EMSL

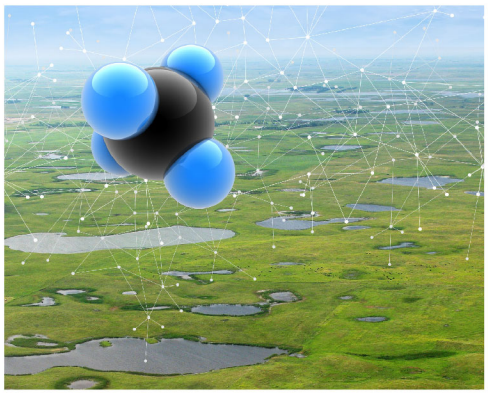
Breakthrough Science & Technology Workshop

Exploring the future of the Environmental Molecular Sciences Laboratory

January 10-11, 2017
EMSL
Richland, WA

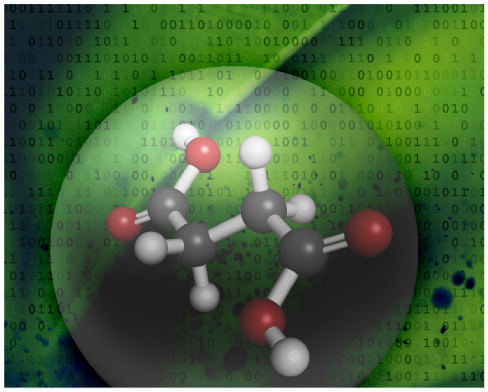
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965
U.S. DEPARTMENT OF ENERGY

About the cover



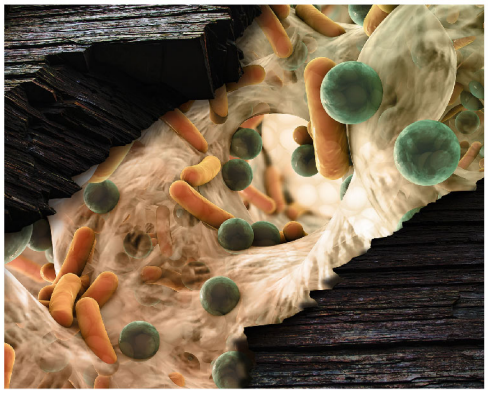
Microbes Drive Methane Release from Wetlands science highlight March 2, 2017

P. Dalcin Martins, D.W. Hoyt, S. Bansal, C.T. Mills, M. Tfaily, B.A. Tangen, R.G. Finocchiaro, M.D. Johnston, B.C. McAdams, M.J. Solensky, G.J. Smith, Y-P Chin, and M.J. Wilkins, "Abundant carbon substrates drive extremely high sulfate reduction rates and methane fluxes in Prairie Pothole Wetlands." *Global Change Biology* (2017). DOI: 10.1111/gcb.13633



How Organic Acids Form Atmospheric Particles science highlight May 11, 2016

G-L Hou, M. Valiev, and X-B Wang, "Deprotonated dicarboxylic acid homodimers: Hydrogen bonds and atmospheric implications." *Journal of Physical Chemistry A* 120(15), 2342-2349 (2016). DOI: 10.1021/acs.jpca.6b01166



Hydraulic Fluids Hospitable for Microbes science highlight September 6, 2016

R.A. Daly, M.A. Borton, M.J. Wilkins, D.W. Hoyt, D.J. Kountz, R.A. Wolfe, S.A. Welch, D.N. Marcus, R.V. Trexler, J.D. MacRae, J.A. Krzycki, D.R. Cole, P.J. Mouser, K.C. Wrighton, "Microbial metabolisms in a 2.5 km deep ecosystem created by hydraulic fracturing in shales." *Nature Microbiology* (2016). DOI:10.1038/nmicrobiol.2016.146

PNNL-26496

Breakthrough Science & Technology Workshop

January 10–11, 2017

Prepared for the U.S. Department of Energy's Office of Biological and Environmental Research under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.
(8/2010)

Executive Summary

The Environmental Molecular Sciences Laboratory (EMSL) is a U.S. Department of Energy (DOE) Office of Science (SC) Office of Biological and Environmental Research (BER) national scientific user facility at Pacific Northwest National Laboratory. The mission of EMSL is to lead molecular-level discoveries that translate to predictive understanding and accelerated solutions for biological and environmental research. Scientific capabilities within EMSL are organized into four science themes: Atmospheric Aerosol Systems (AAS), Biosystem Dynamics and Design (BDD), Terrestrial and Subsurface Ecosystems (TSE), and Molecular Transformations (MT). These science themes provide a framework and strategic direction for critical investments for cutting-edge research, the development of innovative capabilities to support that research, and the prioritization of user access. To fulfill its mission, EMSL must remain at the leading edge of scientific research, which requires continual reassessment of the needs of the scientific community that it supports, an evaluation of current capabilities, and recommendations for development and acquisition of future capabilities that EMSL should provide.

To assess both the current state of the art in the science that EMSL supports and to identify critical technologies and capabilities, a Breakthrough Science & Technology Workshop was held on January 10–11, 2017, at EMSL. Following an orientation focused on the fundamentals of EMSL science and mission areas, the participants broke into four groups, organized around the four science theme areas. There were also several group sessions that focused on general capabilities and cross-cutting themes. From a combination of the general and focused discussion meetings, each of the four groups made recommendations on both the scientific challenges that EMSL should address and the technical capabilities needed to meet those challenges (Figure ES.1).

Figure ES.1. Each year EMSL hosts approximately 700 users, supports approximately 350 projects, and produces over 400 peer-reviewed papers. With this level of relevance and contribution, workshop participants indicated the 20-year old building needs facility upgrades and instrument renewal to deliver new science opportunities.

Recommendations from the workshop participants follow for each science theme.

In the AAS Science Theme, the major scientific challenges identified were understanding the formation and evolution of aerosol particles, their interactions with clouds, and how aerosols affect cloud radiative forcing, lifetime, and coverage. Workshop participants identified three primary areas of research focus: 1) specific chemical reactions that lead to secondary organic aerosol formation and growth; 2) molecular-level processes involved in ice nucleation and how to represent ice-cloud formation in models; and 3) understanding the chemistry of the formation of sea spray particles. Scientific challenges that underlay significant breakthroughs include:

- Understand the origin, development, mechanisms of formation and roles of secondary organic aerosols in the atmosphere.
- Unravel the processes and characteristics that make certain particles efficient at nucleating ice.

- Include and incorporate compositional and mechanistic information into climate models to reduce uncertainties associated with atmospheric aerosols.
- Characterize the temporal, vertical, and horizontal distribution of sea spray particulates, and their chemical fate and transport.
- Improve understanding of the molecular mechanisms of the growth of atmospheric particles.

To achieve breakthroughs in these critical science areas will require new scientific capabilities, especially in the ability to:

- Measure reactive gas-phase species in real time (seconds) in experimental aerosol chambers/systems that can mimic real atmospheric conditions.
- Characterize molecular composition of precursor gases, intermediates, and the full range of products condensing to form secondary organic aerosols, including the ability to measure and observe particle primary structure and functional group location and transformation.
- Measure sodium chloride composition of particles, surfactants on particles, and ice nucleation particle characteristics.
- Develop controlled, highly instrumented atmospheric aerosol chambers that can operate over a range of temperature, relative humidity, and other atmospheric conditions.
- Develop integrated environmental/ecosystem chambers or facilities that allow study of soil/plant/atmosphere interactions and enable simulation and modeling of this system.

The workshop participants in the BDD Science Theme concluded that the major scientific challenge in this area was a deeper understanding of the relationship between the genome of an organism and its phenotype. Deep phenotyping and genome-to-phenome mapping of designed and “redesigned” biological systems is an area EMSL can pioneer and develop into a center of excellence. There is also an urgent need to understand how changes in biological systems at the cellular or higher levels of biological organization can impact large-scale ecological processes. This requires building mechanistic models of how the genomic information of organisms is translated to organism, community, and ecosystem phenotypes. The important science areas that should be addressed to achieve breakthroughs in these areas of research include:

- Define the mechanisms by which specific changes in the environment modulates the relationship between genotype and phenotype.
- Identify the specific regulatory and feedback processes that control biological functions at different scales (intracellular, intercellular, organismal, community, ecosystem).
- Understand the spatial and metabolic modularity of complex biological systems and understand how to represent this using quantitative models.
- Integrate knowledge and data across multiple scales and data types in a meaningful/searchable way.
- Understand the nature of biological unknowns (organisms, genomic regions, genes/proteins/metabolites of unknown function).

To achieve breakthroughs in these critical science areas will require new scientific capabilities. Especially needed are:

- Methods to integrate multimodal data that represent rates as well as states

- Assays for the functional phenotype of biological systems with sufficiently high throughput to support high-density sampling of experimental variables
- Single-cell technologies with sufficient resolution, sensitivity, and throughput to characterize the nature of phenotypic heterogeneity and cellular noise
- A data infrastructure for EMSL that would support the efficient integration of different types of high-throughput data
- Instrumentation for the spatial mapping of protein, protein complexes, and metabolites at the cellular and subcellular scale

There is also a need to develop model microbial communities to facilitate an understanding of how they work in their natural environment, at the molecular and the ecosystem scales.

In the TSE Science Theme, linking genomic information and biological phenotypes to higher-order ecosystem properties was also considered a fundamental scientific challenge. However, discussions were focused more on science at the ecosystem level. Solving scientific challenges in this area will require a scalable mechanistic understanding of complex natural systems and ecosystem feedback processes. This would be facilitated by new approaches for integrating data from field observations with those obtained from laboratory-scale investigations. Also important is understanding mechanisms controlling plant-microbe metabolic interactions and how those interactions respond to environmental changes. Participants identified the following challenges that EMSL should work toward solving:

- Link the genome of organisms to their phenotypes and the impact of these phenotypes on ecosystem properties.
- Understand how local, small-scale environmental properties and interfaces influence larger environmental processes and the ensuing feedback to the small scale.
- Incorporate microbial metabolism and thermodynamics into scalable hydro-biogeochemical models.
- Identify key biogeochemical processes and feedback mechanisms that result in observed ecosystem variations.
- Understand the regulatory interactions, metabolite flux, and dynamics both between and across microbes and plants, and the impact of system heterogeneity.

The new scientific capabilities needed to achieve breakthroughs in these critical science areas include:

- Improved technologies for following elemental pathways (C, N, S, P and metals) through plant-microbe-soil-atmosphere systems (e.g. stable isotope approaches)
- Developing methods for ^{13}C metabolic flux analysis resolvable to the level of individual microbial species within a community
- Multiscale simulations and supporting frameworks to support modelers/users; e.g. those developed in the BER- and Advanced Scientific Computing Research (ASCR)-supported IDEAS project
- Large-scale growth chambers to study natural complexity under controlled environmental conditions
- Field-deployable instruments to generate data to link laboratory experiments to environmental processes

The workshop participants in the MT Science Theme concluded that research in this area can play a significant role in the science supported by BER as well as multiple research agencies. A particularly important problem is understanding the molecular interactions and interfaces in biotic and abiotic systems

that control processes impacting earth systems (i.e., interactions of living cells, minerals and water). Linkages of molecular transformation research with work in the other EMSL science theme areas provide an opportunity for a more comprehensive approach that is not possible at most scientific institutions. Significant progress can be made by focusing on specific areas that include the dynamic nature of interfaces (e.g. biogeochemical), dealing with system heterogeneity, relating molecular processes to large-scale function, and understanding how to identify critical or controlling processes. The participants identified the following specific scientific challenges:

- Understanding the nature of the liquid phase near a surface/interface and its impact on molecular transformations
- Establishing the molecular mechanisms of heterogeneous reactions in large complex enzymatic systems (e.g. lignin breakdown)
- Determining the influence of mineral surfaces on the structure, dynamics, and reactivity of natural organic matter and how the physics and chemistry of these environments are coupled
- Understanding the influence of photochemistry on contaminant and carbon cycling
- Understanding why some long-range weak interactions have a major influence on molecular transformations at the cell-mineral interface

To achieve breakthroughs in these critical science areas will require new scientific capabilities, especially:

- Adding multi-instrument, multimodal capabilities for *in-situ* measurements with increased structural and chemical selectivity and sensitivity
- Improving spatial resolution and sensitivity under more realistic environmental conditions
- Delivering dynamic real-time measurements to understand the nature of interfaces and their surroundings
- Providing analytical techniques that can be applied to single-particle, ultra-fine aerosol, soil, and mineral particles as a function of history, transport, and environment (i.e. source to deposition)
- Increasing sensitivity of imaging and other current capabilities, especially those with sensitivity to see minor species in a complex environment, including solid/liquid interface

Several scientific challenges were noted across multiple themes, suggesting they constitute core challenges that should to be addressed by EMSL. Many of these require the development and deployment of advanced computational models that will require state-of-the-art computing platforms. Specifically, many workshop participants noted dealing with the complexity of biological, chemical, and environmental data and this could be facilitated by a user-friendly data integration, visualization, and management system. A significant challenge was also noted in building predictive computational models that link multiple spatial and temporal scales that characterize much of the research pursued by EMSL users. The synergistic use of experiments and computation is becoming increasingly necessary to understand the complex, multiscale systems studied by EMSL users, so advanced computational capabilities will continue to be essential.

Predictions comprise the heart of EMSL users' science: predicting phenotypes from genotypes, the impact of environmental perturbations, how changes in biological systems impact aerosol and cloud formation, and how changes in interfaces impact biochemical and chemical processes. Accurate predictions require a comprehensive and fundamental understanding of the mechanisms by which dynamic changes at the molecular scale propagate through interconnected biological and abiotic systems. This understanding requires improved instrumentation and experimental capabilities that are more precise, accurate, and

multimodal. Furthermore, it requires the ability to link laboratory experiments to field observations and to quantitatively trace the flow of energy and material through complex systems.

The suite of instrument capabilities and functional application areas (such as multi-omics, bio-imaging, single-cell dynamics, spectroscopic characterization, etc.) currently available within EMSL provides a foundation for the challenging work that lies ahead for BER-supported science. New and enhanced capabilities described in this report will position EMSL to deliver on new science opportunities. However, a recurrent issue throughout the workshop was the need for EMSL to stake out a scientific identity in an important area of research. Leadership in a specific science area would provide a clear example to the world of what the integrated research capabilities of EMSL are capable of achieving. This research identity, together with new capabilities and instrumentation, would greatly increase the impact of EMSL science. It would also extend and enhance the ability of EMSL users to more fully understand and predict the complex natural world.

