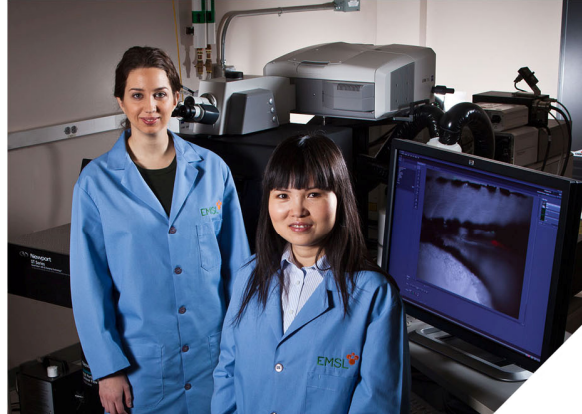


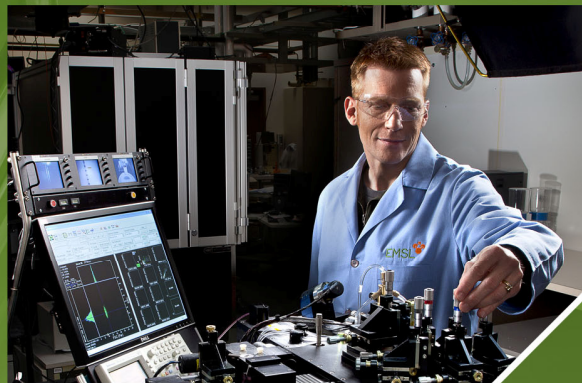
Biological Interactions and Dynamics



Science Theme Advisory Panel (BID-STAP)



May 2011



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,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/ordering.htm>



This document was printed on recycled paper.

(9/2003)

Biological Interactions and Dynamics Science Theme Advisory Panel (BID-STAP)

May 2011
Report: September 2011

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

Meeting Attendees

Thursday, May 5, 2011

Advisory Panel

Birgitte Ahring,
Washington State University, Tri-cities, Washington
 Nitin Baliga,
Institute for Systems Biology, Seattle, Washington
 Jim Frederickson
Pacific Northwest National Laboratory, Richland, WA
 Samuel Kaplan,
*The University of Texas Health Science Center at Houston
 Medical School, Houston, Texas*
 Himadri Pakrasi,
Washington University, St. Louis, Missouri
 Joel Pounds
Pacific Northwest National Laboratory, Richland, WA
 Imran Shah,
*Environmental Protection Agency, National Center for
 Computational Toxicology, Research Triangle Park, North
 Carolina (teleconference)*
 Steven Slater,
University of Wisconsin-Madison, Wisconsin
 Gary Stacey,
University of Missouri, Columbia, Missouri

EMSL

Kenneth Auberry
 Donald Baer
 David Brown Jr.
 David Cowley
 David Hoyt
 David Koppenaal
 Laura Kuprat
 Bryan Linggi
 Karl Mueller
 Galya Orr
 E. “Robby” Robinson
 Raymond Teller
 H. Steven Wiley

PNNL

Charity Plata
 Karin Rodland
 Katrina Waters

Friday, May 6, 2011

Advisory Panel

Birgitte Ahring,
Washington State University, Tri-cities, Washington
 Nitin Baliga,
Institute for Systems Biology, Seattle, Washington
 Jim Frederickson
Pacific Northwest National Laboratory, Richland, WA
 Samuel Kaplan,
*The University of Texas Health Science Center at Houston
 Medical School, Houston, Texas*
 Himadri Pakrasi,
Washington University, St. Louis, Missouri
 Joel Pounds
Pacific Northwest National Laboratory, Richland, WA
 Imran Shah,
*Environmental Protection Agency, National Center for
 Computational Toxicology, Research Triangle Park, North
 Carolina (teleconference)*
 Steven Slater,
University of Wisconsin-Madison, Wisconsin
 Gary Stacey,
University of Missouri, Columbia, Missouri

EMSL

Donald Baer
 Nancy Hess
 David Hoyt
 David Koppenaal
 Laura Kuprat
 Bryan Linggi
 Karl Mueller
 Galya Orr
 Ljiljana Paša-Tolić
 Raymond Teller
 H. Steven Wiley

PNNL

Charity Plata
 Karin Rodland

Executive Summary

On May 5-6, 2011, the Biological Interactions and Dynamics Science Theme Advisory Panel (BID-STAP) met to discuss future directions of the BID Science Theme and how to best align EMSL processes with the more proactive model for user engagement outlined in the EMSL 2.0 research plan. Overall, the STAP thought that EMSL operated well as a user facility, but there were areas that could be improved, especially if the objective is to increase its overall scientific impact. For example, the current Science Theme proposal process, with its yearly calls and funding cycles, is perhaps reasonable for a typical, reactive user facility. However, it is less suited for the new, proactive “EMSL 2.0-style” research paradigm. Because of the rapid pace of modern genomics and systems biology research, it is necessary to respond rapidly and flexibly to science opportunities as they arise. The current system is too slow.

The STAP felt that to attract the best users, it is important to reduce the “barriers to entry” for EMSL. There are many of these barriers, including difficulty in understanding the useful information that an instrument can provide, a rigid proposal process, and lack of seed money to support students/scientists in performing EMSL studies. Biologists are particularly unfamiliar with the usefulness of advanced technologies. The panel suggested that EMSL create a series of “I Didn’t Know You Could Do That!” educational videos to educate the biology community. Also, capabilities descriptions on the website should be in a form understandable and accessible to biologists. Instrument lists are useless to most biologists. Another important outreach effort is the presentation of research by EMSL staff, where they demonstrate the power of the EMSL facility. However, the focus should be on engaging specific users who could deliver the most impactful science, not simply bringing in more users.

