EMSL Research Campaigns

Status Report on the Development of Research Campaigns

June 2013

Pacific Northwest National Laboratory
Richland, Washington 99352

PNNL-22924

EMSL Research Campaigns

Status Report on the Development of Research Campaigns

June 2013

Pacific Northwest National Laboratory
Richland, Washington 99352

Acronyms and Abbreviations

1-D	one-dimensional
AAS	amorphous aluminosilicate
AMT	Accurate Mass and Time
BER	Biological and Environmental Research
CF	continuous-flow
CP	cross-polarization
CPMG	Carr-Purcell-Meiboom-Gill
DFT	density functional theory
DMAPP	dimethylallyl diphosphate
DOE	Department of Energy
DXP	1-deoxyxylulose-D-5-phosphate
EMSL	Environmental Molecular Sciences Laboratory
EPR	electron paramagnetic resonance
GBSS	Geochemistry, Biogeochemistry, and Subsurface Science
GRAS	generally regarded as safe
HPLC-MS/MS	high performance liquid chromatography tandem mass spectrometry
iTRAQ	isobaric tag for relative and absolute quantitation
LV	latent variables
MAS	magic-angle spinning
MBGC	membrane biology grand challenge
MQMAS	multiple quantum magic angle spinning
MS	mass spectrometer or spectrometry
NAD+	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
NMR	nuclear magnetic resonance
PI	principal investigator
PLSR	partial least squares regression
PNNL	Pacific Northwest National Laboratory
POM	polyoxometalate
RNA	ribonucleic acid
SCR	selective catalytic reduction
SEX	starch excess
STAP	Science and Technology Advisory Panel
TRS	trifluorosilane
WSU	Washington State University

Contents

1.0   Overview and Objectives

Research campaigns were conceived as a means to focus EMSL research on specific scientific questions. Campaign will help fulfill the Environmental Molecular Sciences Laboratory (EMSL) strategic vision "to develop and integrate, for use by the scientific community, world leading capabilities that transform understanding in the environmental molecular sciences and accelerate discoveries relevant to the Department of Energy's (DOE's) missions." Campaigns are multi-institutional multi-disciplinary projects with scope beyond those of normal EMSL user projects. The goal of research campaigns is to have EMSL scientists and users team on the projects in the effort to accelerate progress and increase impact in specific scientific areas by focusing user research, EMSL resources, and expertise in those areas.

The concept of campaigns was evaluated by the three Science Theme Advisory Panels held in 2009. Each panel indicated that campaigns could be a useful mechanism for reaching EMSL's strategic objectives. They noted that, "… important research campaigns would likely take different forms and have varying sizes depending on the science area and nature of what is needed for advancement. Some campaigns might be technique oriented while others would address specific science questions," (EMSL Science Theme Advisory Panel Meetings Report, PNNL-19080). We developed a three-phase approach for introducing campaigns in EMSL (Figure 1.1). Because of the importance to EMSL of capability integration and multi-capability approaches to research, the first-generation research campaigns were focused primarily on developing new linkages of capabilities. The second and later generations of campaigns were to focus on specific scientific questions where a multi-institutional effort taking advantage of EMSL capabilities could be expected to lead to important advances.
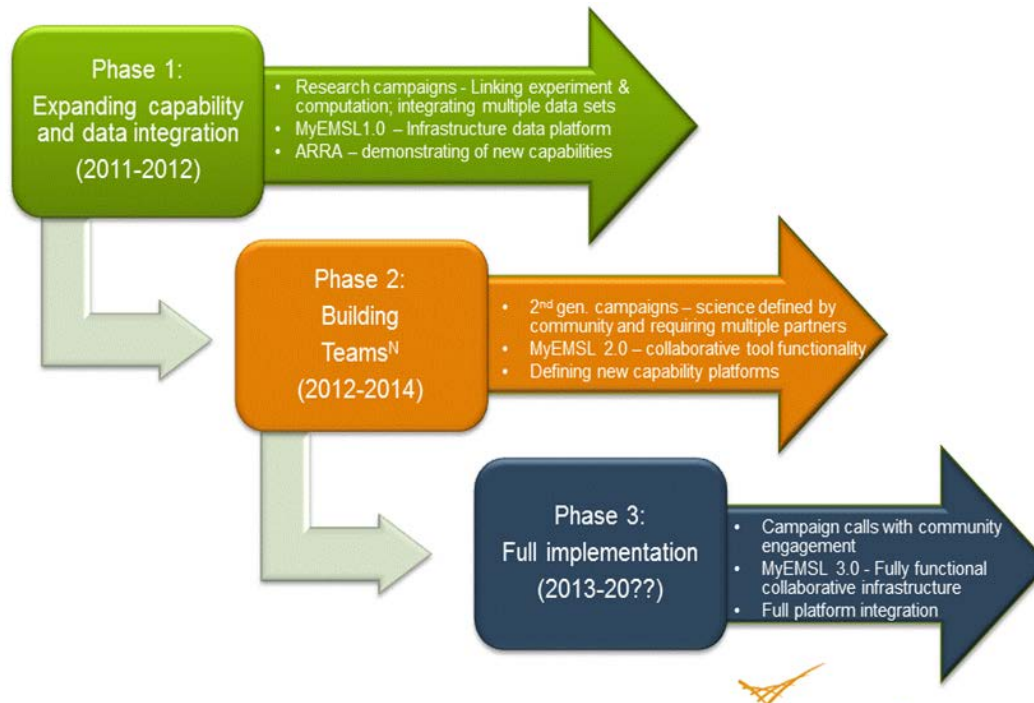