Acronyms and Abbreviations

21T FTICR	21 Tesla Fourier-Transform Ion Cyclotron Resonance
AAS	Atmospheric Aerosol Systems
ACME	Accelerated Climate Modeling for Energy
ARRA	American Recovery and Reinvestment Act of 2009
ASCR	Advanced Scientific Computing Research
BDD	Biosystem Dynamics and Design
BER	Office of Biological and Environmental Research
BERAC	Biological and Environmental Research Advisory Committee
CESD	Climate and Environmental Sciences Division
DOE	U.S. Department of Energy
DTEM	Dynamic Transmission Electron Microscope
EMSL	Environmental Molecular Sciences Laboratory
FICUS	Facilities Integrating Collaborations for User Science
INP	Ice nucleation particles
LA-ICPMS	Laser ablation inductively coupled plasma mass spectrometry
MS	Mass spectrometry
MT	Molecular Transformations
nano-SIMS	Nanoscale secondary ion mass spectrometry
OA	Organic aerosol
PNNL	Pacific Northwest National Laboratory
SC	Office of Science
SIP	Stable isotope probing
SOA	Secondary organic aerosols
TEM	Transmission Electron Microscope
TSE	Terrestrial and Subsurface Ecosystems
VOC	Volatile organic compound

Contents

Executive Summary iii
Acronyms and Abbreviations viii
1.0 Introduction 1.1
1.1 Charge to workshop participants 1.3
1.2 Evaluation of science themes 1.3
2.0 Science Theme: Atmospheric Aerosol Systems 2.1
2.1 Summary of workshop discussions 2.2
2.1.1 Scientific and technical challenges 2.2
2.2 Summary of recommendations 2.4
2.2.1 AAS breakthrough science areas that were identified 2.4
2.2.2 AAS breakthrough capabilities required 2.5
3.0 Science Theme: Biosystem Dynamics and Design 3.1
3.1 Summary of workshop discussions 3.2
3.2 Summary of recommendations 3.4
3.2.1 BDD breakthrough science areas that were identified 3.4
3.2.2 BDD breakthrough capabilities required 3.6
4.0 Science Theme: Terrestrial and Subsurface Ecosystems 4.1
4.1 Summary of workshop discussions 4.2
4.1.1 Metabolite flow through subsurface ecosystems 4.2
4.1.2 Combining laboratory and field studies 4.3
4.1.3 Scalable models 4.4
4.1.4 Scientific workforce development 4.5
4.2 Summary of recommendations 4.5
4.2.1 TSE breakthrough science areas that were identified 4.5
4.2.2 Near-term investments 4.6
4.2.3 Medium-term investments 4.6
4.2.4 Long-term investments 4.6
5.0 Science Theme: Molecular Transformations 5.1
5.1 Summary of workshop discussions 5.2
5.2 Summary of recommendations 5.2
5.2.1 MT breakthrough science areas that were identified 5.2
5.2.2 MT breakthrough capabilities required 5.4
6.0 Workshop Cross-cutting Themes and General Recommendations 6.1
7.0 Conclusion 7.1
8.0 References 8.1
Appendix A – Workshop Agenda A.1
Appendix B – Participants and Breakout Assignments B.1

Figures

ES.1 Each year EMSL hosts approximately 700 users, supports approximately 350 projects, and produces over 400 peer-reviewed papers iii

1.1 EMSL's mission is to lead molecular-level discoveries for BER that translate to predictive understanding and accelerated solutions for national energy and environmental challenges 1.1

1.2 PNNL is advancing the frontiers of science and addressing some of the most challenging problems in energy, the environment, and national security 1.2

1.3 The call for proposals offered two tracks: EMSL's science theme research and Facilities Integrating Collaborations for User Science, or FICUS, with the Department of Energy Joint Genome Institute 1.2

1.4 Workshop participants at the Breakthrough Science & Technology Workshop at EMSL in January 2017 ..... 1.3

2.1 The AAS Science Theme focuses on understanding the chemistry, physics, and molecular-scale dynamics of aerosols for model parameterization to improve the accuracy of climate model simulations and develop a predictive understanding of climate 2.1

2.2 Researchers at PNNL and their collaborators used EMSL and other capabilities to improve atmospheric model predictions 2.2

2.3 Laboratory-generated sea spray aerosol particles were flash frozen to maintain their atmospheric configuration, and then probed by electron microscopy for a new avenue to analyze aerosol particles 2.3

3.1 The BDD Science Theme focuses on regulation of spatial and temporal parameters of metabolic processes in microbes, fungi, and plants 3.1

3.2 The DTEM is being developed to enable significant advances in biology, through studies of protein structures and enzymatic couplings, and environmental studies such as aerosol aging using real-time nanoscale imaging 3.2

3.3 A team of researchers explored single-cell biology and strategies for renewable energy production using brown rot fungi in aspen and spruce trees 3.3

3.4 The first science studies are underway on the 21 T FTICR 3.5

4.1 The TSE Science Theme focuses on the dynamics of nutrients, metabolites, and contaminants at biogeochemical interfaces in heterogeneous environments across multiple scales 4.1

4.2 By bridging the gap between small and large spatial scales, the simulation method could pave the way for more accurate assessments of the risk of groundwater contamination and the development of more effective remediation strategies 4.2

5.1 The MT Science Theme focuses on predictive understanding of molecular transformations in biology and chemistry central to energy production, bioconversion, and other processes that advance the other EMSL science themes 5.1

5.2 Image from a molecular dynamics study of protein interacting with a mineral surface in a water environment 5.2

5.3 The study reveals a newly discovered two-step mechanism by which brown rot fungi degrade plant cell walls 5.5

1.0 Introduction

The Environmental Molecular Sciences Laboratory (EMSL) is a U.S. Department of Energy (DOE) Office of Science (SC) user facility at Pacific Northwest National Laboratory (PNNL). Supported through the SC Office of Biological and Environmental Research (BER), its mission is to lead molecular-level discoveries that translate to predictive understanding and accelerated solutions for national energy and environmental challenges. (Figure 1.1) Since EMSL's opening 20 years ago, users and scientists have conducted groundbreaking research in multiple scientific disciplines using its suites of scientific expertise, state-of-the-art equipment, and mission-ready facilities. The scientific areas addressed by EMSL, as well as its technical capabilities, have evolved as BER and the scientific community have evolved. To maintain EMSL capabilities at the cutting edge of science, it is important to periodically reexamine its current state and direction, and adjust as necessary to better align these with the BER mission and science directions of the user community. The Breakthrough Science & Technology Workshop (Appendix A) described in this document is an important mechanism of EMSL's self-examination.

Figure 1.1. EMSL's mission is to lead molecular-level discoveries for BER that translate to predictive understanding and accelerated solutions for national energy and environmental challenges.

A decade ago, EMSL conducted a similar reassessment, which provided a critical foundation for its current technical and scientific excellence (EMSL 2008). Based in large part on the series of workshops held in 2006, EMSL was expanded to include the Quiet Wing, which houses a suite of advanced microscopes. EMSL also created the Radiochemistry Annex (now known as RadEMSL), designed to accelerate scientific discovery and deepen the understanding of the chemical fate and transport of radionuclides in terrestrial and subsurface ecosystems. Another outcome of this workshop was the development of EMSL's premier 21 Tesla Fourier-Transform Ion Cyclotron Resonance (21T FTICR) ultra-high resolution mass spectrometry capability, completed in 2015. Those workshops provided the blueprint for a $60 million DOE investment in new capabilities for EMSL from the American Recovery and Reinvestment Act of 2009 (ARRA). An important goal of the current workshop was to provide a foundation for future EMSL investments over the coming decade.

EMSL seeks to align with the science missions of DOE, particularly the BER missions. This includes providing the necessary fundamental science to understand, predict, manipulate, and design biological processes that underpin innovations for bioenergy and bioproduct production, and to enhance the understanding of natural environmental processes relevant to DOE. EMSL research also seeks to advance understanding of Earth's biogeochemical processes to enhance the ability to predict future trends and help plan for future energy and resource needs. These scientific aspirations and strategies drive development of new capabilities at EMSL.

To support DOE's research mission, EMSL has established several focus areas to concentrate its resources. These include:

- **Understand the foundational principles for predictive biology**. Gaining systems-level knowledge of genome-to-functional translations in cells and cell systems. Characterizing and analyzing–at the systems level–cells, cell constituents, cell organization, and processes to underpin a predictive understanding or redesign of metabolic and regulatory processes for sustainable bioenergy and environmental purposes.
- **Understand the role of natural and anthropogenic inputs to earth system processes**. Understanding fundamental molecular-scale properties of natural and anthropogenic inputs to improve predictions of key atmospheric and environmental processes. Measuring, characterizing, and analyzing fundamental molecular-scale properties to understand the dynamics of nutrients, metabolites, and contaminants at biogeochemical interfaces in earth systems.
- **Understand molecular transformation mechanisms that are central to bioenergy production, bioconversion, biocatalysts, and other biogeochemical processes in earth systems**. Examining subcellular molecular reactions that influence cell-to-cell, cell-to-mineral interfacial processes in bioconversion of bioproduced intermediates to low-greenhouse gas carbon-based fuels and gaining interfacial processes knowledge for molecular dynamic transformations in earth systems.

Figure 1.2. PNNL is advancing the frontiers of science and addressing some of the most challenging problems in energy, the environment, and national security.

EMSL and its host laboratory, PNNL (Figure 1.2), have strength and depth (scientific and operational) to address significant elements of these challenges, and a culture of multidisciplinary problem-solving to advance scientific goals.

Within each of these focus areas, EMSL has specific science themes: Atmospheric Aerosol Systems (AAS), Biosystem Dynamics and Design (BDD), Terrestrial and Subsurface Ecosystems (TSE), and Molecular Transformations (MT). These science themes provide a framework and strategic direction for critical investments for cutting-edge research, the development of innovative capabilities to support that research, and the prioritization of user access. These areas are where EMSL strives to focus its vision, leadership, and world-class capabilities to create scientific breakthroughs with substantial benefit for national priorities.

Figure 1.3. The call for proposals offered two tracks: EMSL's science theme research and Facilities Integrating Collaborations for User Science, or FICUS, with the Department of Energy Joint Genome Institute. Several projects are underway.

EMSL hosts a wide variety of technical capabilities with which it conducts cutting-edge research. However, EMSL also seeks to extend its research capabilities by actively collaborating with other user facilities. In particular, EMSL has a joint call with the DOE Joint Genome Institute to solicit proposals that exploit the capabilities of both facilities (EMSL 2016) (Figure 1.3). An ongoing pilot program between EMSL and the Atmospheric Radiation Measurement Climate Research Facility supports joint proposals that use both facilities. Other DOE user facilities provide additional resources in advanced computing, imaging, and structural biology. Still, EMSL is unique in the breadth of its capabilities and its ability to obtain and integrate a wide variety of different data to solve particularly complex problems. To maximize its future impact, EMSL strives to maintain its complementarities to the other DOE user facilities by generating unique data sets that can solve complex scientific

problems. A primary purpose of this workshop is to identify areas of research in which EMSL should focus to best advance DOE missions.

The vision outlined in this document is not an extrapolation of ongoing EMSL science, but rather a research community's view from the workshop participants. They suggested new areas of opportunity where advances in environmental molecular sciences and the tools that enable research can have a significant contribution. Furthermore, they looked at current and developing technologies that can be combined in new ways to create a new generation of research tools, which could provide a roadmap for the scientific capabilities within EMSL. The participants believe this could position EMSL to meet challenges in the environmental molecular sciences and have the greatest impact on current and future BER mission areas.

1.1 Charge to workshop participants

Participants were identified both from a pool of users and from suggestions of DOE program managers and EMSL staff (Figure 1.4). All participants were provided background materials and assigned to one of four science themes (Appendix B). They were asked to help EMSL address the following topics:

Strategy Challenges, Gaps, and Opportunities: Within the areas of science identified as part of EMSL's Strategic Plan, what are the most critical science gaps/challenges to be addressed over the next 10 years? What space can/could/should EMSL fill to maximize impact? Are our five- to 10-year leadership targets appropriate for EMSL, DOE, and the nation? How should these leadership areas and strategic directions be refined or refocused to achieve optimum impact?

Science Challenges and Opportunities: What are the barriers to addressing the challenges identified above? Are there specific barriers that if addressed could fill scientific gaps and accelerate progress? What science areas or expertise need to be cultivated to enable progress on the complex challenges?

Figure 1.4. Workshop participants at the Breakthrough Science & Technology Workshop at EMSL in January 2017

Technology and Capability Challenges: What types of investments should be made to address the challenges? What approaches (e.g. new capabilities, capability development, or expertise development/growth) are needed to break through the scientific and technical barriers to progress? What types of teams and skills need to be assembled to enable progress? What near- and long-term capabilities need to be developed or acquired?

1.2 Evaluation of science themes

Following the general overview of the EMSL strategy, the science themes, and the contributions desired, the participants broke into four pre-assigned science theme groups: AAS, BDD, TSE, and MT. Initial discussions were focused on critical science gaps in the EMSL science themes and leadership areas, key unanswered questions, the barriers to addressing the science challenges, and the expertise needed to overcome the barriers. Discussions involved EMSL scientists and external participants, with the aim of defining the most important challenges in which EMSL should focus during the next decade.

The next day, workshop participants were asked to define and prioritize the technical capabilities that would have the greatest impact on the science challenges. Capabilities included instrumentation, methodologies, domain experts, and computational infrastructure. Near- and long-term priorities were set. Following extensive group discussions, the workshop leaders prepared summary slides of their findings, which were then discussed with all participants and used to assemble a list of cross-cutting areas and general conclusions.

The following summaries of the four science theme discussions were assembled from notes and summary slides prepared during the workshop. They were then vetted by all the external participants to ensure accuracy and inclusion of all important points that were discussed. These summaries demonstrate strong support for science that is currently being conducted in EMSL in addition to providing insights and recommendations on new, impactful science directions, and the capabilities needed to achieve them.

2.0 Science Theme: Atmospheric Aerosol Systems

A primary focus for the Climate and Environmental Sciences Division (CESD) within BER is to develop predictive climate models that can support scientific evaluation of the impact of carbon emissions on global and regional climate (CESD 2012). A major challenge is to understand the formation and evolution of aerosol particles, their interactions with clouds, and how aerosols affect cloud radiative forcing, lifetime, and coverage. Important areas of research include organic aerosols (OA), the process of ice nucleation, and sea spray chemistry. Understanding their roles in the formation and properties of clouds will require molecular-scale descriptions extending from the emission and transformation of particles to how physical properties are related to chemical properties (Figure 2.1). For example, the molecular composition of aerosol particles must be known to predict their ability to act as ice and cloud condensation nuclei. In particular, it is difficult to quantify the chemical mechanisms responsible for OA formation, particularly secondary organic aerosols (SOA) arising from the atmospheric oxidation of volatile organic compounds (VOC) emitted by terrestrial ecosystems and by anthropogenic activities. There is growing evidence that growth of aerosol particles to climate-relevant sizes is primarily driven by SOA condensation (Ehn et al. 2014) and that SOA formation is enhanced by anthropogenic pollution related to energy production and use (Weber et al. 2007). Considering the large contribution of OAs to the total aerosol burden, it is important to understand their mechanisms of formation and determine their radiative properties to improve the accuracy of current climate model simulations (Zhang et al. 2007). Incorporating a molecular-scale understanding of the formation and optical properties of different types of aerosols and how they change throughout their life cycle would significantly reduce uncertainties in current climate simulations and would transform our ability to develop accurate predictive global climate models.

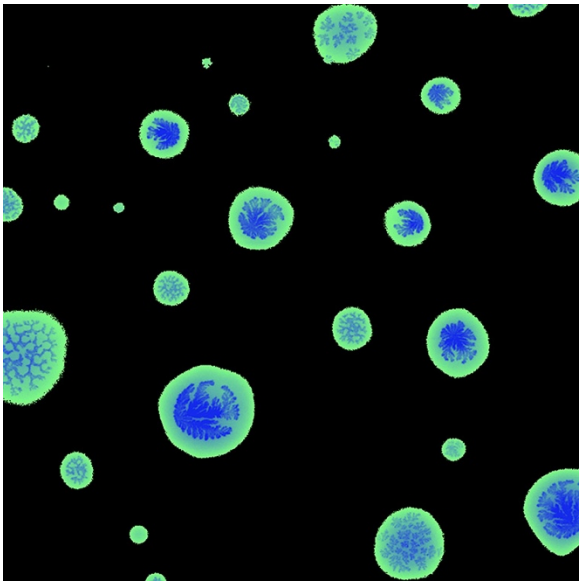


Figure 2.1. The AAS Science Theme focuses on understanding the chemistry, physics, and molecular-scale dynamics of aerosols for model parameterization to improve the accuracy of climate model simulations and develop a predictive understanding of climate.

EMSL's scientists and users focus on understanding and characterizing the chemical properties of aerosols and their link to climate-relevant properties (i.e. optical properties, molecular diffusivity, growth to cloud condensation nuclei sizes) by conducting laboratory studies, participating in field campaigns, and working with modelers to incorporate mechanistic information and constraints from these studies into aerosol and cloud model components that inform earth system models. A major focus of EMSL users is studies to determine specific constituents of aerosols and processes that control their formation, aging, and climate-relevant properties. The knowledge gained from these studies will be used to develop model parameterizations that capture the molecular-level understanding needed to improve DOE climate model simulations of aerosol formation, lifetime, and impacts on climate (Figure 2.2).