Forming partnerships with communities of scientists will be essential for achieving the vision of the higher-impact science of EMSL 2.0. Communities tend to self-assemble around funding sources or attend common topical scientific meetings. Broadly, scientific communities can be classified as “mission-oriented” or “basic-science oriented.” Mission-oriented communities are most likely to be identified by funding source, while basic science areas tend to self-assemble around meetings, such as Gordon or Keystone Conferences. The organizers of these conferences usually are community leaders. Funding agencies, both public and private, also would know the leaders of the communities. The STAP suggests making lists of community leaders and areas of research that would be potential targets of research. Once potential “fits” are identified, the STAP suggests we simply contact several of these leaders and start a dialogue. Successful conversations eventually could evolve into workshops that would serve to define EMSL 2.0 projects.

With respect to U.S. Department of Energy (DOE) mission areas, the obvious communities are those focused on bioenergy research or groups working on systems microbiology. The most organized are the bioenergy research centers (BRCs), and they would be relatively easy to engage because of their well-defined sets of deliverables. One important aspect to consider when engaging DOE community scientists is to also engage the Joint Genome Institute (JGI). Proposing a specific project to JGI in which an important objective of the BRCs was achieved could be a significant win-win situation for both EMSL and JGI and could help win endorsement of the EMSL 2.0 concept from DOE’s Office of Biological and Environmental Research (BER). A good focus of this joint project could be the large-scale characterization of genes of unknown function.

It is important to concentrate on engaging the scientific community outside of Pacific Northwest National Laboratory (PNNL) to drive the EMSL 2.0 objectives. If EMSL 2.0 is seen by DOE as a primary mechanism to enhance PNNL research, it is unlikely to be endorsed. Still, it is necessary to engage PNNL in any EMSL 2.0 project to facilitate both the scientific careers of EMSL scientists and to assure institutional commitment. Regardless, the primary focus should be on the needs of the “outside” scientific community. Ultimately, EMSL’s greatest value to the biological community is likely to be our expertise in using multiple technical capabilities to solve complex problems. Developing this expertise, however,

requires EMSL to have a significant in-house biological research program, which it currently lacks. As a focus of in-house biology capabilities, it was suggested EMSL consider high-throughput functional screening, improved metabolomics, and the ability to integrate and analyze massive data sets.

Another area where EMSL needs improvement is data analysis. It is difficult for most biologists to understand complex data sets. Thus, they are unlikely to extract the maximum useful information. The real experts in data analysis should be EMSL staff, not just collaborators. A real expertise in helping biologists interpret the data generated by EMSL capabilities could significantly enhance EMSL's scientific impact.

Focusing significant resources on the EMSL 2.0 objectives could substantially increase the impact and agency support of EMSL, but it is important not to lose sight of EMSL's traditional role as a user facility. Even if substantial resources are not available to support a traditional user model, ways to make equipment available to individual users with good ideas should be established, especially for underutilized equipment.

Acronyms and Abbreviations

BER	Office of Biological and Environmental Research
BID	Biological Interaction Dynamics
CISA	cell isolation and systems analysis
DOE	U.S. Department of Energy
JGI	Joint Genome Institute
Kbase	Knowledgebase
NIH	National Institutes of Health
NMR	nuclear magnetic resonance
STAP	Science Theme Advisory Panel

Contents

Meeting Attendees	iii
Executive Summary	v
Acronyms and Abbreviations.....	vii
1.0 Opening Remarks	1
1.1 Research Campaigns	2
1.2 Critical Needs.....	2
2.0 EMSL’s Science Campaign Strategy and EMSL 2.0	4
3.0 EMSL’s New Biology Capabilities	5
4.0 New Capability Development in EMSL.....	7
5.0 Potential New Directions for EMSL.....	8
6.0 BID-STAP Discussion.....	10
6.1 The EMSL Proposal Process.....	10
6.2 Community Engagement.....	11
6.3 Science Drivers	12
6.4 Science Themes.....	13
6.5 Concluding Discussions	15
Appendix A.....	A.1

1.0 Opening Remarks

These conversations were conducted throughout the day on Thursday, May 5, 2011.

In introducing the 2011 Biological Interaction Dynamics-Science Theme Advisory Panel (BID-STAP), Steven Wiley, EMSL's lead biologist, provided some background to describe the goals of this current STAP meeting. Specifically, the STAPs were created in response to the U.S. Department of Energy, Office of Biological and Environmental Research (DOE-BER) onsite review in 2008, where it was suggested that the planning process for EMSL's Science Themes needed to be more structured and include input from external advisors. The STAPs were to 1) identify specific research campaigns and other means to focus the user program on specific scientific questions that will drive the development of new capabilities and 2) enhance the ability of EMSL to attract outstanding scientists. Each panel was to be led by a member of EMSL's Science Advisory Committee (SAC) and an EMSL lead scientist.

The first BID-STAP met on June 24, 2009. The panel was chaired by SAC member Professor Sam Kaplan (University of Texas, Houston Medical School) and co-chaired by Wiley. Panel members recommended that the BID Science Theme realign its scientific focus toward fundamental questions of biological complexity that can be productively addressed by the capabilities that EMSL currently has or is actively building. Specifically, it was recommended to focus on questions regarding the nature of cell heterogeneity and the impact of this heterogeneity on biological networks. Such a focus would be unique among national research programs and would address long-standing questions in the biological sciences. Success in enabling science in this research area would be a significant achievement and could drive EMSL biology over the next decade. The three recommended focus areas were as follows:

1. Understanding mechanisms of phenotypic heterogeneity in microbial populations (pure and mixed)
 - a. Discerning the relative roles of genetic variation and environmental variations in the phenotypes of individual cells in a population
 - b. Understanding the physical and chemical environments that are present at the single-cell level (e.g., pH, ion, light)
 - c. Measuring functional genomics at the single-cell level.
2. Understanding interspecies interactions
 - a. Delineating the mechanisms that underlie prokaryotic and eukaryotic interactions
 - b. Understanding how intermolecular networks are organized.
3. Predictive modeling of microbial response to the environment
 - a. Linking inter- and intra-cellular regulatory networks (metabolic, energy flow, signaling)
 - b. Scaling information from the cell to the population level.

These subtopics have served as focus areas for the yearly Science Theme call and EMSL capability development efforts. For example, EMSL has focused efforts on the development of more sensitive and specific analytical assays and new data management capabilities. Overall, this effort will make EMSL more relevant to cutting-edge research in the biological sciences.

