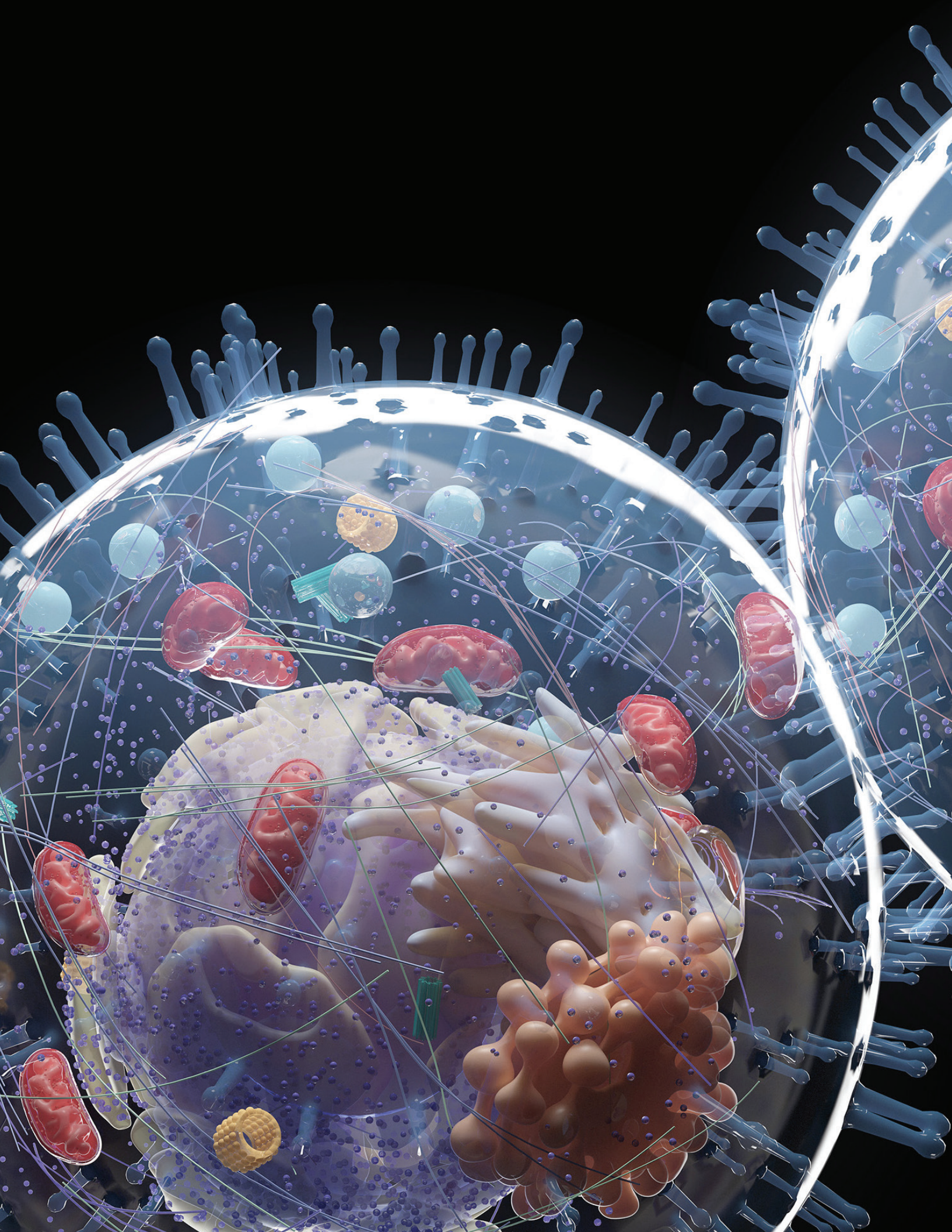
Dr. Hofmockel presented several omics and imaging capabilities for analyzing interactions within the rhizosphere. She emphasized that it is important to be able to scale from field observations to biological mechanisms occurring at the micron scale. One scientific question, for instance, is whether deep rhizosphere metabolism can promote carbon sequestration, especially in marginal soils. Experimental treatments in the field can start to answer this question and provide some data on how metabolites vary in response to wheatgrass treatments under normal or drought conditions. However, these data have low temporal resolution and are very low throughput. To improve resolution and increase throughput, Dr. Hofmockel's group is developing a suite of tools, including MALDI/MSI, ToF/SIMS, confocal microscopy, and activity-based probes, that can be applied on chip and using a SoilBox. The group also develops model soils with well-defined microbial members to help disentangle interactions. Dr. Hofmockel presented examples of using these tools to trace nutrient transport through soil using ^{13}C chitin to visualize spatial distribution of metabolites and to monitor changes in potassium chemistry. Many of these methods help researchers determine how spatial organization leads to interactions.