EMSL strives to support a world-class research program in aerosol science, particularly in OA. A long-term goal is to develop a molecular-scale understanding of ice nucleation, OA formation, and the impact of the emissions from plant ecosystems on chemical processes in the atmosphere. This knowledge will be necessary to understand the effects of anthropogenic and biogenic emissions (gas and particle phases) and their interplay in the global radiation budget. Achieving these goals will require a quantitative, molecular-scale understanding of processes controlling OA formation, the interactions of plants and OAs, brown

carbon, and ice nucleation. Success will lead to new, improved representations of these processes in climate models, including DOE's Accelerated Climate Modeling for Energy (DOE-SC 2015).

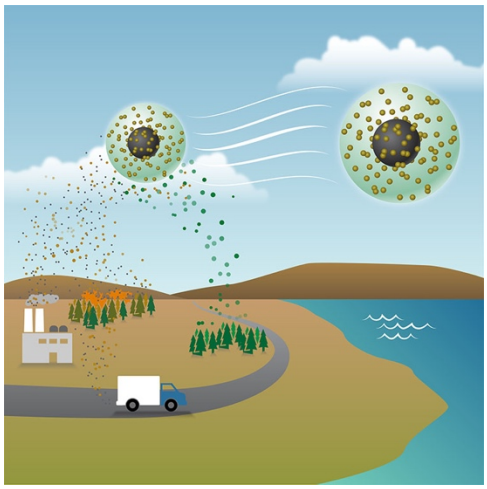


Figure 2.2. Researchers at PNNL and their collaborators used EMSL and other capabilities to improve atmospheric model predictions. For the full story, read Particles Carry Hitchhiking Pollutants Far from Home.

2.1 Summary of workshop discussions

Participants in this workshop identified three areas necessary to advance our understanding of atmospheric aerosol systems: SOA formation and growth, ice nucleation, and sea spray chemistry. Although challenges with each area have significant overlap, there are unique challenges in each.

2.1.1 Scientific and technical challenges

The first topic discussed was related to SOA formation and growth. SOA composes a major fraction of the total non-refractory submicron particle mass; however, there is an incomplete molecular-scale understanding of the specific chemical reactions that lead to SOA formation from VOC precursors, which include both biogenic gases emitted from plants and the ecosystem and gases emitted by anthropogenic activities (Kanakidou et al. 2005). Additional anthropogenic pollutants, such as nitrous oxide and sulfate aerosols, interact with these VOCs and their oxidation products to perturb the molecular chemistry and thus affect SOA formation. These processes regulate the growth of particles from the ultra-fine mode to the accumulation mode and thus affect aerosol optical properties and their ability to form cloud particles. The chemical composition of SOA will in turn determine the microphysical properties, including phase, morphology, and condensed-phase molecular diffusivity (Vaden et al. 2010), although the linkage between chemical and physical properties is still poorly understood. To date, most experimental studies have focused on SOA formation from single precursors, and additional research is needed to elucidate simultaneous reactions of ensembles of molecules, leading to the formation of SOA. Once specific chemical mechanisms important for SOA formation are understood, they will need to be parameterized in a computationally efficient way and implemented into climate and other 3-D earth system models.

The consensus of the panel was that the community is on the cusp of a breakthrough in this area due to the development of new capabilities to measure chemical composition of gas and condensed phase in real time (time scales of seconds to a few minutes) (Lopez-Hilfiker et al. 2014). However, significant challenges remain due to the dynamic nature of the gas-to-particle conversion process. To fully understand the system, precursor VOCs, reactive intermediates, and final products need to be measured, in both the gas and particle phases, in real time, and with minimal interference from the instrument surfaces. Products and intermediates are often short-lived, highly reactive, unstable, present in low concentrations, and difficult to introduce into instruments. Particle chemical composition is also believed to be highly size-dependent with different molecules contributing to particle growth at different sizes. Having multiple instruments co-located and measuring the same aerosol matrix during field campaigns and during laboratory studies would greatly enhance the knowledge gained from the experiments and assist in data integration and analysis. Laboratory campaigns should explore the impact of a wide range of environmental conditions, especially temperature and relative humidity.

The second topic discussed was ice nucleation, which is fundamentally a molecular-level process and should be studied as such. There are still large uncertainties associated with representing ice-cloud formation in climate models. Only a small fraction of the total aerosol particle population nucleates ice

crystals under a particular set of conditions and it is not clear what surface or chemical properties make these particular particles special (Hoose and Mohler 2012). Mineral dust, biological particles, fertile soil particles, SOA, and sea spray might all act as ice nucleation particles (INP), but the conditions that determine this are unclear. Some of these particles, such as mineral dust, are included in 3-D models as INP with various levels of specificity, but others, such as biological and fertile soil particles, are generally not considered as INP. As these particles are transported through the atmosphere, they undergo processing from atmospheric oxidation and coating with secondary species, such as SOA and sulfate. However, the effect of these coating on the particles' ability to act as an INP is not completely clear. Because these particles originate from a wide range of environments and include biological material, fertile soil particles, and dust, experiments that investigate INP will require cross-disciplinary collaboration among AAS and the other three science themes.

The technical challenges with ice nucleation are in many ways similar to those facing SOA-related research. Particles that act as INP represent a small fraction of the aerosol population, and analyzing them quantitatively, in real time, and with minimal interactions with instrument surfaces is very difficult. In particular, there are few to no instruments that can analyze the surface chemical and physical properties of aerosols in real time. Experiments probing the ability of particles to act as INP require precise control of temperature, relative humidity, and chemistry. Few chambers exist that can control all of these variables with the required precision. Characterization of the coating of primary particles with different secondary material is also needed to further understand how INPs age as they are transported.

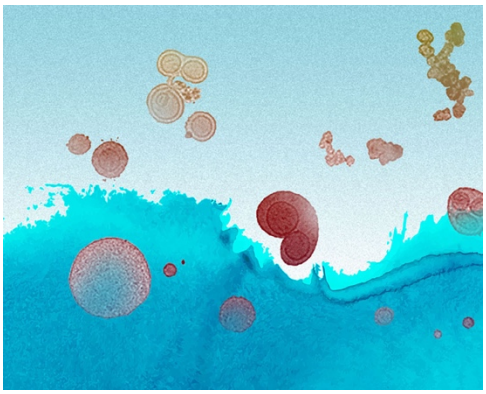


Figure 2.3. Laboratory-generated sea spray aerosol particles were flash frozen to maintain their atmospheric configuration, and then probed by electron microscopy for a new avenue to analyze aerosol particles (Patterson et al. 2016). For the whole story, read Freezing Sea Spray Aerosols to Study their Natural State**.**

The final science challenge the panel focused on was sea spray chemistry (Figure 2.3). Sea spray particles are significant because they may be the primary particle type over the oceans, a region that covers large fractions of the Earth and is particularly sensitive to changes in cloud condensation nuclei concentrations. Particularly over remote ocean regions far removed from anthropogenic influence, sea spray particles may exert significant influence on atmospheric chemistry, for example through heterogeneous chemistry that releases halogens to the atmosphere (Wang et al. 2015). Further effort should be made to understand the processing of sea spray particles as they are transported through the atmosphere. Particular attention should be paid to how organic coating forms and ages on the particles. The technical challenges are again similar to the SOA and ice nucleation areas. There is a need for *in-situ*, rapid, and quantitative measurement of the chemical composition of sea spray particles in real time. While bulk methods of measuring inorganic salts exist, few if any instruments are capable of measuring size-resolved chemical composition of both the inorganic and organic components of sea spray quantitatively and in real time, even for the major sea spray constituent sodium chloride. There are also relatively few measurements of the vertical and spatial distribution of sea spray particles.

In terms of capabilities, the panel noted that EMSL has relatively little capability to measure gas-phase species. Understanding the gas-to-particle conversion process is key to understanding the complete lifecycle of SOA. However, further investment in real-time instruments capable of measuring the molecular composition of condensed-phase organics in real time is also warranted. To be useful for these studies, instrumentation should be capable of *in-situ* sampling at the aerosol source, have time resolution on the order of seconds, be quantitative, and have minimal interactions between the instrument's sampling chamber surfaces and the analytes. There is also a need for size-resolved analysis of particle composition. For the purposes of OA characterization, high-resolution time-of-flight chemical ionization mass

spectrometry capability exists or can be developed or adapted to contribute to the atmospheric science goals of EMSL. Specifically, a new high-resolution chemical ionization mass spectrometer would enable characterization (identification and quantification) of trace biogenic and organic species in atmospheric samples. Instruments capable of measuring the size-resolved aerosol composition of both the inorganic and organic components of sea spray do not exist, but are necessary to advance understanding of sea spray. A basic capability to measure and monitor particulate sodium chloride concentrations in sea spray particles (not in the bulk) does not exist and development of methodology to do this would be invaluable to the community. Sample holders and stages capable of low temperature operation for ice nucleation studies could be developed for the Transmission Electron Microscope (TEM) to contribute to the understanding of INPs.

Many components of atmospheric aerosol are unstable and difficult to probe experimentally and thus need a combination of computational and experimental approaches. Computational prediction of aerosol characteristics such as vapor pressures, reaction potential energy surfaces, collisional cross sections, reaction rates, and molecular structures can contribute to the interpretation and validation of the molecular data from the experiments and instrumentation. The general feeling among the panel was that computational chemistry software packages such as NWChem, which was developed and is maintained at EMSL, are well suited to tackle these calculations. However, expertise to bridge the atmospheric and computational chemists to enable effective collaboration is largely lacking. The addition of aerosol and microphysics/chemistry modelers would help fill this gap and provide the ability to combine mechanistic models at the molecular scale with high-quality data. It would also be useful to have a research version of ACME (Accelerated Climate Modeling for Energy) focused on more molecular complexity. The resulting scientific contribution could underpin models at regional and global climate scales.

Finally, the panel discussed the possibility of EMSL developing a reaction chamber and/or environmental chamber facilities. A laboratory atmospheric chamber facility with highly precise control over temperature, relative humidity, and chemical conditions that is highly instrumented with molecular composition instrumentation would provide a breakthrough capability for AAS research and make a unique facility that would attract users to EMSL. First, the chamber should be operational at wide temperatures (200 – 300 K) and relative humidity (30 to greater than 100 percent) ranges. Such a chamber could be used for both ice nucleation studies and understanding the chemistry and phase transitions of SOA particles at cold temperatures. The panel also recommended that a potential EMSL chamber not focus on new particle formation, but rather particle growth phenomena. In the end, multiple chambers would potentially be necessary; e.g. a plant chamber to collect VOC emissions from live plants with a separate chamber to oxidize emissions. Also discussed was the need for an expanded chamber, or facility, of this type that would enable study of combined soil, plant, and atmospheric processes and effects. Several chambers and facilities of this type exist in Europe, but not in the U.S. Those in Europe often only monitor atmospheric conditions and constituents, and do not have the full range of molecular-level measurement capabilities that EMSL possesses. Thus a clear opportunity exists for EMSL to develop unique, controlled and instrumented atmospheric chambers as well as larger environmental/ecosystem chambers, or perhaps even a new facility combining experimental chambers or facilities at different scales, characteristics, and conditions. The panel suggested a future workshop could be held to further refine these concepts. This recommendation echoes a similar recommendation from the EMSL Atmospheric Aerosol Science Theme Advisory Panel workshop held in May 2014 (EMSL 2015).

2.2 Summary of recommendations

2.2.1 AAS breakthrough science areas that were identified

- Understanding the origin, mechanisms of formation and roles of SOA in the atmosphere

- Unraveling the processes and characteristics that make certain particles efficient at nucleating ice
- Including/incorporating compositional and mechanistic information into climate models to reduce uncertainties associated with atmospheric aerosols
- Characterizing the temporal, vertical, and horizontal distribution of sea spray particulates, and their chemical fate and transport
- Improving understanding of atmospheric particle growth phenomena

2.2.2 AAS breakthrough capabilities required

- Capabilities to measure reactive gas-phase species in real time (seconds) and implement in experimental aerosol chambers/systems that mimic real atmospheric conditions
- Capabilities to characterize molecular composition of precursor gases, intermediates, and the full range of products condensing to form SOA, including the ability to measure and observe particle primary structure and functional group location and transformation
- Design, development, and construction of controlled, highly molecular instrumented atmospheric aerosol chambers that can operate over a range of temperature, relative humidity, and other atmospheric conditions
- Design, development, and construction of integrated environmental/ecosystem chambers or facilities for study of soil/plant/atmosphere interactions to enable simulation and modeling of this system
- Expertise in aerosol and microphysics modeling to enable integration of mechanistic models at the molecular scale with high-quality data
- Integration of computational chemistry methods for calculating/predicting reaction mechanisms relevant to aerosol chemistry and optical properties of atmospheric chromophores with experimentally measured data

3.0 Science Theme: Biosystem Dynamics and Design

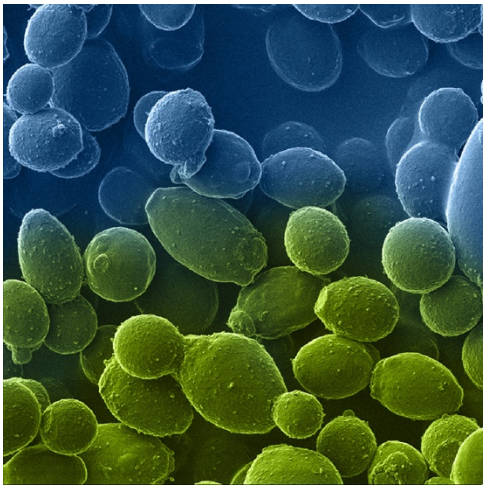


Figure 3.1. The BDD Science Theme focuses on regulation of spatial and temporal parameters of metabolic processes in microbes, fungi, and plants.

The Genomics Science Program at BER supports systems biology research identifying foundational principles driving biological systems of plants, microbes, and multispecies communities relevant to DOE missions in energy and the environment (BER 2015a). A major challenge is to understand dynamic interactions (e.g. among cells in a microbial community or within plant tissues) that regulate the function of living systems (Figure 3.1). This requires structural and functional understanding at multiple scales, including individual protein functions and macromolecular interactions with other cellular components, the regulation of pathways and dynamic networks in cells, and workings of cells in communities. Understanding foundational principles governing living organisms is expected to enable novel approaches for large-scale manipulation of functional properties that would not be possible via more traditional metabolic engineering approaches (Nielsen et al. 2014; Nielsen and Keasling 2016).

Leveraging EMSL's strengths in mass spectrometry and multimodal imaging and analysis, the BDD Science Theme is focused on understanding how biological systems are regulated in time and space by applying omics and microscopy measurements across multiple scales. For example, EMSL user projects have generated proteomics, metabolomics, and bio-imaging data with spatial and temporal information to understand microbial partnerships and interspecies metabolite exchanges (Bernstein et al. 2017), identify cellular compartments involved in synthesizing natural products in fungal cells (Boenisch et al. 2017), and the genesis of cellular compartments and their associated metabolic pathways (e.g. lipid droplets (Pomraning et al. 2015)). These data are used to develop new and updated metabolic models for microbes and plants in nutrient cycling, and biofuel and bioproduct production.