As a consequence of the previous STAP meeting, EMSL has revisited and revised its current BID Science Themes to be the following:

- The dynamics of cellular composition and the localization and assembly of multiprotein complexes

-
- Protein modifications and how they impact cell regulatory networks
 - Molecular mechanisms that define and control the interactions between and within prokaryotic and eukaryotic cell communities
 - Understanding mechanisms of phenotypic heterogeneity in cell populations and the relative roles of genetic versus environmental factors
 - Characterizing and linking inter- and intra-cellular regulatory networks from the cell to the population level, especially those that control cell response to their environment.

1.1 Research Campaigns

One of the major recommendations of the previous BID-STAP was to launch “research campaigns,” which were seen as a way to mobilize EMSL resources to work on a set of larger, multidisciplinary problems with higher scientific impact. The other STAPs had similar recommendations. As such, a decision was made to launch several campaigns in fiscal year (FY) 2010 with the intent to work through the logistical challenges involved. Consistent with BID-STAP recommendations, the call for a research campaign was broadly advertised and had a requirement to integrate multiple EMSL capabilities. The initial call was for “concept papers” rather than full proposals to foster collaboration with EMSL staff. The choice of campaign was made using a peer review process. Currently, we are conducting a campaign in the biofuels area with Birgitte Ahring at Washington State University (WSU). This research requires the integration of transcriptomics, proteomics, and metabolomics data, so it is a good test case for large-scale data integration. Thus far, progress has been good, and we have learned much about the logistics of forming teams to work on multidisciplinary projects. It also is clear that to get the best proposals, we need to actively engage the scientific community before these campaign calls are issued.

1.2 Critical Needs

At the previous STAP meeting, there were seven points identified as being critical to EMSL’s mission outcomes in the area of biology:

1. Enhanced metabolomics focus and staff
2. Data management and integration infrastructure to handle anticipated increase in proteomics and transcriptomics data
3. Staffing the new cell isolation and systems analysis (CISA) capability
4. Top-down proteomics support
5. Imaging mass spectrometry support
6. Improved user support for data integration and analysis
7. Advertise widely.

EMSL has made significant progress in all of these areas. For example, we have installed a data analysis pipeline for transcriptomics and established methods to integrate it with proteomics data. EMSL now has a presence at several biology meetings a year, such as the DOE Genomics Science meeting. The challenge will be to continue making progress in these areas in the face of funding that likely will remain flat in the foreseeable future.

The STAP also was asked to provide advice on how EMSL can do better, in particular, how to sharpen the focus on scientific impact and the user experience. To become more expert, we need to decide what we are—and are not—going to do, such as:

- How do we enhance EMSL’s profile among DOE and the scientific community?
- Where is EMSL good at bringing value that best contributes to scientific progress?
- How do we move from being a reactive to a proactive facility?

The BID-STAP’s initial reaction was that EMSL should keep in mind how it differs from other DOE user facilities, such as the Joint Genome Institute (JGI) in California. A critical difference is that other facilities have an easily identifiable uniqueness. EMSL’s uniqueness comes from our combination of resources. To effectively use these, it is necessary to maintain a balance between the EMSL Science Themes. To maintain focus, it was suggested to pick organisms or flagship genomes and stay within those organisms or take a theme or problem and go deeper in its investigation. It also is important to remain engaged long-term with specific scientific communities.

2.0 EMSL's Science Campaign Strategy and EMSL 2.0

Donald Baer, EMSL's Chief Science Officer (interim), discussed EMSL's evolution, its progress toward EMSL 2.0, and how it ties in with DOE's vision as outlined by Energy Secretary Steven Chu. The general vision of EMSL 2.0 is to increase the pace and impact of EMSL science. In particular, to demonstrate the impact of teams of globally engaged investigators collaborating across disciplines, agencies, and organizations to address long-standing scientific challenges. Part of the goal of these STAP meetings is to add the specific scientific targets for such efforts and to obtain advice on how to best engage the scientific community.

Baer described the pillars of excellence to accelerate science that involve using advanced tools, including: 1) a closer link of theory and experiment; 2) analyzing more realistic environments; 3) innovation in data collections, analysis, storage, and sharing; and 4) exploring new paradigms and partnerships. Ultimately, EMSL wants to move away from a focus on individual users and emphasize multidisciplinary research teams focused on national problems. This could expedite research, capitalize on synergy and expertise, and improve knowledge transfer.

We still need to identify the bottlenecks that inhibit scientific advancements. Research campaigns should be a mechanism to work toward eliminating such bottlenecks. For example, the biofuel campaign between EMSL and WSU is attempting to identify the gene product responsible for isoprene biosynthesis using a variety of high-throughput data. Identifying the function of uncharacterized genes is one of the most significant bottlenecks in inferring biological functions from genomics data. Thus, this campaign seeks to develop a general approach for tackling this problem. The catalysis campaign that started this year is an integrated, multi-capability approach for high-throughput analysis of heterogeneous catalysts. It couples advanced pulse and *in situ* nuclear magnetic resonance (NMR) methods with high-level computational analyses to examine active site distributions and structural changes in support structure, as well as transformations of small molecules. Effective coupling of theory with experiment is one of the greatest challenges in surface science.

The STAP pointed out a major challenge is in knowing how to convert extreme amounts of data into information and translating that information into knowledge. When how EMSL was addressing this problem, Baer affirmed that EMSL has held workshops on handling such issues. However, the STAP pointed out it was important to define specific areas where EMSL can succeed in data management and integration. In biology, the challenges are not only related to the volumes of the data but the incomplete nature of much of it. Right now, there is an opportunity for EMSL to take a leadership position with data integration in biology.

The Panel also pointed out that EMSL should not lose sight that it is a *user* facility. The challenge is not only to choose higher impact projects but to increase the scientific impact of the projects currently being done. Do all of the group leaders come together and drill down on projects they work on? To what extent can more information be derived from specific technologies, such as NMR? What will the use of different instruments bring to a project? How can the instrumentation be improved? Learning how to enhance the quality of current projects will provide insights into the types of projects that EMSL technologies can best do in the future.