Figure 1.1. Staged approach to developing research campaigns (from 2011 Biological and Environmental Research (BER) Review Document).

1.1 First-Generation Campaigns

We recognized that although many user-proposed research projects in EMSL use multiple instruments and technical capabilities, there are additional opportunities to truly advance science and speed scientific progress in several scientific areas by enhancing the ability to rapidly integrate the significant amounts of multi-instrument experimental data that can now be collected often in combination with the advanced modeling that is now available. Both the progress and challenges were highlighted in a review of changes in science between 2000 and 2010 when *Science Magazine* noted that *"In the past 10 years, new ways of gathering, analyzing, storing, and disseminating information have transformed science. Researchers generate more observations, more models, and more automated experimentation than ever before, creating a data-saturated world."* Because of the need to enhance our ability to process and integrate data of various types, we decided that as we were learning how to best formulate and conduct research campaigns it was appropriate to focus the first efforts on projects designed to expand our experience at integrating data from multiple-technologies (experiment and computation).

Two specific areas were identified for which capability integration was considered important both for advancing science and enhancing EMSL experience and capability. As highlighted in a 2010 call for campaign proposal white papers, these were: 1) biological studies in which proteomics data would be integrated with transcriptomics data to answer fundamental questions regarding the dynamics of cellular networks and how the flow of energy, matter, and information is regulated or 2) interfacial studies that linked computational capabilities with dynamic measurements at molecular interfaces to address important questions related to catalysis, atmospheric chemistry, geochemistry, biogeochemistry, emission sequestration, or energy production, transformation, and storage.

In spring 2010, EMSL issued a formal call for first-generation campaign proposals. Specifically, EMSL requested concept papers for projects that would exploit multi-technologies and data integration capabilities to significantly advance understanding of specific scientific problems in either biology or the interfacial sciences. EMSL worked with two teams to develop full proposals in the areas of linking advanced nuclear magnetic resonance (NMR) capabilities and computation for catalysis studies (Penn State University, University of California [UC]-Berkeley, and Pacific Northwest National Laboratory [PNNL]) and integration of multiple omics data for microbe studies related to isoprene production (PNNL and Washington State University [WSU]). As a result of strong peer reviews, campaigns were initiated in each area in December 2010.