To move biology from observation to a functional understanding, EMSL strives to support research to understand the spatial organization and functional characteristics of plant and fungal organelles and bacterial/archaeal microcompartments. For example, *in-situ* studies of spatial and temporal changes to metabolism in microbial communities, the dynamics of bacterial microcompartments, and organelle formation and function will enable new approaches for optimizing microbial consortia and redesigning organisms for optimized or novel metabolic applications (Heinig et al. 2013; Frank et al. 2013). To enable predictive biology, EMSL users will need gene, transcript, protein, and metabolite data to model and simulate the current biological complexity of specific systems. Moreover, thermodynamic potential of a cell—determined by its redox state—is a crucial factor in determining whether reduced carbon compounds can be synthesized from precursors that are at a higher oxidation state. Modeling is required to use information from systems biology to facilitate synthetic biology. Models that include spatial localization are especially needed. The available flux-based models are valuable for initial estimates of activity of metabolic pathways, if growth rates and biomass measurements have been made. However, computational models that can predict biomass production and growth rates from genome data and environmental conditions are greatly needed.

With BER support, EMSL scientists have been pioneers in proteomics using both "bottom-up" characterization of the protein composition of cells and "top-down" characterization of combinatorial posttranslational modifications of intact proteins. The multiple mass spectrometry-based imaging capabilities at EMSL have supported users in defining the spatial distribution in biological systems for a

variety of molecules. EMSL has developed capabilities to investigate primary mechanisms that regulate cell activities, including gene transcription and protein

modification, with planned extension to single cells to address questions related to heterogeneity within cell populations. BER is funding the development of additional bioimaging capabilities to enable combining structural information on biomolecules with visualization techniques and computer simulation to potentially obtain a deeper understanding of cellular metabolism or engineered metabolic processes occurring *in vivo*. In the future, EMSL will endeavor to build visual proteomics expertise that leverages and combines its microscopy and proteomics capabilities to enable visualization of proteins and protein assemblies within the whole cell context. To do this will require advances in structural biology, cryo-electron microscopy and tomography, native mass spectrometry, and their integration using advanced informatics and computational approaches. The development of the Dynamic Transmission Electron Microscope (DTEM) and future ultrafast microscopy capability will enable dynamic *in-situ* observation of macromolecular systems at nanoscale spatial and temporal resolutions to visualize conformational changes in real time to reveal molecular mechanisms and inform protein engineering for synthetic biology and biodesign, e.g. of isozymes, transcription factors, and transporters (Figure 3.2).

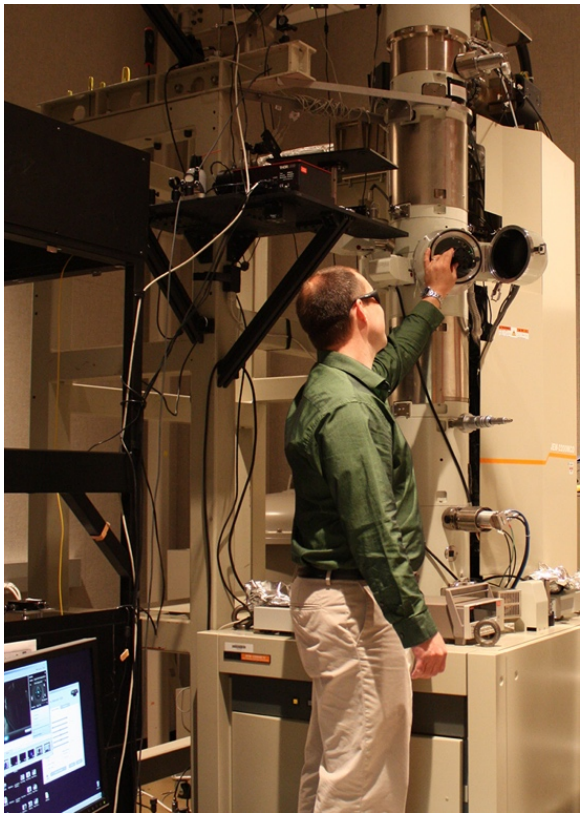


Figure 3.2. The DTEM is being developed to enable significant advances in biology, through studies of protein structures and enzymatic couplings, and environmental studies such as aerosol aging using real-time nanoscale imaging.

Measurements of metabolite levels (metabolomics) and their localization are used to determine the rate of change and refine predictive models of plant and microbial metabolism (both single-cell models and community models). Advances in metabolomics measurements and approaches that can circumvent the traditional bottlenecks of standards availability and identification uncertainties are desperately needed. New standard-free metabolomics and stable isotope probing (SIP) techniques being pioneered or developed at EMSL add promise that metabolite identification, quantification, and flow information can soon be realized in higher-throughput fashion for many biological systems.

3.1 Summary of workshop discussions

BER-relevant science is focused on processes that work across vast scales (BER 2014b, 2015a). To understand this type of scientific problem requires a systems perspective and working with extremely large and complex data sets. Growth of biological data is outpacing the ability of scientists to interpret it. Thus, data analytics are of increasing importance to science in this area of research.

There is a strong interest in developing a functional understanding of communities, especially microbial communities—how organisms work in their natural environment, at both the molecular and the ecosystem scale. Important questions include how to identify the keystone species in an environment. What are the primary metabolic roles of those organisms and how can meta-omics tease out those relationships? One important challenge is to identify good model communities that support investigations of global processes, yet are simple enough to be experimentally tractable.

DOE aspires to understand important global processes at the molecular level because this information would support future efforts to reengineer those systems. One important challenge is how to go from observational to predictive science at the molecular scale. The key is building models that can be tested, but this requires ways to simplify models so they are computationally tractable. Synthetic biology is a powerful approach for testing fundamental concepts in biology, but one still needs to be able to measure the phenotype/function of these systems. There was a discussion on whether building synthetic biology capabilities to complement those already available at the DOE Joint Genome Institute would be appropriate for EMSL, or would a better focus be on measuring the phenotype of modified systems? One possible solution is to develop synthetic consortia for the scientific community to use as experimental test-beds. Another would be to tailor measurement capabilities to encompass both synthetic or bioengineered cells and cell-free systems in addition to phenotyping and characterizing environmental samples and communities.

EMSL should continue to support biological research with single-cell resolution because it provides a foundation for understanding the basis of natural system variation (Figure 3.3). Organisms often show phenotypic heterogeneity, which could be a strategy of "hedging bets" on which phenotypes are optimal for a given environment. Thus, there is not a single genotype-phenotype relationship (Forsman 2015). The mechanisms that cause such phenotypic diversity are not well understood, yet are critical for reengineering biological processes. Phenotypic heterogeneity also needs to be understood in the context of heterogeneous environments, which are highly characteristic of terrestrial systems. Are there mechanisms that control population responses, such as feedback or cell signaling (e.g. quorum sensing)? What is the best way to determine the presence of such processes and their underlying mechanisms? Extending both multi-omics and imaging technologies to the single-cell level could be a tremendous advance for characterizing phenotypic heterogeneity in microbial systems and plants, and addressing such fundamental scientific questions.

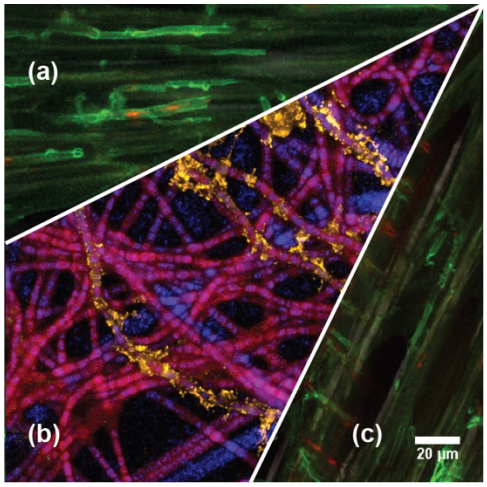


Figure 3.3. A team of researchers explored single-cell biology and strategies for renewable energy production using brown rot fungi in aspen and spruce trees (Zhang et al. 2016).

Connecting genotype to phenotype in a predictive manner will require new approaches to measure phenotype variations across different spatial and temporal scales (Lehner 2013). These measurements could be continuous, allowing assessment of response dynamics, or discontinuous, allowing sampling of wide ranges of different environmental conditions or genetically modified organisms. New technologies have greatly increased the sensitivity, depth, and coverage of analytical measurements (e.g. nucleic acid sequencing, metabolomics and proteomics), but there is a need to increase their throughput. As the ability to modify the genetic structure of cells rapidly accelerates because of technologies such as CRISPR/Cas9, there is an increasing need to rapidly screen for the resulting cellular phenotypes (Estrela and Cate 2016). There are always tradeoffs between sensitivity, coverage, and throughput, so scientific priorities are needed to define future capabilities. Nevertheless, new technologies should be sought that improve both sensitivity and throughput. During the discussions, there seemed to be strongest support for increased throughput to support screening approaches. High-throughput screening of different genetic variants of cells or cell systems, such as microbes, plants or microbial consortia, at the physiological level would also allow selection of subsets of specimens for more detailed phenotyping using EMSL's molecular and imaging capabilities.

To connect the genome to phenotype, analytical approaches that can inform across scales are needed (Baldazzi et al. 2012). How do transcripts cause specific protein levels, how do these proteins assemble into complexes, and how do these complexes shift the energy balance of cells or modulate cellular

interactions in communities? This will require high-throughput RNA measurement technologies and spatially resolved measurements of mRNA, proteins, and metabolites. This could be accomplished by multimodal imaging technologies in conjunction with novel approaches to increase their spatial resolution (e.g. expansion microscopy; Chozinski et al. 2016), or using multiplexed ion beam imaging (Angelo et al. 2014). Urgently needed are new analytical technologies that support high-density sampling with high sensitivity and throughput.

DOE research is focused on the fundamental science that underlies an emerging "bioeconomy" (NRC 2015). In this area, biocatalysis (the use of biomolecules to speed up chemical reactions) is likely to be a significant research area that could be addressed by new EMSL capabilities. Work in this area would help connect synthetic biology, structural biology, and the catalysis expertise at PNNL and link the BDD and MT Science Themes. EMSL would need to help develop the community for this area and identify the capabilities needed to drive the science. One area of potential development is surface chemistry and using proteins as scaffolds for work on complex biochemical and metabolic conversions *in vitro*. This obviously would cross over to the other science themes.

An attractive way of increasing EMSL's impact is enhancing the use of the data it generates. One of the challenges that the BDD research area faces is analyzing and interpreting the increasing amounts and complexity of its data. EMSL should provide more services in data management and interpretation, and ways to connect it to other work in the field. This will require expansion of EMSL's expertise in data visualization, implementing new software and hardware systems to support metabolic modeling studies, and developing new algorithms relating metabolite dynamics to changes in enzyme abundance. In particular, there is an urgent need for improved platforms for managing and interpreting metabolomics data. There is also a need for more expertise in organismal physiology and molecular/metabolic function to help analyze and interpret the data.

Finally, a system for capturing and managing sample metadata in EMSL was considered a significant priority. Data management is obviously challenging because of the wide range of different data types generated in EMSL, but a robust system is urgently needed. There was also significant discussion on the need for increased interaction with EMSL scientists for data interpretation as well as for discussing results. Ideally, EMSL should develop an easier way to share data in a cohesive fashion, for example, to overlay proteomics, sequencing, and metabolomics on a single interface.

3.2 Summary of recommendations

3.2.1 BDD breakthrough science areas that were identified

3.2.1.1 Scientific challenges

One of the greatest challenges faced by BER scientists is understanding at the molecular level how biological systems impact large-scale processes. Meeting this challenge will require building mechanistic models of how the genomic information of organisms is translated to organismal, community, and ecosystem phenotypes. To accomplish this, EMSL should identify and focus on a specific, representative scientific problem on which it could build an international reputation.

Broadly, EMSL needs:

- Approaches to understand biological complexity at both the spatial and temporal level
- Improved approaches to represent/validate complex biological systems with computational models

- The ability to define/identify the key (limiting) molecules and processes necessary to model and understand biological systems
- Ways to represent the multiscale nature of biological systems in a useful way
- Integration of multimodal data that represents rates as well as states

There are many key unanswered scientific questions related to building multiscale models of biological systems. These include:

- What is the nature of biological unknowns (i.e., unknown functions in organisms, genomic regions, genes/proteins/metabolites)?
- What is the relationship between genotypes and phenotypes, and how does the environment change this relationship?
- How can one predict structure-function relationships at the molecular level?
- How are the different components of a cell, organism, or community regulated and coordinated? Who interacts with whom, including biomolecules and organisms?
- What are the functions of membrane protein systems, and how they are regulated?
- How can one integrate knowledge and data across multiple scales and data types in a meaningful/searchable way?
- What is the nature of the interfaces that connect organisms with each other and the abiotic environment, and what are the molecular mechanisms that facilitate energy, material, and information transfer?

3.2.1.2 Technical challenges

Enormous technical challenges must be overcome to understand biological complexity. EMSL is well positioned to tackle many of these challenges with the breadth of its instrumentation as well as its expert staff (Figure 3.4). However, there are specific gaps in its portfolio of capabilities, including:

- Phenotypic assays with sufficiently high throughput to support high-density sampling of experimental variables, particularly the extension of conventional imaging instruments to include more molecular-level measurements
- Single-cell technologies with sufficient resolution, sensitivity, and throughput to understand the nature of phenotypic variation
- Data infrastructure that would allow reliable integration of high-throughput data
- Being able to measure dynamics of biological systems at the appropriate time intervals (i.e. less than its characteristic or reaction timescale)
- Spatial mapping and visualization of protein and metabolites at the cellular scale (<1 µm)

Figure 3.4. The first science studies are underway on the 21 T FTICR. These studies and resulting insights could lead to better-informed technical and policy decisions affecting environmental remediation, waste processes, energy production, and associated health impacts.

- Rapid and high-throughput approaches for characterizing and imaging macromolecular complexes, interactions and dynamics

Although EMSL staff have a broad range of technical expertise, there are several science and technical areas that could be enhanced with additional staff or training. These include:

- Molecular modeling
- Data management and integration and interpretation
- Biological engineering (biodesign)
- Structural biology (cryo-electron microscopy and native mass spectrometry)
- Plant and microbial physiology
- Biocatalysis

One suggestion was to create a limited-term program for technically astute scientists to bring new capabilities to EMSL. The EMSL Scientific Partner Proposals could provide the process to accomplish this if funding levels permit. There were also significant discussions about how the user community could most productively interact with EMSL scientists. There is a need to educate users on how they can access and use the complex technologies in EMSL to solve important biological problems. One approach would provide consultation services by EMSL scientists through social media.

3.2.2 BDD breakthrough capabilities required

Technologies that are driving scientific progress in biology are being developed at an ever-increasing rate. To remain at the forefront of research, it is necessary to incorporate new capabilities as rapidly as possible. However, EMSL also needs to update existing bio-imaging, mass spectrometry and NMR capabilities to stay current with new technology. These can expand the range of functional studies that can be done while reducing costs and increasing data quality. For instance, a new generation of thermally and mechanically stable cryoTEM with automated sample loader and pondermotive phase plate would enable high-throughput structural biology studies. Coupled with EMSL's computational and mass spectrometry capabilities, this capability would enable quantitative visual proteomics and structural imaging of systems relevant to BER. Advanced mass spectrometry technologies (e.g., the newest generation Orbitrap and SYNAPT technology) have enhanced sensitivity and throughput, both critical in multi-omics applications and enable more efficient characterization of protein complexes. Similarly, EMSL transcriptomics capabilities need to be upgraded as existing ones are becoming obsolete. It is essential for EMSL to keep current in all these critical technologies.