So, what does success look like? There was general agreement that if EMSL could identify important bottlenecks to progress in the science being tackled by scientific communities and successfully eliminate those bottlenecks, it could be considered "success." Success could also be measured in terms of making the broader scientific community aware of what could be accomplished by the integration of types of resources that exist at EMSL. EMSL could be a world leader in this area.

3.0 EMSL's New Biology Capabilities

EMSL Scientist Robby Robinson provided a short discussion regarding EMSL's mass spectroscopy capabilities, including the Advanced Proteomics Capability, Imaging Mass Spectrometry Capability, Ion Mobility Mass Spectrometry Capability, and next-generation Metabolomics and Metallomics Capability. Several STAP participants asked how EMSL is managing its capability increases without hiring additional personnel. The general answer was that EMSL was "coping the best it could." The STAP then pointed out that spreading the scientific expertise too thin would reduce the potential impact of the unique instrumentation. More importantly, it is important to demonstrate the unique science that can be done with the new instrumentation. Why should a user go to EMSL when she or he can go to a core facility across the hall in their own institution? In many cases, it is not obvious why EMSL's capabilities are so much better than what is generally available to users at their home institutions. Obviously, to overcome users' natural resistance to seek a distant facility to do their work, it is necessary to demonstrate the power of EMSL's capabilities.

CISA Capability Lead, Galya Orr, described CISA's new capabilities, such as the multi-photon, laser scanning confocal microscopy and fluorescence lifetime imaging (FLIM) capability, which can measure molecular interaction dynamics in live cells. A photoactivated localization microscopy (PALM) was built within EMSL and is used to characterize the interaction and fate of nanomaterials in three-dimensions with nanometer resolution. The laser-capture micro-dissection microscope enables the isolation of cells and organelles from complex cultures, communities, or tissues, while the SOLiD4 system offers transcriptomics analysis. The influx flow cytometer uses five lasers simultaneously for high-sensitivity detection and diverse sorting capabilities. Orr also mentions the helium ion microscope (HIM), the first one available at a national user facility.

David Hoyt, EMSL's NMR Capability Lead, provided a technically detailed description of EMSL's varied NMR capabilities and described how they relate to all three Science Themes. EMSL has three NMRs—600 MHz, 500 MHz, and 750 MHz—used for metabolomics. There are five NMRs used for protein structure studies of liquids. Three are used for membrane/crystalline proteins (solids) and four for metalloproteins focus (solids). Hoyt describes various NMR capabilities, including identification strategies and how to use specific NMR instruments, such as the 750 MHz wide bore NMR. Hoyt explains how NMR can examine metabolic processes as a function of time and changes in the environment and how diffusivity within biofilms can be determined as a function of depth.

Kenneth Auberry, a senior research scientist at EMSL, introduced the MyEMSL concept as a centralized data repository designed to facilitate team-based science. Core requirements of the system include:

- Universal, centralized data storage and metadata tracking
- Web interface to manage project information and data
- Flexible, easy to use, and robust infrastructure.

Auberry then provided an architecture overview and explained how MyEMSL will be used in facilitating collaboration and open-science development.

Several STAP members commented on the need for data policies and searchable metadata. Ultimately, users will not use a new system unless they are confident of data security. There also appeared to be overlap between MyEMSL and the Kbase (Knowledgebase) project currently being developed by DOE. It was noted that MyEMSL would need to be integrated into the Kbase project in some way eventually.

Auberry then shared the current MyEMSL roadmap. Beta testing showed MyEMSL could handle uploads exceeding 30 gigabytes per file. The release of MyEMSL version 1.0 occurred in summer 2011. As teams are assembled for EMSL 2.0 projects, MyEMSL is expected to be part of the core supporting technologies. The intent is eventually to have MyEMSL as an open-source project with a large community working to develop it.

4.0 New Capability Development in EMSL

David Koppenaal, EMSL's Chief Technology Officer, discussed EMSL's continuing capability development efforts. In particular, he explained the role of planning workshops and forming capability development plans. EMSL is expected to continuously upgrade its capabilities and maintain a roadmap for doing so. When the Recovery Act opportunity arose, EMSL's well-developed capability plans allowed it to successfully apply for nearly \$60 million in funding. This provided a tremendous boost in EMSL's technical capabilities. Now, we are asking the BID-STAP to support further development planning efforts by providing ideas about technical capabilities that should be pursued over the next five to 10 years.

In looking beyond our current capabilities, there are four Major Items of Equipment (MIE), or items that exceed \$2 million in costs, in the pipeline right now. One is a 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) for \$17.5 million and is expected for delivery by the end of FY 2013. A nanoSIMS, or secondary ion mass spectrometer, will be installed in early FY 2012. It will be only one of eight in the U.S. and 20 worldwide. EMSL scientists will be using that capability for localization of specific isotopes within materials and biological samples. A new molecular beam epitaxy (MBE) system is a vapor-phase capability that will be installed in FY 2012. An environmental transmission electron microscope (ETEM) is being planned that can conduct dynamic system studies and is expected by FY 2012.

Koppenaal moved into discussions regarding the various programs that EMSL offers. For example, the Intramural Development Program assembles capability development teams to develop items for EMSL. The Scientific Partner proposals engage external investigators to come to EMSL and develop new technologies and capabilities. For example, Ron Heeren, of the FOM Institute for Atomic and Molecular Physics, or AMOLF (in the Netherlands), is assisting EMSL in building a C60 imaging mass spectrometer that will take SIMS to another level. EMSL also is working on a scanning near-field microscopy system with Markus Raschke, from the University of Colorado. Raschke had the original concept and wanted to help EMSL build it. It is being perfected in his lab, and, once completed, it will be installed in EMSL and other scientists will have access to it.

Gary Stacey (University of Missouri) then proposed that EMSL think seriously about how to bring genetic screening approaches aboard because of their proven power. For instance, the National Science Foundation (NSF) recently had a workshop in automated phenotyping. One outcome was the idea that if you can measure phenotypes with extreme accuracy and make many measurements rapidly, you can look for the rare outliers. Within his research group, Stacey wants to screen 50,000 samples to look for mutations. There is a multitude of different "phenotypes" that potentially could be measured. By coupling robotics to many different types of accurate measurements, EMSL could have a significant enabling effect on biology.