Robert Egbert, Pacific Northwest National Laboratory (PNNL)

Rhizosphere Biocontainment through Persistence Control Engineering

Dr. Egbert described the research area of rhizosphere biocontainment through persistence control engineering, with the long-term goal of bioengineering the rhizosphere environment to increase biomass activity, especially in marginal environments. His group is interested in understanding what governs

microbe persistence and resilience under stress and drought conditions and in the presence of pathogens in the rhizosphere environment. They are engineering microbial genomes to control the microbes' environmental niche. A central hypothesis for niche reconstruction is that genome reduction and metabolic addiction can be used to control environmental persistence. Dr. Egbert described one approach involving extracting persistence-control hosts from the sorghum root microbiome to identify metabolically active cells under different growth environments. He then discussed specific work to drive microbial growth on sorgoleone, a successful attempt to identify microbes that use sorgoleone as their sole carbon source, and efforts to assess the efficacy of the persistence-control species. To identify hotspots for plant-microbe interactions and complementation of persistence-control strains, a 3D cartography platform is used for imaging. The platform makes use of standard 3D printed "rhizogrids" in which sorghum can be grown, then employs x-ray tomography, metabolomics, and 16S sequencing to get spatial resolution of microbe metabolite interactions. Early results show that metabolic profiles cluster better to individual plants than to depth profiles, though there were also depth-specific trends.