Specific recommendations for capability investments in the BDD Science Theme include:

- Instrumentation for high-content, high-throughput phenotype screening (e.g. automated imaging systems); field-deployable instruments
- High-throughput mass spectrometry-based multi-omics instrumentation capable of ~100 samples per day per instrument; high throughput for discovery (depth) or for screening (narrow analyses over many samples)
- Extension of multi-omics measurements to the single-cell level to specifically address phenotypic heterogeneity, community interactions, and the role of the microenvironment
- High-throughput characterization and visualization of protein complexes/molecular machines in the cellular context (visual proteomics)

- Automated single particle cryo-transmission electron microscopy for high-throughput structural biology of protein complexes in their native state without requiring crystallization and advanced native mass spectrometry capabilities to study molecular machines (e.g. protein complexes)
- High-throughput automated sample preparation pipelines
- Bioinformatics pipelines to support high-throughput multi-omics (especially metabolomics and integrative omics) and multimodal imaging
- An interactive data management system
- Cutting-edge metabolomics capabilities, including expanded compound libraries, development of standards for metabolomics, and development of standard-free approaches
- Label-free ultrastructural 3-D imaging of intact cells up to 15 µm thick without chemical fixation or physical sectioning via addition of X-ray nanotomography capabilities to address a critical gap for 3-D probing that exists with current EMSL capabilities

Enhanced computational capabilities are also needed to support the anticipated role of EMSL in advancing science relevant to DOE mission areas. These include:

- Approaches and workflows for integration of multi-omics data at different levels of granularity
- Improved software for modeling enzyme active site reactions at the molecular level
- Simulations at all scales (molecular, network, cell, organism, etc.)
- High-throughput image analysis pipelines
- Image data management and access system as well as improved software for integration of multimodal imaging
- Combined experimental-theory approaches to eliminate characterization barriers; current work exploring computationally assisted metabolomics (using collisional cross-section analysis/modeling to reduce the requirement for individual metabolite standards)

4.0 Science Theme: Terrestrial and Subsurface Ecosystems

This science theme supports BER research to advance a robust predictive understanding of terrestrial surface and subsurface ecosystems, extending from the bedrock to the top of the vegetated canopy and from molecular to global scales (Figure 4.1). Specific BER-supported research is focused on understanding the biogeochemical interactions and transport processes involving nutrients and contaminants in diverse landscape systems (CESD 2012). This information is being used to tackle the grand challenge of understanding process interdependencies and how these processes quantitatively scale from the molecular to ecosystem levels. Success in this research endeavor will require bold new research approaches aimed at linking global-scale climate phenomena, ecosystems, biogeochemical processes, microbial communities, and functional activities encoded in the genomes of plants and microbes (BER 2014c). Understanding how multiscale heterogeneity and environmental perturbations impact ecosystem responses represents an additional challenge.

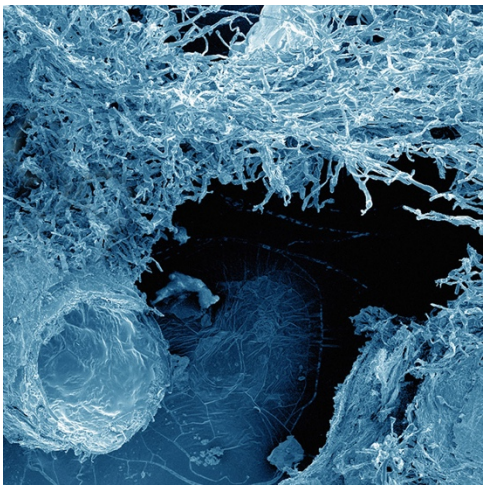


Figure 4.1. The TSE Science Theme focuses on the dynamics of nutrients, metabolites, and contaminants at biogeochemical interfaces in heterogeneous environments across multiple scales.

The response of ecosystems to environmental change represents the collective responses of many organisms whose functions are encoded in genomes. Although fully understanding how localized genomes dictate the global response of ecosystems is not yet achievable, it is possible to establish connections at intermediate scales using new systems biology approaches that combine powerful analytical technologies and modeling techniques (BER 2014c). Expanding the spatial and temporal scales and resolution of those approaches is necessary to fully understand the effect of environmental perturbations on earth systems and how to mitigate the impact of these perturbations.

A formidable disparity exists between the temporal and spatial scales of the fundamental processes that govern terrestrial carbon and water cycling and those represented in current models. Large-scale models include a terrestrial (land surface) component that represents key fluxes of energy, water, and carbon, as well as greenhouse gases (predominately CO_2 and CH_4) across the terrestrial-atmosphere interface. However, predicting ecosystem stability requires foundational knowledge of molecular processes that describe electron transfer between molecules, nutrient dynamics, and elemental (including contaminant) fluxes. It also requires knowing where these processes occur within the terrestrial environment and their impacts on systems of increasing scales through the complete ecosystem (BER 2014c). This is because near- and below-surface biogeochemistry is driven by reactions occurring at the molecular scale of enzymes, cell membranes, organic substrates, and mineral surfaces. On the other hand, key processes may involve emergent phenomena that are difficult to identify and study by reductionist approaches. The mismatch in scale and lack of a whole-systems perspective contributes to the inability of current models to predict the response of ecosystems to perturbations or disturbances.

Previous research at EMSL has focused on linking processes at multiple scales to improve molecular-level understanding of terrestrial and subsurface ecosystems. These have included the development of micromodels for investigating subsurface geochemical reactions (Zhang et al. 2010a), studies on the bioreduction of minerals (Bose et al. 2009) and the effects of pore-scale heterogeneity on bacterial growth in model soil systems (Zhang et al. 2010b). In parallel with efforts to define the molecular events occurring in the subsurface, EMSL scientists have focused on developing more highly resolved subsurface reactive transport models that incorporate spatially resolved processes (Scheibe et al. 2015a)

(Figure 4.2). These models suggest that the fate and transport of abiotic and biotic compounds are dependent on the interplay of a complex array of variables such as the oxidation state of the materials, the redox conditions of the subsurface sediments, the native microbial community, and the hydrologic characteristics of the sediments (e.g., Tartakovsky et al. 2013; Lee et al. 2014).

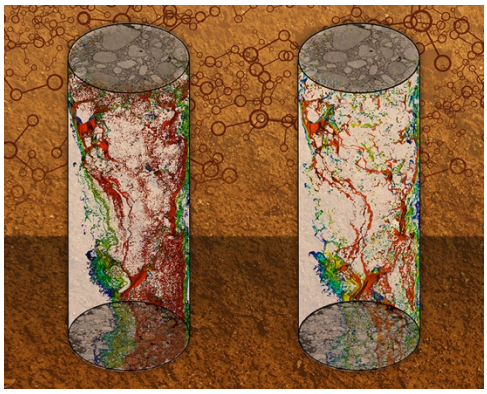


Figure 4.2. By bridging the gap between small and large spatial scales, the simulation method could pave the way for more accurate assessments of the risk of groundwater contamination and the development of more effective remediation strategies (Scheibe et al. 2015b). The figure shows a comparison of solute transport simulations using binary segmentation (left image) and ternary segmentation (right image). For the whole story, read *Multiscale Simulations*.

Although EMSL users have made significant advances in understanding nutrient and contaminant transformation and transport in the near- and subsurface ecosystems, there is a significant need to include additional macroscale interactions between ecosystem components (e.g. between plants and microbes or plants and minerals) and between ecosystem subdomains (e.g. along groundwater-surface water interfaces). There is also a critical need for an improved understanding of "keystone" processes, their behavior across ecosystem boundaries, and their connections to larger-scale phenomena. This will require the development of novel strategies to combine field and laboratory studies to test, predict, manipulate, and control subsurface biogeochemical processes in key subsurface environments. By combining laboratory, field, and modeling studies, EMSL users can improve molecular- to pore-scale mechanistic understanding of the contribution of plants, microbes, biogeochemical controls, and hydrologic inputs to elemental cycling. This will lead to a better understanding of the feedbacks between the water cycle and ecosystem biogeochemistry, which can be used to improve the biogeochemistry components of earth system models.

4.1 Summary of workshop discussions

Participants had an active discussion on the inability to link genomic information to biological phenotypes and then to link this to higher-order ecosystem properties (BER 2015b). This science gap greatly impedes the ability to predict ecosystem processes and responses from data collected in the field. Although many groups are working toward linking the genome to phenome, additional efforts are needed to advance this linkage to the ecosystem level and to include abiotic factors. Phenomes influence local environments and thus influence larger environmental processes. Conversely, local ecosystem properties also influence phenomes (Houle et al. 2010). Thus, there is a need to develop a scalable mechanistic understanding of feedback processes to fully characterize ecosystems.

Workshop participants felt there were multiple scientific areas in which EMSL could make significant contributions to TSE research areas: 1) a better understanding of metabolite flow through subsurface ecosystems, particularly between plants and microbes, 2) improved understanding of elemental flow through the environment by better combining field studies and laboratory experiments, and 3) developing dynamic models of terrestrial and subsurface ecosystems that include both spatial heterogeneity and feedback processes.

4.1.1 Metabolite flow through subsurface ecosystems

The ability to predict ecosystem biogeochemistry requires knowledge of molecular processes and how they impact electron, nutrient, and contaminant fluxes. Plant-microbe metabolic interactions are crucial in dictating biogeochemical cycling of carbon, nutrients, and contaminants through ecosystems and their environmental responses (BER 2014a). Both the range of metabolic interactions and their response to

environmental perturbations are encoded in the genomes of the interacting species, but the community lacks the ability to predict phenotypes from genotypes. Progress in this area would be accelerated by improved *in-situ* metabolomics profiling measurements to assess the state of the ecosystem. Key metabolites could be used as biomarkers of ecosystem health, especially in response to a rapidly changing environment, and could be used to effectively identify and potentially mitigate harmful impacts.

The SIP is a particularly attractive approach for following the dynamics of specific elements through systems and can generate quantitative measurements needed to calculate the rates of important biological processes in the environment. These measurements not only can identify active microbial community members, but also be used to measure the flow and flux of metabolites. Because of the expertise at EMSL, stable isotope measurements are areas in which EMSL could excel and provide important information on the flow of energy through complex ecosystems. In particular, SIP can be used in conjunction with proteomics, metabolomics, and imaging technologies to identify which organisms are actively metabolizing specific compounds and the spatial localization of the conversion.

To connect the genome of environmental organisms to their phenome, it will be necessary to map metabolic processes to specific organisms. The complexity of microbial communities in the soil makes this a daunting challenge. Initial work should focus on a select set of model microbial communities and important interfaces, such as between plants and microbes. In exploring potential model systems, their natural genetic variations can be used to determine which range of genotypes can result in informative system properties that can be explored further. It will be important to link genome composition of microorganisms and plants with phenotypic traits at cellular, organismal, and community levels.

Interpreting metabolic flux data will require improved computational models of underlying plant and microbial metabolic processes, particularly in response to rapidly changing environments. Most current models are focused on metabolic potential rather than dynamics and cannot predict how changes in gene expression are translated into altered metabolism. Knowledge of thermodynamics and biological regulation should be included in improved computational models as well as the impact of spatial arrangements and biological compartmentation. Here, imaging technologies can provide a foundation for building these improved models. For example, ^{11}C positron imaging can be used to follow real-time carbon allocation to plant tissues (and exudates). The nature of interaction surfaces and high-resolution metabolite transfer can also be studied with nano-SIMS. Eventually, studies of metabolite flow should be extended to include multiple microbial species (including fungi), plants, and the impact of different minerals and soil types on metabolite transfer.

4.1.2 Combining laboratory and field studies

Integrating across scales, there also is the classical challenge of extrapolating from laboratory results to field studies and the reverse while simultaneously accounting for the complexity and heterogeneity of the real world. The activity of subsurface organisms is highly dependent on soil, nutrient, and moisture conditions, and these factors in turn can determine the number, variety, and interactions of the organisms. Knowing all the important factors of the environment under study can be difficult to predict. Thus, even though exploring model systems in the laboratory under tightly controlled conditions is required for defining important molecular mechanisms, it will be important to use information from field studies to properly interpret them. Addressing science questions across the span of molecules to organisms to ecosystems is best accomplished by integrating field observations and laboratory investigations. This will inform the choice of model systems and ensure they contain important aspects of the complexity inherent in the native environment. As an intermediate step in defining important environmental parameters to explore, an “eco-simulator” can be constructed in which intact native plant-soil-microbe environments are brought into large-scale and highly controlled growth chambers.

How to design combined laboratory and field studies so they are synergistic is not well established, and EMSL could make significant contributions in this area. It will be important to use field-derived hypotheses to design experiments to elucidate mechanisms using a range of spatial and temporal scales. This is because the heterogeneity that exists at every level in the environment (e.g. the composition and functions of microbial communities, the geochemistry and physical properties of earth materials, genome expression dynamics, moisture and pore size distribution, etc.) can have a profound effect on ecological functions. Thus, using natural phenomena to direct critical questions from both top-down, as well as bottom-up, is likely to be most effective in identifying emergent biogeochemical properties that regulate the system. However, applying this type of multiscale analysis to field studies would require developing more powerful field-deployable instrumentation, particularly those that can evaluate multiple parameters simultaneously (e.g. multimodal). Unless the key parameters can be simultaneously captured, reproducing field observations in the laboratory will be difficult. Using instrumented drones is an attractive approach for surveying field sites to assess their spatial heterogeneity and to ensure representative sampling.

Model systems for combined field and laboratory studies should represent a range of biotic and abiotic processes that modulate the fate and transport of carbon, nutrients, and contaminants. Plant rhizosphere systems are particularly attractive models because they encapsulate a range of environmental organisms together with their functional interactions. Such a system could be used to develop isotopic fractionation methods/models that could better constrain natural abundance isotope studies for biogeochemistry and SIP. Also, it would be useful to develop an enclosed research greenhouse for studying interactions between plants and the environment (phytotron) and is coupled to other EMSL analytical capabilities. Phytotrons could also support the development of imaging-based noninvasive root characterization and sampling systems for laboratory and field studies.

4.1.3 Scalable models

To enable robust predictions, quantitative modeling approaches informed by microbial metabolism and hydro-biogeochemical processes are needed. It is also important to develop models that can work across a range of spatial and temporal scales. Parameterization of metabolic and abiotic processes in existing models has mostly been based on empirical observation or assuming linear reaction kinetics. The new generation of microbial-informed, reactive transport models are needed that are based on thermodynamic calculations of the reaction kinetics. These models need to include key processes that cause observed ecosystem variations as well as the feedback and heterogeneity that characterize natural ecosystems. Ideally, these would be fully coupled models of subsurface environmental processes that incorporate metabolic modeling of microbial processes, and molecular-scale understanding of geochemical stability, speciation, and biogeochemical reaction kinetics. Such models could be used to identify diagnostic signatures of the system response at varying spatial and temporal scales.

A challenge in building scalable models of natural ecosystems is how to include biological complexity at a useful level of abstraction. The DOE has supported efforts to build computational models (regulatory, flux, interactions) that describe intercellular and intracellular microbial interactions as well as how they are regulated. Because of the complexity and heterogeneity of natural environments, however, it is not currently feasible to include such details in an ecosystem-scale model. Thus, new approaches need to be developed to incorporate microbial metabolism and thermodynamics into scalable hydro-biogeochemical models. Identification of critical regulatory processes and feedback mechanisms across a range of environmental conditions is likely to be key to developing broadly useful models. Such models should be developed through collaboration between ecosystem modelers, domain scientists, and developers who can develop/improve simulation codes. These models could use multiscale simulation frameworks, such as those being developed in the BER and ASCR-supported Interoperable Design of Extreme-scale Application Software project.

There is a broad need to provide a computational framework that can integrate a wide variety of data types to create predictive models that are critically needed in this area of research. The CESD within BER has promoted the idea of a community cyberinfrastructure that can advance the predictive understanding of earth and environmental systems. They have proposed the creation of a "Virtual Laboratory" that would integrate disciplines and scientists, field sites and field data, and experimental and laboratory data at different temporal and spatial scales (BERAC 2013). This cyberinfrastructure would enhance the development of conceptual and numerical models so scientists could effectively develop a predictive understanding of natural, managed, and disturbed terrestrial, aquatic, and subsurface systems across multiple scales and processes. EMSL should consider aligning with this effort so data it generates are available to the community to test and validate the models being developed.