5.0 Potential New Directions for EMSL

This discussion was intended to create a framework for the following day. A series of high-level questions were posed and the STAP was asked to think about them overnight and perhaps discuss them at dinner that evening.

In leading this discussion, Wiley first asked the group: *What issues do we need to consider for pushing things forward?* A variety of current EMSL efforts can be used to advance EMSL 2.0 objectives. These include new capabilities investments, intramural staff/capability development efforts, Science Theme-focused calls, and research campaigns. There can be many different types of campaigns. Our initial effort was a “simple” capability-focused campaign that provided a framework for data integration. A second-generation campaign could be developed with the specific objective of advancing the EMSL 2.0 idea. However, it is important to consider the lessons learned from the first campaign.

The first campaign was advertised widely (e.g., ads in *Science*), but there was minimal response. It seems the scientists did not understand what EMSL was offering or did not understand what it could accomplish. The first call yielded only eight white papers with the most promising ones coming from scientists already quite familiar with EMSL. Thus, it is clear EMSL must be more proactive in engaging the research community. Still, it is not clear on how that engagement should be pursued or can be achieved.

Several STAP members remarked that most research projects start with an idea. However, projects can be accomplished only after the idea is funded. Unfortunately, there is no natural coupling of EMSL to the research funding process. As such, if EMSL is asking for white papers with novel ideas, either the project with the idea must already be funded or EMSL needs to provide some method of funding support. For a project to be funded by an outside agency, the proposal usually must state the capabilities that are available and will be used as part of the project. By the time the project is funded, the needed capabilities already are set. With the current user proposal process, EMSL cannot commit to providing resources upfront. Thus, identifying the changes that may be required in the way EMSL engages users is paramount to moving forward.

Wiley pointed out that EMSL currently functions mostly as a “reactive” user facility that mostly responds to whatever proposals are submitted. In addition, the individual capabilities offered to support the biological sciences are already available at top-ranked research institutes. He also discussed some observations on the biology proposals that EMSL currently receives, including:

- Request for use of individual capabilities are the dominant type of EMSL proposal.
- The top scientists in the fields that use those individual capabilities rarely submit EMSL proposals—most likely because they already have access to them at their home institutions.
- Resource-limited investigators, mostly at second-tier institutions, drive demand for most individual capabilities.
- Demand for multi-capability projects is relatively low and mostly driven by EMSL engagement.

Thus, unless EMSL changes the way we engage users or can offer revolutionary new capabilities, this user facility will continue to have a relatively low impact in biology.

There was a consensus that scientists tend to focus on what they have at their disposal, i.e., “what’s across the hall.” Highly successful established scientists already have what they need at their disposal. As such, EMSL must offer something unique. An obvious unique aspect of EMSL is having so many different capabilities under one roof, but there must be an easy way to use these diverse capabilities as part of a single research project. Stacey made the point that he

knew of many different ways that he could use the multiple capabilities available at EMSL, but he lacked money and staff to do so. Access barriers exist both within and outside of EMSL, so it is insufficient to offer unique capabilities without addressing issues regarding user access and the ability to effectively use multiple capabilities simultaneously.

Several STAP members also mentioned that the simplest way to lower the barriers to EMSL would be to offer funding, such as for a joint postdoctoral position. If this was an incentive, users would be more willing to engage further. Another approach would be to offer a valuable service, such as data analysis. The National Institutes of Health (NIH) funds “centers of excellence” in certain areas that provides such services. The resources and specific projects they support are all peer reviewed together. This ensures that supported resources are focused on specific research projects. However, these centers offer funding for the scientists to do actual collaborative research, and these “center” users already are expected to have NIH-funded grants.

Wiley stated that the logistics of EMSL offering something similar to the NIH’s method was challenging. The research that EMSL focuses on should be something compelling to DOE or have intrinsic scientific value. It also must engage EMSL scientific staff. A largely untapped resource is EMSL’s considerable staff expertise. Organizationally, EMSL is highly matrixed with PNNL. Thus, we need to leverage both PNNL and EMSL strengths. Collaborations with users outside of PNNL need to be a win-win situation for all.

With respect to biology, EMSL’s strength and leadership is in technology development (primarily proteomics) and systems biology. The primary PNNL biology research strengths are: 1) environmental microbiology and microbial communities, 2) systems toxicology, 3) the effect of low-dose radiation on multicellular systems, and 4) biological forensics (national security). There also is growing interest in the data integration area, driven in part by the large amounts of omics data that EMSL and PNNL generate and by the recent Kbase program calls from DOE. Putting together complex networks drives much DOE research, and omics data is increasingly used to reconstruct metabolic and signaling networks. However, the best approaches for accomplishing this have not been wholly worked out.

Eventually, Wiley posed the question: *What are the opportunities in large-scale data generation, integration, and analysis that could have a significant impact on the biological sciences?* There are many opportunities in data integration and for adding value to the data generated in EMSL. One challenge is how to capture the data needed for systems biology. The data should be accessible and easy to integrate immediately after it is generated. To accomplish this, we need to link data through meaningful relationships. It also is necessary to provide tools to operate on integrated data sets. This will require organizing data into a coherent framework. If EMSL can solve the problem of structuring data in a transparent way, it would be a huge contribution to science. The advantage that EMSL has is a diverse collection of instruments, well-characterized samples, control over the resulting data and metadata, and an experienced data management group.

The afternoon session ended with the agenda for the next day, which is a discussion around three areas:

- Science Theme and Campaigns
- Alignment with Mission Areas
- Research Directions.

All subjects, areas and options should be considered. Nothing should be considered off the table.

6.0 BID-STAP Discussion

Day Two of the BID-STAP meeting, held Friday, May 6, 2011, took on a more open forum type of discussion, a record of its important discussions and outcomes follows herein.