BERKELEY LAB

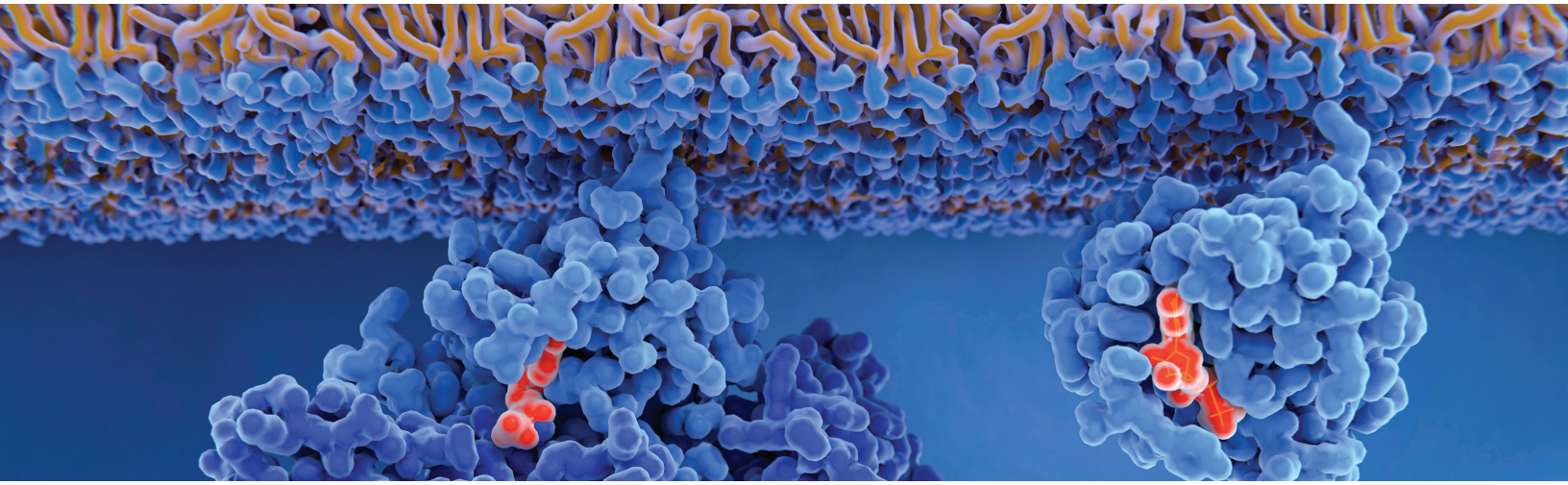
OAK RIDGE National Laboratory

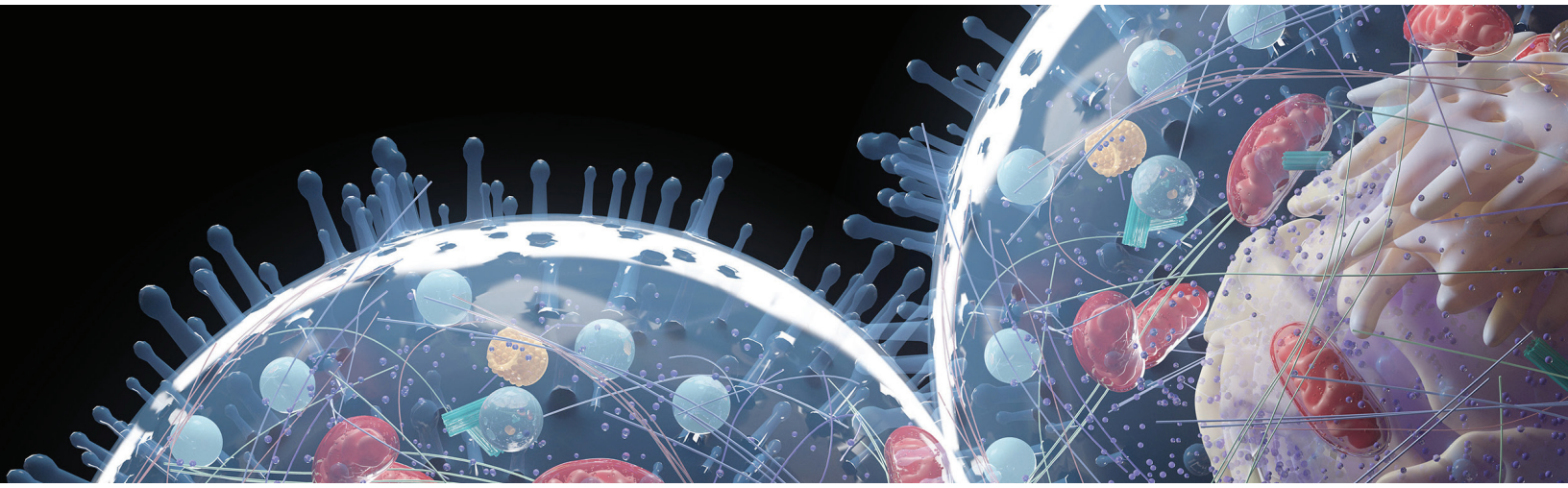
Argonne NATIONAL LABORATORY

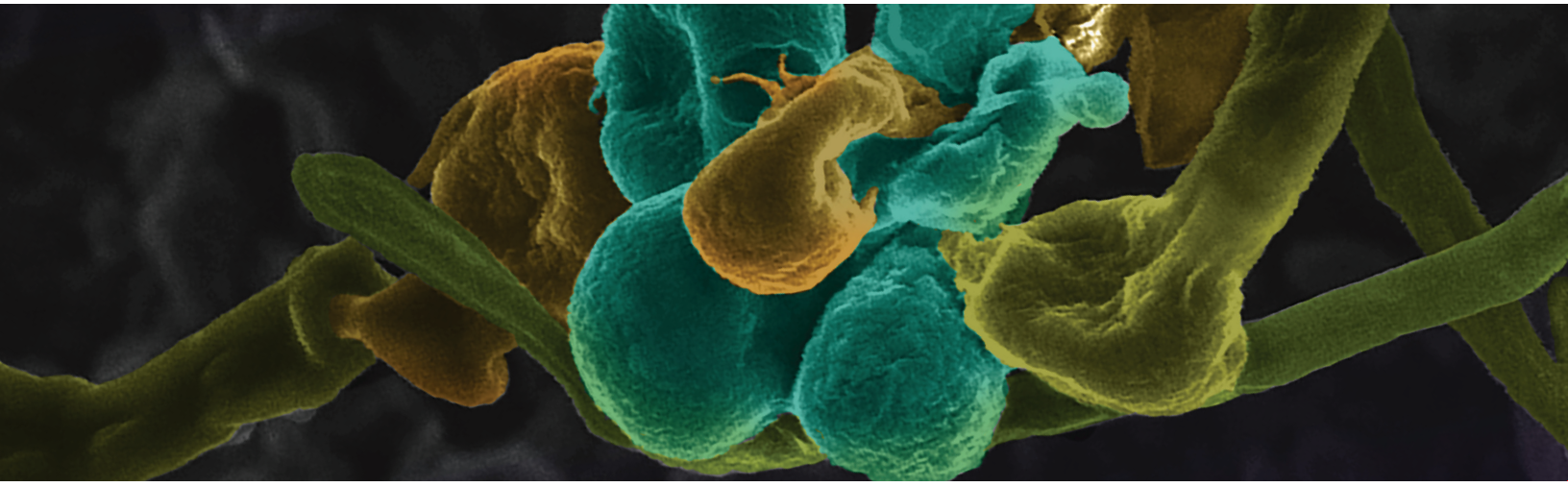
Pacific Northwest NATIONAL LABORATORY

Brookhaven National Laboratory

SLAC NATIONAL ACCELERATOR LABORATORY







22-JGI-23118