NMR/Modeling/Catalysis − The catalysis-related campaign focused on coupling advanced pulse and in-situ NMR methods with high-level computational analyses to examine active site distributions and structural changes in support structure and transformations of small molecules. This campaign aimed to expand the application of advanced NMR methods to catalysis by combining:

- New and advanced NMR pulse sequences (multiple quantum magic angle spinning, or MQMAS; cross-polarization Carr-Purcell-Meiboom-Gill, or CP-CPMG)

- Application of new in-situ probes (high-temperature, continuous flow MAS and high-pressure/high-temperature [350°C])

- Computation to understand reactive site distributions (e.g., amorphous aluminosilicates and aluminas) and water attachment at specific sites

- Data integration and informatics.

This campaign was successful in bringing new capabilities to EMSL that are of high importance to catalysis research and applied these capabilities to specific research problems. As a result of this campaign:

Four new capabilities have been developed, applied as part of the campaign, and are now incorporated into the user program:

- CP-CPMG pulse sequences implemented and applied. These pulse sequences will enable increased resolution of similar sites and detection of sparse spin systems (e.g., low analyte concentration)

- New elevated temperature flow probe established for NMR. This probe for the 850 MHz spectrometer can be used to monitor chemical reactions as they occur and structural changes.

- New elevated temperature flow cell developed for electron paramagnetic resonance (EPR). This cell enables *in-situ* monitoring of chemical reactions, structural changes as a function of hydration, and redox changes.

- New codes implemented in NWChem and validated for EPR tensor calculations. These new codes allow more detailed interpretation of experimental EPR data.

In response to the annual EMSL call for proposals, multiple user proposals requested use of the constant flow (CF) NMR probe (on the 850-MHz solid-state NMR instrument) and/or CF EPR capability. The new NWChem codes will be integral parts of successful proposals.

- Four manuscripts have been published or are in advanced stages of preparation:

  o Andersen A, ED Walter, F Gao, N Washton, N Govind, J Autschbach, CHF Peden. In preparation (2013). "Combined electron paramagnetic resonance (EPR) and density functional theory (DFT) characterization of Cu(II)-SSZ-13 for selective catalytic reduction (SCR)."

  o Washton NM, A Andersen, KT Mueller. In preparation (2013). "Silanization binding sites on aluminosilicate gels determined by solid-state nuclear magnetic resonance and *ab initio* studies."

  o Mei D, VM Lebarbier, RJ Rousseau, VA Glezakou, KO Albrecht, L Kovarik, MD Flake, and RA Dagle. 2013. "Comparative Investigation of Benzene Steam Reforming over Spinel Supported Rh and Ir Catalysts." *ACSCatalysis* 3(6):1133−1143

  o Mei D. 2013. "First principles characterization of formate and carboxyl adsorption on the stoichiometric $CeO_2(111)$ and $CeO_2(110)$ surfaces." *Journal of Energy Chemistry* 22(3):524−532.

Integrating Biological Data − The second campaign aimed to push development forward for methods of integration and analysis of complex biological data sets including transcriptomics, proteomics, and metabolomics to facilitate the discovery of the key enzymes related to isoprene biosynthesis. The long-term goal of the research is to substantially increase the yield of isoprene using genetically modified bacterium, *Bacillus subtilis*. Central hypotheses being tested include the following:

- The formation of isoprene is tightly controlled by unknown enzymes (e.g., isoprene synthase) and unidentified regulatory factors (the ongoing work).

- By knocking out repressors and overexpressing activators in the 1-deoxyxylulose-D-5-phosphate (DXP) pathway, significantly higher levels of isoprene production could be achieved in *B. subtilis* (future work).

Although this campaign addresses an important scientific challenge in understanding key regulatory steps in isoprene production, the major goal was to accelerate the development of methods for integrating multiple types of omics data in order to accelerate the discovery of key enzymes. Significantly, this approach could provide an improved method for functional genome annotation and functional characterization of biological dark matter.

EMSL is a highly unique and advantageous environment for integrating the data from multiple instruments. Having the instruments for sample preparation and analysis under one roof simplifies the logistics of sample handling and eliminates the need for multiple sample derivations. However, the degree to which data integration was applied had not been previously attempted in EMSL and was accomplished by establishing workflows for data analysis of combined data sets. The campaign highlighted both advantages and challenges (both technical and logistical) of the multi–modal analysis efforts. The campaign produced results and helped develop capabilities that would not have been available without the campaign.