6.1 The EMSL Proposal Process

The discussion started with the general topic of the user proposal process and how well it serves EMSL. At present, EMSL waits for proposals to be submitted. This is a reactive user facility mode. There are fundamental limitations with this, especially when trying to set up partnerships with groups of people. In essence, it requires external users to organize before they come to EMSL, which requires a considerable amount of effort. Without assurance that EMSL would cooperate, the motivations to do this degree of self-organization are considerable.

The alternative model would have EMSL reach out and form partnerships with external investigators. This also requires some alterations in our proposal model because, without some assurance that EMSL could provide resources in return for their collaboration, it will be hard to motivate collaboration. One variant of this model is what JGI does, which is identify communities that work in a particular area and invite those communities to work with them. EMSL needs to identify the communities we want to serve and invite those communities to come and discuss what EMSL has to offer. If resources are provided, that will be an enticement, but it still is necessary to lower the entry and access barriers to EMSL. This will be easier to do if you can identify people from the community you can serve uniquely.

An important consideration is the cost of doing a specific research project. EMSL itself cannot provide direct funding to investigators to perform work in EMSL, but there are certainly cost savings from using EMSL resources. The problem is that all projects require funding for the external investigators. How can the external investigators obtain funding to specifically do projects that require the unique resources of EMSL if EMSL cannot guarantee access to its facilities as part of the proposal itself? The best way would be to coordinate this with the funding agencies, such as DOE, and make an upfront commitment contingent on funding. One way this could be accomplished is through the DOE “Science Focus Area” mechanism that provides funding for staff at both national laboratories and universities to work on a common problem of interest to DOE.

One of the sensitivities with respect to coordinating research between DOE and EMSL is the potential for PNNL to have a competitive advantage relative to other groups. However, excluding PNNL from such arrangements also would be unfair. There will need to be some arrangement with DOE to focus on broad-based projects of particular interest (such as biofuels) or to work with larger organizations (such as JGI or Bioenergy Research Centers) that would eliminate issues of fairness. The important issue to keep in mind is that without some stable source of funding for collaborative projects, they are unlikely to get far.

Fundamentally, DOE-BER has a vested interest in having its funded investigators use EMSL. However, they have not taken full advantage of it. One plan is to propose to BER specific ways that EMSL could facilitate research in their programs, talk actively with BER program managers at BER, and come up with ways that this could work. The money does not have to go to EMSL, but it could be used to support external collaborators. Again, there are possible sensitivities with respect to making sure that PNNL does not take unfair advantage of such opportunities.

One potential way to identify research areas where EMSL could have a significant impact on BER programs is to hold a workshop on that specific topic. It would be most effective if there was some prospect of a funded program in this area or in identifying some areas where BER already is planning to enhance its current programs and discussing how EMSL

could enhance research projects in these areas. This approach could be used with other sponsors as well, but it would be best to start with BER.

The typical lag between the time where a user is engaged and when a project actually can be submitted, reviewed, accepted, and started can require up to a year, which many STAP members cited as a significant barrier to entry at EMSL. The ability to quickly access EMSL capabilities would increase the chances of doing cutting-edge science and lower the barriers of entry to EMSL. In addition, feasibility studies often are needed for user proposals to be competitive. Often, proposal reviewers are negative if a proof-of-concept study is not done, especially for some of the more complex capabilities in EMSL. Thus, there should be more encouragement for the “rapid access” proposal mechanism and the provision of funds to support it.

6.2 Community Engagement

To have a major, sustained impact on the biological sciences, EMSL must be engaged in a community-based science problem. However, the first step toward community engagement is to define the scientific problem that will be addressed. If it is interesting and sufficiently impactful, community leaders will get involved and encourage others to join. The problem must be something that EMSL itself is engaged in pursuing. If the community does not perceive EMSL as a full-fledged partner in the project, it will be difficult to get buy-in.

Another significant issue is the lack of a substantial biology program in EMSL. Scientists like to collaborate with other scientists. It is much more difficult to use a technical capability effectively at a remote user facility if the appropriate scientific expertise is not available there. Identification of scientific problems for team science must be done in conjunction with identification of the target communities and the building of a biology program in EMSL. In particular, biologists at EMSL should show what is possible with the capabilities here.

The best science problems for EMSL are those that BER actively is working toward solving. Thus, one approach to picking an optimal problem is to look at the proposals BER is currently funding, identify the investigators working on those proposals, and invite them to a workshop to determine the most important problems they need to solve. There should be an emphasis on long-term problems and those that require data integration and team science to solve. Currently, integration across different DOE projects is not being done effectively. EMSL can foster that integration and show people how to work together. Bringing JGI into the process also would be useful—not only because of its complementary capabilities but because of its importance to BER.

Another approach for problem identification is to engage the DOE Bioenergy Research Centers. These centers have a clear set of deliverables and are conducting multidisciplinary research in support of DOE mission areas. It seems likely they would welcome any efforts that EMSL made to support their research program. Because of their applied focus, it might be challenging to find a significant scientific problem to solve rather than just a technical one.

Before engaging any community of scientists, it will first be necessary to define our expectations. What will EMSL get out of this project? What do we expect the deliverables to be? What will be the timeline? From a strategic perspective, EMSL should be in a leadership position in a particular area of the biological sciences in the next five to 10 years. What makes sense in the context of where EMSL and PNNL could or should go? Are there new problems and challenges or aggressive mission areas done in a way that scientific communities have not done before? Address these issues and you will go a long way toward identifying the communities that EMSL should interact with.

During the meeting, a discussion ensued regarding how to define the “scientific communities” with whom EMSL could productively interact. There are the scientific communities that have grown around a set of fundamental biological

problems, such as cell division in microbes. These types of problems tend to have organized meetings associated with them, such as Gordon or Keystone conferences or Cold Springs Harbor Symposia. If you look at the organizers of these meetings and the attendees, you should have a good idea of the community membership or at least a way to engage some of its leaders. There are other communities that have organized around specific funded areas of research, such as the Energy Frontier Research Center (EFRC) membership. There also are foundations funding research in specific areas (Burroughs Wellcome Fund, Bill & Melinda Gates Foundation, Howard Hughes Medical Institute, etc.). These entities would have lists of the problems, investigators, and a timeline of the expected progress. Potentially, EMSL could work with these groups if their science driver made sense with EMSL's development plans. EMSL also could help develop scientific communities focused on solving important problems. For example, JGI played a role in developing a soybean genomics community to support the sequencing of the soybean genome. A similar effort in community development might be necessary to get the right community for EMSL interaction.