4.1.4 Scientific workforce development

Besides novel instrumentation, methods, and computational models, there is also a critical need for training a new generation of scientist to use these sophisticated technologies to meet DOE research needs. The multidisciplinary nature of research that crosses multiple spatial and temporal scales requires a better approach for identifying appropriate collaborators and then bringing them together for focused research studies. Specifically, there is a need to:

- Integrate experimental/observation/modeling training courses that discuss strengths and weaknesses of specific analytical technologies, and introduce appropriate modeling approaches
- Develop a network of scientists interested in the TSE research area that could expand the EMSL user base and help develop the scientific workforce
- Identify and engage scientists who are experienced in integrating knowledge across spatial and temporal scales and those that bridge fields; communication across disciplines is essential
- Implement methods for dynamically assembling teams and identifying mechanisms to enhance the support of EMSL teams working in this research area
- Form longer-term EMSL collaborations with users and research teams

4.2 Summary of recommendations

4.2.1 TSE breakthrough science areas that were identified

- Linking the genomes of organisms to their phenotypes and the impact of these phenotypes on ecosystem properties
- Understanding how local, small-scale environmental properties influence larger environmental processes and the ensuing feedback to the small scale
- Understanding how to incorporate microbial metabolism and thermodynamics into scalable hydro-biogeochemical models
- Identifying key processes and feedbacks that result in observed ecosystem variations
- Understanding the regulatory interactions, metabolite flux, and dynamics both between and across microbes and plants
- Understanding the role of natural genetic variation and environmental heterogeneity in the systems-level properties of microbial communities and ecosystems

4.2.2 Near-term investments

- Improved coupling of current EMSL capabilities (MS, nano-SIMS, LA-ICPMS, imaging techniques) are needed to provide focus and emphasis in establishing and developing recognized SIP capability and expertise
- A new high-intensity ion source for nanoSIMS to increase sensitivity and lateral resolution for trace metals in biological and environmental systems while dramatically increasing throughput and decreasing required maintenance
- X-ray nanotomography capability for high-resolution 3D imaging to help BER scientists connect research between the molecular and pore scales and enable better integration of EMSL capabilities with light-source and synchrotron facilities
- New 600 and 800 MHZ NMR systems and upgrades for following elemental pathways (C, N, H, O, S, P, etc.) through the plant-microbe-soil-atmosphere system and new mass spectrometry systems to better support SIP
- Improved methodologies for isotope-based metabolic flux analysis to quantify pore-scale soil metabolic processes such as substrate use efficiency, energy production and consumption for growth and maintenance, substrate consumption, nitrogen demand, oxygen consumption, and microbial carbon isotope composition
- Improved resources for collaborations focused on linking domain scientists with modelers and software developers

4.2.3 Medium-term investments

- Development of metabolite libraries and experimental capabilities to enable metabolomic profiling of ecosystem status or health (with high spatial and time resolution)
- Development of methods for isotope-based metabolic flux analysis to the level of individual microbial species within microbial communities
- *In-situ* imaging capabilities to characterize the molecular-level interaction of organisms with both biotic and abiotic interfaces

4.2.4 Long-term investments

- Mobile technologies for phenotype analysis in the field (e.g. multispectral imaging using drones, field-deployable GC, LC, or X-band radar systems)
- A large-scale soil-plant-atmosphere growth chamber for studying natural complexity under controlled environmental conditions and verifying predictions of ecosystem response to environmental changes

5.0 Science Theme: Molecular Transformations

The MT Science Theme was articulated in 2016 and has replaced EMSL's Energy Materials and Processes Science Theme (Figure 5.1). Because molecular transformations play fundamental roles in each of the EMSL science target areas in biology, environment, and energy, the MT Science Theme is expected to play a cross-cutting and integrating role for EMSL science.

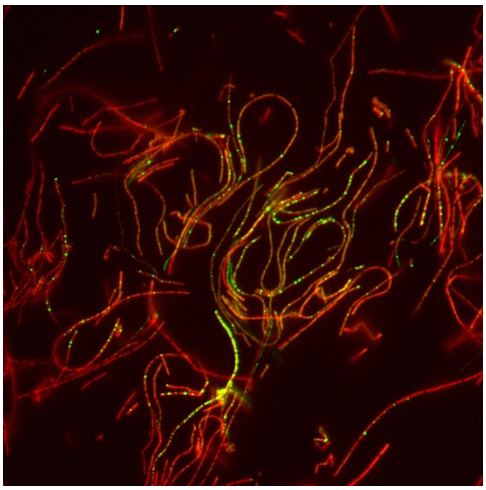


Figure 5.1. The MT Science Theme focuses on predictive understanding of molecular transformations in biology and chemistry central to energy production, bioconversion, and other processes that advance the other EMSL science themes.

Many research areas supported by BER require an understanding of molecular transformations. BER has emphasized the need for new methods for characterizing chemical reactions at natural, mineral surfaces; the organization and structural components in molecular complexes; and tracking molecules to view cellular processes (BER 2014c) as they are occurring. Fundamental knowledge of the molecular-scale processes from laboratory studies is essential for understanding the dynamics of chemical and biological processes in the different environmental media. This information will also inform larger-scale studies aimed at understanding earth system responses to changes in environmental emissions. There is also a need to understand solvent-mediated transformations involving charge, mass, and electron (energy) flow at interfaces. This knowledge will impact EMSL studies in biology, the terrestrial subsurface and atmosphere, as well as those central to the production of biofuels, bioproducts (BER 2014c; BES 2015; BES/BER/ASCR 2016). Ultimately, the goal is to understand the dynamic and emergent behaviors of interfaces in natural and engineered systems sufficient to predict and, where possible, control these important processes.

Achieving this predictive understanding requires integration of many types of experimental information with modeling and simulation. This will require dynamic studies at molecular or macromolecular scales, computational chemistry, and multiscale modeling methods that can identify key features and processes with the goal to enable the use of molecular-scale information to understand emergent mesoscale (molecule to systems) phenomena.

Because of the importance of dynamic processes that occur at solvated interfaces in the environment, biology, and energy systems, EMSL has established a long-term goal of significantly advancing knowledge in this area of research (BER 2014c; BES 2015; BES/BER/ASCR 2016). Current efforts are making significant progress in both the science and the new capabilities (experimental and computational) needed to significantly advance understanding (Wang 2015; Zhou et al. 2016). This should enable the prediction of transformation mechanisms and resulting properties that occur in both natural and engineered systems, as well as prediction of environmental responses to the energy-generation associated byproducts. In particular, biocatalysis (the use of biomolecules to speed up chemical reactions) is likely to be a significant research area that crosscuts to BDD and uses synthetic biology, structural biology, and surface chemistry. Using proteins as scaffolds for work on complex biochemical and metabolic conversions *in vitro* will help to advance bio/conversion of bioproduced intermediates or waste material to carbon-based fuels and chemicals.

5.1 Summary of workshop discussions

This workshop followed immediately and built upon a one-day Science Theme Advisory Panel that discussed the nature and useful directions MT should pursue. Because MT is a recently revised science theme, much of the discussion included assessments of the role and importance MT could or should play in EMSL science. The overall conclusion was that molecular transformations play a significant role in all science areas supported by DOE in general and BER specifically. This science theme targets important interactions that control biological processes that impact energy, water, and the environment. Linkages of MT with the other EMSL science themes create an opportunity for a more comprehensive scientific approach that could lead to major scientific advances that would not be possible at most institutions.

It was also affirmed that molecular transformations at solid/liquid interfaces impact many natural (Brown and Calas 2012) and engineered systems (Siretanu et al. 2014), and advancing the understanding of solvent-mediated molecular transformations at solid/liquid interfaces is an appropriate and challenging goal for EMSL. The challenges associated with the solid/liquid interface are also common with those associated with interfaces between hard and soft matter (including interactions with biological molecules on surfaces) (Walsh et al. 2011) and interfaces in soft materials (Pincus and Tirrell 2014). There was consensus that the greatest impact would be achieved by focusing on a subset of potentially controlling molecular transformations. Significant progress can be made by focusing on the areas identified in the workshop including: 1) the dynamic nature of the interfaces; 2) improved ability to experimentally and computationally deal with heterogeneity, highly active low-population sites, and increased complexity; 3) important processes occurring outside the immediate molecular scale; and 4) relating molecular processes to large-scale function or behavior.

5.2 Summary of recommendations

5.2.1 MT breakthrough science areas that were identified

5.2.1.1 Scientific challenges

To develop a predictive understanding of solvent-mediated molecular transformations at interfaces, it is necessary to recognize the challenges and opportunities inherent in specific interfaces and processes (Figure 5.2). Several challenges that were identified that are important opportunities include:

- Dynamic systems – Interfacial systems are dynamic. Progress requires recognizing dynamic relationships among biomolecule and solid (natural mineral) interfaces and solvents (such as water) and how these mutual interactions influence molecular transformations or other functional properties.
- Mesoscale effects – Molecular transformations will be impacted by processes occurring beyond the immediate molecules of interest or active site of these molecules. These may include a variety of environmental impacts (temperature, pH, water moisture content, redox potential), charge, and mass and energy availability and transport, as

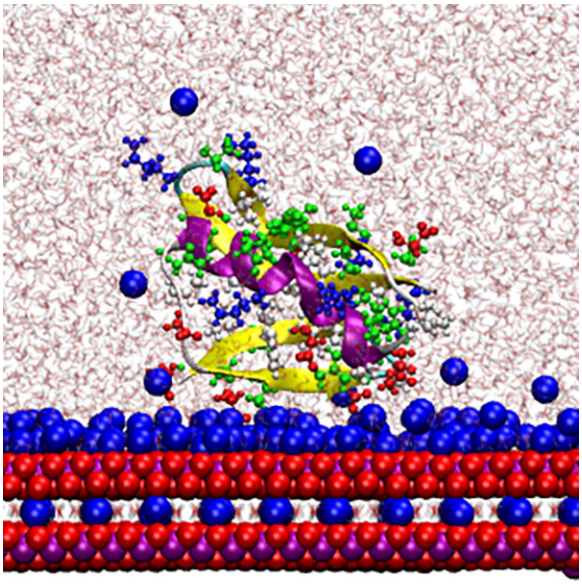


Figure 5.2. Image from a molecular dynamics study of protein interacting with a mineral surface in a water environment. The mineral birnessite, in contrast to other mineral surfaces, was found both experimentally and by modeling to provide an abiotic pathway for protein fragmentation by soil minerals (Andersen et al. 2016).

well as long-range, non-covalent interactions around the active sites.

- Complexity – Many systems have different levels of complexity, including compositional and environmental complexity, heterogeneity of reaction mechanisms, and a distribution of important sites (e.g. defects, boundaries, etc.). Weak long-range interactions can also play an important role in some processes and transformations.
- Linking data across multiple scales – It is important to relate measurements and models at multiple spatial and time scales; e.g. understanding the molecular causes of large-scale behavior. In the case of biological systems, it will be important to understand how to convert genomic information into functional (bio)molecules and materials.

There were also several questions that should be addressed in this area of research:

- What is the nature of the liquid phase near a surface/interface and its impact on molecular transformations?
- What are the molecular mechanisms of heterogeneous reactions in large complex enzymatic systems (e.g. lignin breakdown)?
- What is the influence of (mineral or catalyst) surfaces on the structure, dynamics, and reactivity of organic matter? How are the physics and chemistry of these environments coupled?
- What is the influence of photochemistry on contaminant and carbon cycling, and alternative energy?
- Why do some long-range weak interactions have major influences on molecular transformations at cell-mineral interface?

5.2.1.2 Technical challenges

The discussion considered technical challenges that need to be addressed to make progress on the scientific challenges. Complexity of the systems that needs to be understood and/or controlled is an overarching challenge that is particularly relevant for solvent-mediated interfacial processes, but is also a broadly relevant issue. In general, the technical challenges are instrumental and computational.

Instrumental challenges

- Apply EMSL's state-of-the-art tools for interrogating interfaces in controlled environments to more heterogeneous and complex systems. *In-situ* monitoring capabilities of active molecular transformational systems at multiple scales are of particular importance.
- Combine techniques to obtain orthogonal information on interfaces that can directly link species with function.
- Use *in-situ* real-time measurements during synthesis, processing, and operation (aging, cycling, and exposure) for the dynamic nature of many systems.
- Access to the unique *in-situ* and support capabilities in EMSL and comparable *in-situ* measurements using highly penetrating probes such as X-rays and neutrons at other user facilities could help address measurement challenges. Solvent-mediated interfacial transformations often take place in confined structures and at buried interfaces.
- Develop additional approaches for gathering detailed information about soft matter in heavy hard-material environments.

- Access instruments and methods capable of quantifying long-range weak interactions, such as electrostatic, hydrogen bonding, and hydrophobic interactions.
- Increase the number, types, sensitivities, and data collection rates of field-deployable instrumentation that can collect molecular-level data.

Computational challenges

- Bridge the scales of fine-grained models of isolated phenomena and coarse-grained models for larger-scale systems.
- Develop scalable frameworks that combine state-of-the-art theoretical methods with novel programming models that can exploit peta/exascale computing platforms.
- Integrate and analyze the results from multiple types of complementary measurements and experiments, including those involving large data sets.
- Mine data from dissimilar data sets. This will require physical models for integration of the data that include and deal with uncertainty and false results.
- Increase the length and time scales of systems for which rigorous computation can be done.
- Enhance the accuracy of computation models.
- Improve accessibility and use of advanced computation by researchers lacking theoretical or computational expertise.

5.2.2 MT breakthrough capabilities required

In the course of this workshop, it became clear that no single, new capability would meet the identified challenges. Therefore, instrumental development must occur on several fronts:

- Facilitate the development of multi-instrument, multimodal capabilities and information processing, including *in-situ* abilities, increasing spatial and time resolution, structure, chemical selectivity, and sensitivity. While EMSL and PNNL have significant experience, more development is needed.
- Improve spatial resolution, sensitivity, and the range of environmental conditions to make measurements closer to realistic conditions and/or multimodal operation. New instrumentation and methods are always stretching the limits of sensitivity, resolution, and measurement simultaneity. EMSL needs to refresh many of its capabilities in order to stay cutting-edge and in demand for users.
- Several upgrades and additions to existing microscopy capabilities to keep existing systems at state-of-art and to make them more amenable to soft-tissue, biological sample experiments
- Dynamic real-time measurements to understand the nature of interfaces and their surroundings.
- Instrumentation that can effectively deal with the liquid side of the interface
- Multiple scale measurements that relate molecular behaviors to a larger scale (natural systems) or function (engineered systems), preferably using a single multimodal instrument
- Analytical techniques that can be applied to single-particle, ultra-fine aerosol, soil, and engineered particles as a function of history, transport, and environment (i.e. source to deposition). These techniques again require the ultimate in sensitivity and time-resolved analysis.
- Increased sensitivity of imaging and other capabilities, especially those with sensitivity to see minor species in a complex environment, including the solid/liquid interface

- Expand field-based molecular capabilities for lab-based studies, especially analytical capabilities that can be remotely monitored and controlled
- Apply expertise to take advantage of important and unique capabilities at other user facilities. Certain new EMSL capabilities (i.e. X-ray spectroscopy, microscopy, tomography (Figure 5.3) would make EMSL scientists better, more efficient, and more competitive users at X-Ray Light Sources.

More powerful computational approaches could also have a significant impact on research in molecular transformations. There is a need for:

- Theoretical approaches (quantum, classical, mixed quantum/classical), force field development (reactive and non-reactive), novel theoretical models and algorithms, and sufficient computational resources for simulation of complex mechanisms/reactions across multiple length and time scales
- Increased fidelity of computational models by employing new computer technologies and resources
- Modeling to assist experimental design, especially for complex systems that include metalloenzymes, which can play important roles in molecular transformation, such as biomass conversion

Figure 5.3. The study reveals a newly discovered two-step mechanism by which brown rot fungi degrade plant cell walls. The findings could be harnessed to develop highly efficient and cost-effective strategies for renewable energy production. For the whole story, read Brown Rot Fungi Efficiently Degrades Wood.