Once EMSL has identified a potential science problem(s) that makes sense with respect to where it wants to go and has identified the relevant community leaders (or strategic partners if EMSL is driving the community development effort), the next step would be to put conduct a workshop. The purpose of such a workshop would be to identify the critical "bottlenecks" in the field and what EMSL could do to eliminate such bottlenecks. However, it is important that EMSL clearly position itself as a user facility that can assist the community in its efforts. This aspect is essential because it is not desirable to change EMSL's image or DOE's perception of it. Remember the cautionary note: "Do not let the cart drive the horse." Make sure that the people are not confused as to what EMSL is trying to accomplish.

6.3 Science Drivers

It became apparent that the choices of science drivers for EMSL were going to be crucial for community engagement. Thus, an active discussion ensued on what types of science drives made the most sense. One obvious area was the interaction of nanomaterials with biological systems. It combines biology, chemistry, and interfaces and plays to current PNNL strengths. However, defining specific questions to be addressed would be key. More broadly, the whole area of predictive toxicology is one in which EMSL could play a seminal role. This area is focused on predicting the toxicity of a chemical compound from first principles. It uses chemistry, computer simulations, metabolic modeling, physiology, and an understanding of gene regulatory networks. This is an extremely important area of research, but the timelines for progress in this area might be too long. Community funding also may be an issue.

"Synthetic biology" is another potential science driver area. BER is becoming interested in synthetic biology because of its implications for redesigning biological systems for biofuel production (among other applications). However, we would need to define EMSL's role in such an area. There is a need for standardized "platform microbes" and characterization of its response to manipulation—but is this a science driver? More generally, the need to understand manipulated systems, either natural or man-made, is a problem in systems biology. We need a comprehensive understanding of systems properties. Then, we can predict how to put the right part into the right chassis. There are consequences if we do not understand what happens if we put a foreign part into an organism. When that is understood, it will be much easier to go into the iterative science process.

An underlying problem associated with synthetic biology is the annotation of genes of unknown function. If people do not know what the natural parts of cells actually do, how can intracellular molecular networks be redesigned with the expectation of a predictable result? What contribution could EMSL make to solving this problem? What are the specific approaches that relevant communities are taking to achieve this? What are their bottlenecks? JGI is actively involved in this area of research, so partnering with them on this problem would make sense. JGI also would have a good idea regarding the challenges that the field faces in this area. Together, we could better determine the critical research

bottlenecks in this area. Regardless of the approach that other investigators are taking, the STAP felt that EMSL's analytical and mathematical aspects are ideally suited for tackling this problem. However, if this problem is pursued, capabilities for functional and molecular characterization of microbes will be needed. The specifics of the problem being tackled will also help define the new capabilities needed at EMSL.

Regardless of the science driver, EMSL can serve as a "platform" for driving high-risk, high-impact science. For scientific communities, EMSL could offer not only the technical infrastructure for tackling difficult problems but also the organizational infrastructure. Identifying and bringing together teams of appropriate scientists is difficult and time-consuming. By taking on this function, EMSL can be a catalyst for change.

6.4 Science Themes

The panel also discussed the current BID Science Themes and how well they were helping EMSL to effectively focus on high-impact science. The current Science Themes are:

- The dynamics of cellular composition and localization and assembly of multiprotein complexes
- Protein modifications and how they impact cell regulatory networks
- Molecular mechanisms that define and control the interactions between and within prokaryotic and eukaryotic cell communities
- Understanding mechanisms of phenotypic heterogeneity in cell populations and the relative roles of genetic versus environmental factors
- Characterizing and linking inter- and intra-cellular regulatory networks from the cell to the population level, especially those that control the response of cells to their environment.

Wiley asked whether these were too broad or if there were too many of them. On some, such as in the area of protein modifications, EMSL receives many proposals. In others, such as understanding mechanisms of phenotypic heterogeneity, EMSL receives few (if any). Should the themes be changed?

The panel was asked whether EMSL was receiving the types of proposals it was hoping for and if EMSL felt it was engaging the appropriate community of users. Specifically, how many of the proposals EMSL received were focused on the Science Themes? If the themes were broadened, would that increase the number and variety of proposals that were received? Would this be a good thing?

Most of the proposals that EMSL receives are Science Theme proposals because EMSL issues a specific call for these. Considering that only about 60% are funded, broadening the theme would not appear to be a reasonable approach. The discussion then centered on why the Science Themes should be important. They should focus EMSL's efforts on areas of science in which it could have a significant impact. If there are few proposals submitted to a particular topic, this suggests that more effort needs to be expended on engaging the community working in that area. For example, in the area of understanding mechanisms of phenotypic heterogeneity, EMSL needs to decide if it is going to abandon or pursue it. If the decision is to pursue it, we need to hold a workshop and push the problem. EMSL needs to get people enthusiastic to submit a proposal in this area. This type of targeted push also could assist in defining research campaigns.

Not all biological capabilities in EMSL are fully utilized. For example, proteomics is oversubscribed, but imaging capabilities are underutilized. The question then became if the Science Theme calls should be directed toward these capabilities. This would help with the awareness as to their existence and could shift demand from more heavily used

capabilities. One reason why the proteomics capabilities are in high demand is because biologists are aware of what they can do. However, to shift demand to new instrumentation may not be a simple task. First, there is an issue of having sufficient staff. EMSL tends to hire staff to cover high-demand capabilities, not vice versa. Second, it is difficult to shift work to other capabilities after a proposal is submitted because of the way resources are allocated to proposals. The short time window between the Science Theme call and proposal submission means that investigators must be aware of the underutilized capabilities and encourage others to use them well ahead of time. It was suggested that EMSL initiate an outreach effort to inform potential users on not only the types of instrumentation within EMSL but what can be done with them.

Educating potential users on what can be done with EMSL capabilities would help build a foundation for recruiting higher-impact users. One suggestion would be to film short YouTube videos on what users could learn from different capabilities (i.e., the “Wow, I didn’t know you could do that at EMSL!”). If EMSL could integrate information from these lesser-used capabilities with some of the more commonly used ones, it could provide something important to show potential users. This type of “integration education” also could assist and inform the EMSL 2.0 team science effort.

The discussion then moved toward the issue of data analysis and integration. The STAP felt that to attract the average biologist to work effectively within EMSL, the data EMSL provides must be made “biologist friendly.” There must be an interface mechanism with a staff person available to help biologists interpret their data. Primarily, this is due to the types of biological data that EMSL generates (large amounts of complex data), which is quite different than data biologists likely generate from their own laboratories. Currently, data analysis at EMSL is ad hoc.

There was general agreement that data analysis and integration needs to be more of a priority within EMSL. It was pointed out by several STAP members that one of the greatest difficulties in coordinating the efforts of teams of biologists was the problem with data exchange and integration. If EMSL had a centralized capability to store, manage, analyze, integrate, and visualize data from multiple groups, this could help drive team-based scientific problems. However, it is also important to have trained staff who can interpret this type of data appropriately. JGI is starting to confront this problem as well and has made some progress in remedying it over the last year.

Data integration is a bottleneck for many of the biological sciences because of the newness of the multiple omics technologies and the general lack of experience on how to best use multiple types of data to solve biological problems. To tackle this, EMSL will have to appropriate domain experts in each of the different types of biological data. Perhaps, EMSL 2.0 biology should be focused on integration and analysis of large-scale data sets. The way that biologists currently tackle this issue is simplistic (top five and bottom 10 genes that change). Clearly, this is an area of intense interest to JGI as well. Again, a partnership between EMSL and JGI to work on this problem would make sense.

It was noted that the two Science Theme topics that receive proportionally fewer proposals require a good deal of data integration and analysis. Perhaps, the lack of proposals on these topics is due to the difficulty of managing the type of data they generate. The higher-level issue is how to turn data into biological understanding. This always will be domain-specific, so it will be necessary to choose the area that EMSL wants to concentrate its effort on and either hire the right staff or create necessary partnerships.

High-throughput phenotyping could be important in interpreting genomics data. Genomics technologies provide tools to find differences between cells and communities, but you need to know how these genetic differences translate into functional differences. EMSL could develop the tools to evaluate these differences and analyze the data.

Phenotypes could be proteomics, metabolomics, and transcriptomics differences, but EMSL should not restrict itself to these. These omics technologies are restricted to the number of samples that can be analyzed. Here, high-throughput

imaging technologies could be important because of the ability to analyze many different types of samples rapidly (e.g., flow cytometry). If a rapid, imaging-based phenotype tool identified the outliers in a population, you then could use the omics tools to delve deeper into the underlying phenotypic differences. However, the specific capabilities needed to assess cell functions will depend on the user group engaged and the problem(s) they are investigating.

6.5 Concluding Discussions

It will be important for EMSL to be proactive in identifying scientific communities to form partnerships. However, it also is important for EMSL staff to go out and present the work they do at conferences. Scientists need to know about the incredible capabilities at EMSL, and we need to get people thinking about new ways of approaching problems and integrating capabilities.

EMSL staff should not simply think of themselves as helping users. Instead, they should consider themselves scientific collaborators who contribute knowledge and keen insight. They should be expected to share authorship on user papers. This also will increase their overall visibility and help them obtain funding for their own research.

One of the STAP's primary directives was: do not wait until the "perfect" opportunity is identified. For EMSL, it is more important for our staff to initiate engagement. We need to talk to the Bioenergy Research Centers because they already are well aligned with DOE research areas. We must identify one or two basic science targets to pursue, as well as the experimental and computational capabilities required to move research forward.

One final point: JGI is an obvious partner. Everywhere JGI wants to go in the future overlaps with where EMSL already is. Instead of waiting for conflict, we should be proactive and start a partnership with them that makes sense. DOE should be in favor of that. Because of JGI's involvement in the Bioenergy Energy Centers, it might be possible to create a joint project with both JGI and one or more centers to solve a problem of high importance to DOE.

Appendix A

Appendix A

EMSL BID-STAP Meeting Agenda May 5-6, 2011

May 5, 2011

EMSL Boardroom

- 12:30 pm Badging – ETB (foreign nationals), EMSL (U.S. citizens)**
- 1:00 pm Welcome and Plans for the Meeting – Steven Wiley**
- Introductions
- 1:30 pm EMSL's Science Campaign Strategy and EMSL 2.0 – Don Baer/Steven Wiley**
- Science Campaigns and EMSL 2.0
- 2:00 pm New Biology Capabilities in EMSL**
- Proteomics – Robby Robinson
 - Cell Isolation and Systems Analysis – Galya Orr
 - NMR – David Hoyt
 - MyEMSL – Kenneth Auberry
- 3:45 pm New Capability Development in EMSL – Dave Koppenaal**
- 4:15 pm Potential New Directions for EMSL – Steven Wiley**
- Science Themes
 - Project engagement versus community engagement
 - Role of computation in EMSL 2.0
 - Funding challenges
- 5:00 pm Wrap-up**
- 6:30 pm Dinner at Tagaris**

May 6, 2011

EMSL Boardroom

- 7:30 am Morning Refreshments**
- 8:00 am Science Themes and Campaigns**
- How well do our science themes drive excellent science?
 - Going from reactive to proactive science
 - Biology stakeholders and their needs
- 9:00 am Alignment with Mission Areas**
- Fundamental science (e.g., systems biology)
 - Bioenergy research
 - Other DOE mission areas
 - Computational toxicology
 - Medical research

10:00 am Research Directions

- What is working well?
- Potential new opportunities
- Partnerships for new opportunities

11:00 am Summary and Recommendations

- Identify 1 or 2 Science Campaign targets
- Review Science Theme focus areas
- Identify needed experimental, computational capabilities

12:00 pm Working Lunch

2:00 pm End

Optional – Tour of new EMSL capabilities