- Improved data analytics and data integration of results for many types of complementary measurements and experiments, such as integrating spectroscopic data with simulations
- Development of physical models that could serve as a framework for integration of the data, including the ability to include uncertainty and “false” results
- Near real-time data analysis to optimize experimental efficiency

6.0 Workshop Cross-cutting Themes and General Recommendations

The breakout discussions were framed by the need to be relevant to current DOE mission areas and long-term goals in the future. During the two-day meeting, several themes arose that cross-cut all science areas. One idea that was brought up repeatedly was the need for EMSL to identify specific areas of science in which to build an international reputation. This area (or areas) would better establish EMSL's recognition within the scientific community, increase its user base, and ultimately raise the scientific impact of DOE's investments at EMSL. Several cross-cutting science needs (gaps) that tie science themes together were also discussed and could serve as a nucleation point for development of a specific scientific challenge on which EMSL should focus. For instance, one of the greatest challenges faced by BER scientists is decoding how genomic information is translated to cell, organismal, community, and ecosystem phenotypes. This theme could serve as a unifying "experiment" that brings together multiple science themes, scientific disciplines, and EMSL's capabilities.

Several critical gaps in the scientific areas were identified and could serve as initial cross-cutting focus areas. These included:

- Robust approaches to deal with complexity (especially in regulatory processes), both conceptually and experimentally
- Methods for reliably linking and integrating data across different spatial and temporal scales
- General approaches for combining qualitative and quantitative data or integrating dissimilar data types
- Capturing spatiotemporal dynamics at sufficient resolution to support computational modeling
- Identifying key/critical steps and feedback in complex processes to simplify modeling of the system
- Conceptual frameworks for understanding the impact of the environment (abiotic factors) on living systems
- Approaches for accurately predicting cell/organismal/community/ecosystem phenotypes from their genotype
- Approaches for designing laboratory studies to better connect with field studies
- More efficient approaches for implementing the understand-predict-control concept for complex systems

Participants suggested several generic capability improvements as well as some specific capabilities in which EMSL should invest to move its agenda forward. These included:

- Increased sensitivity of current measurement capabilities
- Enhanced precision and accuracy of current instrumentation
- Novel molecular-level instrumentation for field measurements
- *In-situ* and (near) real-time analytical measurements
- High-throughput phenotype assays, both screening and molecular level, could be realized as specialized facilities
- SIP and spatially resolved omics measurements

- Integrated experiments (i.e. performing multiple analyses using the same sample)
- New facilities, including and eco-simulator (i.e. coupled atmospheric and ecosystem chamber) systems or facilities
- Developing model systems of tractable complexity to facilitate dynamic measurements, and scaling from molecules to system-level processes
- Improved user engagement so users know better what EMSL has to offer and how the capabilities at EMSL can help solve their problems

There were also multiple discussions regarding the types of specific scientific expertise that could be added to EMSL to facilitate its mission. These included:

- Experiment-theory integration
- Data management, analysis, visualization, and integration
- Gas-phase chemistry
- Certain bio-disciplines (e.g. microbial and plant physiology; plant systems modeling; metabolomics; data management, integration and interpretation)

In every session, there were discussions about computational capabilities in which EMSL needs to focus and expand to meet users' needs. Mostly, discussions focused on how computing and experiments could be used synergistically to improve both the experimental process and predictive capabilities. The following areas of computational focus were suggested:

- Computational biology – Improved modeling at the molecular, cell, organism, and community scales. Biotic and abiotic system expertise are needed.
- Computational chemistry – There are multiple opportunities in which computational chemistry simulations could provide atomistic-level descriptions of chemical transformations and processes, but greater communication is needed between computational chemists and domain scientists working in BER-relevant science.
- Data and metadata management – Integrate data management and metadata capture into scientific workflows. To accomplish this, software engineers are needed who can work with scientists to develop novel software tools and workflows.
- Experimental data interpretation – This includes bioinformatics and the ability to run simulations to provide insight to users, not just data (e.g. using pre-calculated collisional cross sections to identify metabolites measured by ion mobility mass spectrometry). Integrated omics and image analysis were also specifically called out along with computational molecular spectroscopies to bridge the gap with experiments.
- Multiscale modeling – Enable concurrent simulations or other forms of information passing across scales (both up and down). Improve accuracy of larger-scale simulations by linking to high-resolution simulations (e.g. tying optical properties of aerosols to atmospheric energy adsorption or reflection).

It is critically important to maintain the cutting-edge nature of EMSL's current custom and commercial instrumentation. Technical progress has been rapid in specific fields, such as mass spectrometry and RNA/DNA sequencing, and some EMSL capabilities run the risk of becoming obsolete if they are not refreshed. The average age of EMSL's instrumental capabilities is nearing seven years, meaning these capabilities are at least two to three generations beyond state of the art. It was suggested that EMSL identify at-risk capabilities that are of high value to the user community and prioritize for upgrades.

Making such upgrades will ensure that EMSL continues to support users to perform the highest-impact science and makes essential contributions to BER science mission areas.

The panel also discussed sociological, staffing, training, education, and integration needs and opportunities:

- Develop dedicated or focused portfolios of capabilities to address relevant scientific challenges.
- Identify and partner with strategic researchers who can bridge scales to link molecular- and large-scale systems and to foster the development of integrative scientists.
- Develop user or staff expertise linking BER domain questions to computational tools. Develop software (an app) to help match a technique to a specific problem. Train users with little computational expertise to effectively use PNNL/EMSL computational resources.
- Implement near real-time analysis (data, processing, and modeling) to increase productivity.
- Provide mechanisms for pre-experiment modeling.
- Enable EMSL staff development with deeper domain expertise and increase visibility through additional outreach.
- Implement a graduate program at PNNL/EMSL or, at the least, workshops for staff.

7.0 Conclusion

The EMSL Breakthrough Science & Technology Workshop that took place on January 10–11, 2017, provided the science community's visionary look at new areas of opportunity where advances in molecular sciences and associated state-of-the-art capabilities can have a significant impact. Several cross-cutting themes were identified, creating the foundation for EMSL investments over the next decade. Dealing with the complexity and spatiotemporal dynamics across the scales were identified as critical gaps in the scientific areas necessitating advanced tools for lab and field measurements with improved sensitivity, accuracy, resolution, specificity, and throughput. Parallel with this, computational tools for data analysis and simulation need to be developed to handle and interpret expanded amounts of data that will be generated. Current and developing capabilities need to be combined in creative new ways to build the next generation of research tools. The intersection of the four science themes presents major opportunities for the greatest scientific impact on current and future DOE-BER mission areas.

8.0 References

Andersen A, PN Reardon, SS Chacon, NP Qafoku, NM Washton and M Kleber. 2016. "Protein-Mineral Interactions: Molecular Dynamics Simulations Capture Importance of Variations in Mineral Surface Composition and Structure." *Langmuir* 32(24):6194-6209. DOI: 10.1021/acs.langmuir.6b01198.

Angelo M, SC Bendall, R Finck, MB Hale, C Hitzman, AD Borowsky, RM Levenson, JB Lowe, SD Liu, SC Zhao, Y Natkunam and GP Nolan. 2014. "Multiplexed Ion Beam Imaging of Human Breast Tumors." *Nature Medicine* 20(4):436-+. DOI: 10.1038/nm.3488.

Baldazzi V, N Bertin, H de Jong and M Genard. 2012. "Towards Multiscale Plant Models: Integrating Cellular Networks." *Trends in Plant Science* 17(12):728-736.

BER – Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy. 2014a. *Data-Model Needs for Belowground Ecology*. Washington, DC. Available at https://science.energy.gov/~/media/ber/pdf/workshop%20reports/Belowground-Workshop-Report-Nov2014.pdf.

BER – Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy. 2014b. *Doe Genomic Science Program 2014 Strategic Plan*. Washington, DC. Available at http://genomicscience.energy.gov/strategicplan/GenSciStrategic%20Plan_Oct2014LR.pdf.

BER – Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy. 2014c. *Molecular Science Challenges*. DOE/SC-0172. Washington, DC. Available at http://genomicscience.energy.gov/molecularscience/molecularsciencechallengesreport.pdf.

BER – Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy. 2015a. *Biological Systems Science Division Strategic Plan*. Washington, DC. Available at http://genomicscience.energy.gov/pubs/BSSDStrategicPlanOct2015.pdf.

BER – Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy. 2015b. *Building Virtual Ecosystems: Computational Challenges for Mechanistic Modeling of Terrestrial Environments*. Accessed March 3, 2017 at http://www.doesbr.org/BuildingVirtualEcosystems/

BERAC – Biological and Environmental Research Advisory Committee. 2013. *Ber Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges* DOE/SC-0156. Available at https://science.energy.gov/~/media/ber/berac/pdf/Reports/BER_VirtualLaboratory_finalwebHR.pdf.

Bernstein HC, RS McClure, V Thiel, NC Sadler, YM Kim, WB Chrisler, EA Hill, DA Bryant, MF Romine, JK Jansson, JK Fredrickson and AS Beliaev. 2017. "Indirect Interspecies Regulation: Transcriptional and Physiological Responses of a Cyanobacterium to Heterotrophic Partnership." *mSystems* 2(2). DOI: 10.1128/mSystems.00181-16.

BES – Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy. 2015. *Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science*. Washington, DC. Available at https://science.energy.gov/~/media/bes/besac/pdf/Reports/Challenges_at_the_Frontiers_of_Matter_and_Energy_rpt.pdf.

BES/BER/ASCR – Offices of Basic Energy Science, Biological and Environmental Research, and Advance Scientific and Computing Research; Office of Science; U.S. Department of Energy. 2016. *Basic Research Needs for Environmental Management: Report of the Office of Science Workshop on Environmental Management July 8–11, 2015*. Washington, DC. Available at https://science.energy.gov/~/media/bes/pdf/reports/2016/BRNEM_rpt.pdf.

Boenisch MJ, KL Broz, SO Purvine, WB Chrisler, CD Nicora, LR Connolly, M Freitag, SE Baker and HC Kistler. 2017. "Structural Reorganization of the Fungal Endoplasmic Reticulum Upon Induction of Mycotoxin Biosynthesis." *Scientific Reports* 7. DOI: 10.1038/Srep44296.

Bose S, MF Hochella, YA Gorby, DW Kennedy, DE McCready, AS Madden and BH Lower. 2009. "Bioreduction of Hematite Nanoparticles by the Dissimilatory Iron Reducing Bacterium Shewanella Oneidensis Mr-1." *Geochimica Et Cosmochimica Acta* 73(4):962-976. DOI: 10.1016/j.gca.2008.11.031.

Brown GE and G Calas. 2012. "Mineral-Aqueous Solution Interfaces and Their Impact on the Environment." *Geochemical Perspectives* 1(4-5):483-742. DOI: 10.7185/geochempersp.1.4.

CESD – Climate and Environmental Sciences Division, Office of Biological and Environmental Research, U.S. Department of Energy. 2012. *Biological and Environmental Research, Climate and Environmental Sciences Division Strategic Plan*. DOE/SC-0151. Washington, DC. Available at https://science.energy.gov/~/media/ber/pdf/CESD-StratPlan-2012.pdf.

Chozinski TJ, AR Halpern, H Okawa, HJ Kim, GJ Tremel, ROL Wong and JC Vaughan. 2016. "Expansion Microscopy with Conventional Antibodies and Fluorescent Proteins." *Nature Methods* 13(6):485-+. DOI: 10.1038/Nmeth.3833.

DOE-SC – U.S. Department of Energy Office of Science. 2015. *Regional and Global Climate Modeling Program Overview and Status Report*. DOE/SC-0184. Washington, DC. Available at https://science.energy.gov/~/media/ber/pdf/community-resources/RGCM-ProgramStatusandOverview-FY2016web.pdf.

Ehn M, JA Thornton, E Kleist, M Sipila, H Junninen, I Pullinen, M Springer, F Rubach, R Tillmann, B Lee, F Lopez-Hilfiker, S Andres, IH Acir, M Rissanen, T Jokinen, S Schobesberger, J Kangasluoma, J Kontkanen, T Nieminen, T Kurten, LB Nielsen, S Jorgensen, HG Kjaergaard, M Canagaratna, M Dal Maso, T Berndt, T Petaja, A Wahner, VM Kerminen, M Kulmala, DR Worsnop, J Wildt and TF Mentel. 2014. "A Large Source of Low-Volatility Secondary Organic Aerosol." *Nature* 506(7489):476-+. DOI: 10.1038/nature13032.

EMSL – The Environmental Molecular Sciences Laboratory. 2008. *Recapitalizing Emsl: Meeting Future Science and Technology Challenges*. PNNL-17653. Richland, Washington. Available at https://www.emsl.pnl.gov/emslweb/sites/default/files/reports/goldbook_08.pdf.

EMSL – The Environmental Molecular Sciences Laboratory. 2015. *Atmospheric Aerosol Science Theme Advisory Panel Report*. PNNL-24039. Richland, WA. Available at https://www.emsl.pnl.gov/emslweb/atomspheric-aerosol-science-theme-advisory-panel-report

EMSL – The Environmental Molecular Sciences Laboratory. 2016. *Facilities Integrating Collaborations for User Science (Ficus)*. Accessed March 3, 2017 at https://www.emsl.pnl.gov/emslweb/facilities-integrating-collaborations-user-science-ficus

Estrela R and JHD Cate. 2016. "Energy Biotechnology in the Crispr-Cas9 Era." *Current Opinion in Biotechnology* 38:79-84. DOI: 10.1016/j.copbio.2016.01.005.

Forsman A. 2015. "Rethinking Phenotypic Plasticity and Its Consequences for Individuals, Populations and Species." *Heredity* 115(4):276-284. DOI: 10.1038/hdy.2014.92.

Hoose C and O Mohler. 2012. "Heterogeneous Ice Nucleation on Atmospheric Aerosols: A Review of Results from Laboratory Experiments." *Atmospheric Chemistry and Physics* 12(20):9817-9854. DOI: 10.5194/acp-12-9817-2012.

Houle D, DR Govindaraju and S Omholt. 2010. "Phenomics: The Next Challenge." *Nature Reviews Genetics* 11(12):855-866. DOI: 10.1038/nrg2897.

Kanakidou M, JH Seinfeld, SN Pandis, I Barnes, FJ Dentener, MC Facchini, R Van Dingenen, B Ervens, A Nenes, CJ Nielsen, E Swietlicki, JP Putaud, Y Balkanski, S Fuzzi, J Horth, GK Moortgat, R Winterhalter, CEL Myhre, K Tsigaridis, E Vignati, EG Stephanou and J Wilson. 2005. "Organic Aerosol and Global Climate Modelling: A Review." *Atmospheric Chemistry and Physics* 5:1053-1123.

Lee JH, JM Zachara, JK Fredrickson, SM Heald, JP McKinley, AE Plymale, CT Resch and DA Moore. 2014. "Fe(Ii)- and Sulfide-Facilitated Reduction of Tc-99(Vii) O-4(-) in Microbially Reduced Hyporheic Zone Sediments." *Geochimica Et Cosmochimica Acta* 136:247-264. DOI: 10.1016/j.gca.2013.08.017.

Lehner B. 2013. "Genotype to Phenotype: Lessons from Model Organisms for Human Genetics." *Nature Reviews Genetics* 14(3):168-178. DOI: 10.1038/nrg3404.

Lopez-Hilfiker FD, C Mohr, M Ehn, F Rubach, E Kleist, J Wildt, TF Mentel, A Lutz, M Hallquist, D Worsnop and JA Thornton. 2014. "A Novel Method for Online Analysis of Gas and Particle Composition: Description and Evaluation of a Filter Inlet for Gases and Aerosols (Figaero)." *Atmospheric Measurement Techniques* 7(4):983-1001. DOI: 10.5194/amt-7-983-2014.

Nielsen J, M Fussenegger, J Keasling, SY Lee, JC Liao, K Prather and B Palsson. 2014. "Engineering Synergy in Biotechnology." *Nature Chemical Biology* 10(5):319-322. DOI: 10.1038/nchembio.1519.

Nielsen J and JD Keasling. 2016. "Engineering Cellular Metabolism." *Cell* 164(6):1185-1197. DOI: 10.1016/j.cell.2016.02.004.

NRC – National Research Council. 2015. *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals*. Washington, DC.

Patterson JP, DB Collins, JM Michaud, JL Axson, CM Sultana, T Moser, AC Dommer, J Conner, VH Grassian, MD Stokes, GB Deane, JE Evans, MD Burkart, KA Prather and NC Gianneschi. 2016. "Sea Spray Aerosol Structure and Composition Using Cryogenic Transmission Electron Microscopy." *Acs Central Science* 2(1):40-47. DOI: 10.1021/acscentsci.5b00344.

Pincus F and M Tirrell. 2014. *Grand Challenges in Soft Matter*, University of California, Santa Barbara.

Pomraning KR, SW Wei, SA Karagiosis, YM Kim, AC Dohnalkova, BW Arey, EL Bredeweg, G Orr, TO Metz and SE Baker. 2015. "Comprehensive Metabolomic, Lipidomic and Microscopic Profiling of Yarrowia Lipolytica During Lipid Accumulation Identifies Targets for Increased Lipogenesis." *Plos One* 10(4). DOI: 10.1371/journal.pone.0123188.

Scheibe TD, EM Murphy, XY Chen, AK Rice, KC Carroll, BJ Palmer, AM Tartakovsky, I Battiato and BD Wood. 2015a. "An Analysis Platform for Multiscale Hydrogeologic Modeling with Emphasis on Hybrid Multiscale Methods." *Groundwater* 53(1):38-56. DOI: 10.1111/gwat.12179.

Scheibe TD, WA Perkins, MC Richmond, MI McKinley, PDJ Romero-Gomez, M Oostrom, TW Wietsma, JA Serkowski and JM Zachara. 2015b. "Pore-Scale and Multiscale Numerical Simulation of Flow and Transport in a Laboratory-Scale Column." *Water Resources Research* 51(2):1023-1035. DOI: 10.1002/2014WR015959.

Siretanu I, D Ebeling, MP Andersson, SLS Stipp, A Philipse, MC Stuart, D van den Ende and F Mugele. 2014. "Direct Observation of Ionic Structure at Solid-Liquid Interfaces: A Deep Look into the Stern Layer." *Scientific Reports* 4. DOI: 10.1038/Srep04956.

Tartakovsky GD, AM Tartakovsky, TD Scheibe, Y Fang, R Mahadevan and DR Lovley. 2013. "Pore-Scale Simulation of Microbial Growth Using a Genome-Scale Metabolic Model: Implications for Darcy-Scale Reactive Transport." *Advances in Water Resources* 59:256-270. DOI: 10.1016/j.advwatres.2013.05.007.

Vaden TD, C Song, RA Zaveri, D Imre and A Zelenyuk. 2010. "Morphology of Mixed Primary and Secondary Organic Particles and the Adsorption of Spectator Organic Gases During Aerosol Formation." *Proceedings of the National Academy of Sciences of the United States of America* 107(15):6658-6663. DOI: 10.1073/pnas.0911206107.

Walsh T, D Quigley and M Allen. 2011. *Grand Challenges in Understanding Interfaces between Hard and Soft Matter*, CECAM, Lausanne, Switzerland.

Wang BB, RE O'Brien, ST Kelly, JE Shilling, RC Moffet, MK Gilles and A Laskin. 2015. "Reactivity of Liquid and Semisolid Secondary Organic Carbon with Chloride and Nitrate in Atmospheric Aerosols." *Journal of Physical Chemistry A* 119(19):4498-4508. DOI: 10.1021/jp510336q.

Wang CM. 2015. "In Situ Transmission Electron Microscopy and Spectroscopy Studies of Rechargeable Batteries under Dynamic Operating Conditions: A Retrospective and Perspective View." *Journal of Materials Research* 30(3):326-339. DOI: 10.1557/jmr.2014.281.

Weber RJ, AP Sullivan, RE Peltier, A Russell, B Yan, M Zheng, J de Gouw, C Warneke, C Brock, JS Holloway, EL Atlas and E Edgerton. 2007. "A Study of Secondary Organic Aerosol Formation in the Anthropogenic-Influenced Southeastern United States." *Journal of Geophysical Research-Atmospheres* 112(D13). DOI: 10.1029/2007jd008408.

Zhang CY, K Dehoff, N Hess, M Oostrom, TW Wietsma, AJ Valocchi, BW Fouke and CJ Werth. 2010a. "Pore-Scale Study of Transverse Mixing Induced $Caco_3$ Precipitation and Permeability Reduction in a Model Subsurface Sedimentary System." *Environmental Science & Technology* 44(20):7833-7838. DOI: 10.1021/es1019788.

Zhang CY, QJ Kang, X Wang, JL Zilles, RH Muller and CJ Werth. 2010b. "Effects of Pore-Scale Heterogeneity and Transverse Mixing on Bacterial Growth in Porous Media." *Environmental Science & Technology* 44(8):3085-3092. DOI: 10.1021/es903396h.

Zhang JW, GN Presley, KE Hammel, JS Ryu, JR Menke, M Figueroa, DH Hu, G Orr and JS Schilling. 2016. "Localizing Gene Regulation Reveals a Staggered Wood Decay Mechanism for the Brown Rot Fungus Postia Placenta." *Proceedings of the National Academy of Sciences of the United States of America* 113(39):10968-10973. DOI: 10.1073/pnas.1608454113.

Zhang Q, JL Jimenez, MR Canagaratna, JD Allan, H Coe, I Ulbrich, MR Alfarra, A Takami, AM Middlebrook, YL Sun, K Dzepina, E Dunlea, K Docherty, PF DeCarlo, D Salcedo, T Onasch, JT Jayne, T Miyoshi, A Shimono, S Hatakeyama, N Takegawa, Y Kondo, J Schneider, F Drewnick, S Borrmann, S Weimer, K Demerjian, P Williams, K Bower, R Bahreini, L Cottrell, RJ Griffin, J Rautiainen, JY Sun, YM Zhang and DR Worsnop. 2007. "Ubiquity and Dominance of Oxygenated Species in Organic Aerosols in Anthropogenically-Influenced Northern Hemisphere Midlatitudes." *Geophysical Research Letters* 34(13). DOI: 10.1029/2007gl029979.

Zhou YF, J Yao, YZ Ding, JC Yu, X Hua, JE Evans, XF Yu, DB Lao, DJ Heldebrant, SK Nune, B Cao, ME Bowden, XY Yu, XL Wang and ZH Zhu. 2016. "Improving the Molecular Ion Signal Intensity for in Situ Liquid Sims Analysis." *Journal of the American Society for Mass Spectrometry* 27(12):2006-2013. DOI: 10.1007/s13361-016-1478-x.

Appendix A

Workshop Agenda

Environmental Molecular Sciences Laboratory
902 Battelle Boulevard • P.O. Box 999 • Richland, WA 99352

Breakthrough Science & Technology Workshop
AGENDA
January 10–11, 2017

Charge Questions	1. Within the areas of science identified in EMSL's Strategic Plan, what are the most critical science challenges facing the research community over the next 10 years? 2. Which key technical barriers need to be overcome to provide the research tools to enable progress on the scientific challenges? 3. What specific capabilities should EMSL invest in to address impending science and technical challenges? What near- and long-term capabilities need to be developed or acquired? What computational capability (e.g. simulation, data analytics, visualization, and imaging processing software tools) and data will be required?	
Tuesday, 01/10/2017		
EMSL Lobby	**Topic**	**Presenter/Facilitator**
07:30 - 08:00	Badging	Mary McGown
EMSL Auditorium		
08:00 - 08:10	Welcome	Allison Campbell
08:10 - 08:15	Charges, expected outcomes, and the workshop agenda	Liyuan Liang
08:15 - 08:30	Participant introductions (name, affiliation, area of expertise)	Liyuan Liang
08:30 - 10:00	Plenaries: (OPEN SESSION) - Overview of EMSL strategy (15 min) - EMSL: Integrated experimental, theoretical, computational and data science (15 min) - Microbiology and ecosystem carbon cycling (20 min) - Aerosol research to advance atmospheric science strategies for BER (20 min) - Applications of high resolution functional genomic to study biological complexity (20 min)	Lili Paša-Tolić Tim Scheibe Kirsten Hofmockel Ruby Leung Gary Stacey
10:00 - 10:15	Break	
10:15 - 12:00	**Breakout Session 1: Identify key science challenges** - What are the critical gaps in the EMSL Science Themes/Leadership areas? - What are the key unanswered questions in each research area? What are the barriers to addressing these challenges? - Are there specific barriers that, if overcome, could advance science? - What expertise is needed to enable progress on the challenges?	**Leads/Rapporteur**

EMSL 1019	Atmospheric Aerosol Systems (AAS)	Leads: John Shilling & Joel Thornton Rapporteur: Steven Ghan
EMSL Boardroom	Biosystem Dynamics and Design (BDD)	Leads: Scott Baker & Claudia Schmidt-Dannert Rapporteur: Steve Wiley
EMSL 1044	Molecular Transformations (MT)	Leads: Don Baer & Robert Hamers Rapporteur: Mark Bowden
EMSL 1075	Terrestrial and Subsurface Ecosystems (TSE)	Leads: Nancy Hess & Joy Ward Rapporteur: Kirsten Hofmockel
12:00 - 13:00 (Breakout Rooms)	Working lunch (Breakout leads to prepare 3-5 slides with key outcomes)	
EMSL Auditorium		
13:00 - 15:00	Breakout 1 outcomes: Reports and discussion	Lili Paša-Tolić & Dave Koppenaal
15:00 - 15:15	Break	
15:15 - 15:45	Overview of EMSL experimental capabilities	Dave Koppenaal
15:45 - 17:00 CISA/Lab 1102 SSFT Labs EMSL 1621/29	EMSL Tours: - Overview of EMSL biological science capabilities - Overview of subsurface transport capabilities - Overview of mass spectrometry for bio & enviro research	 Galya Orr Nancy Hess & Mart Oostrom Mary Lipton & Malak Tfaily
EMSL Auditorium		
17:00 - 18:00	Summary and general discussion	Lili Paša-Tolić
Dr. Bill's Bistro		
18:00	Dinner and Presentation: *Imaging capabilities at EMSL*	James Evans
Wednesday, 01/11/2017		
EMSL Auditorium		
08:30 - 09:00	New capabilities for new science challenges (OPEN SESSION)	Dave Koppenaal & Tim Scheibe

09:00 - 10:00	Plenaries (OPEN SESSION) - New Methodological Advances in Biological and Environmental Analytics (20 min) - Plant system biology capabilities at EMSL (20 min) - Computational strategies for BER biology (20 min)	Mary Lipton Christer Jansson Jeremy Smith
10:00 - 10:15	Break	
10:15 - 12:00	**Breakout 2: Identify key technology/computation challenges, needs and investment options in context of science challenges** - Which key technical barriers need to be overcome to provide the research tools to enable progress on the scientific challenges? - What specific new capabilities are needed to address key science challenges? - What computational capability (e.g. simulation, data analytics, visualization, and imaging processing software tools) and data will be required?	**Leads/Rapporteur**
EMSL 1019	Atmospheric Aerosol Systems (AAS)	Leads: John Shilling & Joel Thornton Rapporteur: Steven Ghan
EMSL Boardroom	Biosystem Dynamics and Design (BDD)	Leads: Scott Baker & Claudia Schmidt-Dannert Rapporteur: Steve Wiley
EMSL 1044	Molecular Transformations (MT)	Leads: Don Baer & Robert Hamers Rapporteur: Mark Bowden
EMSL 1075	Terrestrial and Subsurface Ecosystems (TSE)	Leads: Nancy Hess & Joy Ward Rapporteur: Kirsten Hofmockel
12:00 - 13:00 (Breakout Rooms)	Working lunch (Breakout leads to prepare 3-5 slides with key outcomes)	
EMSL Auditorium		
13:00 - 15:00	Breakout 2 outcomes: Reports and discussion	Dave Koppenaal & Tim Scheibe
15:00 - 15:15	Break	
15:15 - 17:30	Brainstorming / Wrap-up Identify 3-5 key near- & long-term technology/computational needs: investment prioritization	Lili Paša-Tolić
17:30	Conclusion	Liyuan Liang

Appendix B

Participants and Breakout Assignments

B.1 Atmospheric Aerosol Systems, EMSL/1019

Leads: John Shilling, ASGC-PNNL and Joel Thornton, University of Washington

Rapporteur: Steven Ghan, ASGC-PNNL

Edoardo Apra, EMSL

Susannah Burrows, ASGC-PNNL

Kolby Jardine, Lawrence Berkeley National Laboratory (via Skype)

Dave Koppenaal, EBSD-PNNL

Karol Kowalski, EMSL

Scott Lea, EMSL

L. Ruby Leung, ASGC-PNNL

Ron Pindak, Brookhaven National Laboratory (via Skype)

Doug Worsnop, Aerodyne Research

Alla Zelenyuk-Imre, PSD-PNNL

B.2 Biosystem Dynamics and Design, EMSL/Boardroom

Leads: Scott Baker, EMSL and Claudia Schmidt-Dannert, University of Minnesota

Rapporteur: Steve Wiley, EMSL

Cindy Bruckner-Lea, EMSL

James Evans, EMSL

Louise Glass, University of California, Berkeley

Samuel Hazen, University of Massachusetts Amherst

Janet Jansson, EBSD-PNNL

Ananth Kalyanaraman, Washington State University (skype)

Ryan Kelly, EMSL

Jon Magnuson, EPMD-PNNL

Lee Ann McCue, EMSL

Galya Orr, EMSL

Lili Paša-Tolić, EMSL

Lloyd Smith, University of Wisconsin-Madison

Gary Stacey, University of Missouri

Marat Valiev, EMSL

B.3 Terrestrial and Subsurface Ecosystems, EMSL/1075

Leads: Nancy Hess, EMSL and Joy Ward, University of Kansas

Rapporteur: Kirsten Hofmockel, EMSL-PNNL

Vanessa Bailey, BSD-PNNL

Harvey Bolton, EMSL

Eric Bylaska, EMSL

Zoe Cardon, Marine Biological Laboratory, Woods Hole

Dave Cowley, EMSL

Christer Jansson, EMSL

Mary Lipton, EMSL

Richard Tran Mills, Intel Corporation

Tom Metz, BSD-PNNL

Dave Moulton, Los Alamos National Laboratory

Tim Scheibe, EMSL

Alan Stone, Johns Hopkins University

Kathe Todd-Brown, BSD-PNNL

Susannah Green Tringe, DOE Joint Genome Institute

Xinning Zhang, Princeton University

B.4 Molecular Transformations, EMSL/1044

Leads: Don Baer, EMSL and Robert Hamers, University of Wisconsin-Madison

Rapporteur: Mark Bowden, EMSL

Jim De Yoreo, PSD-PNNL

Niri Govind, EMSL

James Kubicki, University of Texas, El Paso

Liyuan Liang, EMSL

Yi Lu, University of Illinois at Urbana-Champaign

Karl Mueller, PCSD-PNNL

Andrew Payzant, Oak Ridge National Laboratory

Wendy Shaw, PSD-PNNL

Jeremy Smith, University of Tennessee, Knoxville

Nancy Washton, EMSL

Todd Yeates, University of California, Los Angeles

Pacific Northwest National Laboratory (PNNL) organizations:

Atmospheric Sciences & Global Change, ASGC

Biological Sciences Division, BSD

Earth & Biological Sciences Directorate, EBSD

Energy Processes & Materials Division, EPMD

Environmental Molecular Sciences Laboratory, EMSL

Physical & Computational Sciences Directorate, PCSD

Physical Sciences Division, PSD