Two new capabilities have been developed, applied as part of the campaign, and are now available to the user program:

- High performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) to detect the DXP pathway intermediates. This capability enables detection of low levels of metabolites that have not been detected without the use of custom probes.

- Detection of low abundance proteins using isobaric tags for relative and absolute quantitation (iTRAQs) proteomic approaches. This significantly expands the sensitivity of protein detection to allow for quantitation of all of the DXP pathway proteins as well as a large view of global protein abundance.

Four manuscripts have been published or are in advanced stages of preparation:

- Hess BM, J Xue, LM Markillie, RC Taylor, HS Wiley, BK Ahring, and B Linggi. 2013. "Coregulation of terpenoid pathway genes and the predictive capacity of transcriptomics profiles related to isoprene production in *B. subtilis*." *PLoS ONE*. doi:10.1371/journal.pone.0066104

- Xue J, N Isern, BM Hess, B Linggi, HS Wiley, and BK Ahring. In preparation (2013). "Characterization of metabolic profiles using 1H NMR spectroscopy and HPLC-MS/MS facilitates the strain improvement to overproduce isoprene in *Bacillus subtilis.*"

- Hess BM, J Xue, J Aldrich, S Wu, L Pasa-Tolic, BH Ahring, HS Wiley, and B Linggi. In preparation (2013). "Overexpression of the geranyltransferase IspA in *Bacillus subtilis* decreases genetic competence."

- Hess BM, S Wu, L Pasa-Tolic, BH Ahring, HS Wiley, and B Linggi. In preparation (2013). "Regulation of sporulation pathways in Bacillus subtilis overexpressing IspA."

Reports on these two campaigns are provided in Sections 2.0 and 3.0.

1.2 Second-Generation Campaigns

The expanded objective for second-generation campaigns starting in 2012 was assembling multi-institutional teams to focus on critical scientific issues or challenges. Several campaign possibilities had been suggested by EMSL users, the EMSL Science and Technology Advisory Panels (STAPs), and members of the Science Advisory Committee. Small workshops were conducted as part of the process of identifying the important science questions and the possible relevance of multi-institutional involvement in addressing these challenges. The nature of these second-generation campaigns reflects the science challenges unique to each area. Each of these campaigns represents a true collaborative-team effort with each of the participating institutions participating through the use of its current research activities with the objective to achieve a research impact beyond what each group could individually achieve.

High-level descriptions of three campaigns initiated in 2012 are provided below.

1.2.1 Pore-Scale Campaign – Accelerate and Validate the Development of Pore-Scale Modeling

The December 2010 Geochemistry, Biogeochemistry, and Subsurface Science (GBSS) STAP identified a potential campaign related to modeling flow and reactive transport at the pore scale. Several independent groups have developed simulations of multiphase flow and reactive transport at the pore scale—most often without the benefit of experimental data to validate their models. The STAP suggested that a focused effort with EMSL providing data using its in-house flow and transport capabilities to enable pore-scale modelers to test, calibrate, and validate computer codes would be a worthy campaign. After discussing this idea with modeling groups around the country to assess interest and importance, EMSL hosted a two-day Pore-Scale Modeling workshop on August 9−10, 2011, with 40 participants from 14 institutions to assess the need and possible nature of such a campaign. (See Appendix B for a short workshop summary).

The workshop participants concluded that a model benchmarking effort associated with 1) transverse fluid mixing and 2) fluid displacement would be valuable for validating and understanding the limits of the range of pore-scale models that have been developed. The effort should include definition or "learning sets" of mutually beneficial pore-scale experimental studies to calibrate numerical models and "challenge sets" designed to numerically predict results without knowing the experimental results. The experimental data would be generated using micromodels fabricated and flow experiments conducted using EMSL's new microfabrication capability in tandem with EMSL's integrated subsurface flow and transport capability. The benchmark experiments would focus on fundamental concepts, from pre-experiment modeling to hydraulic characterization, analytical chemistry, numerical modeling, and analysis.

Based on this workshop recommendation, a research campaign was designed to generate five micromodel experiments that will produce data necessary to test and verify both pore-scale and continuum-based simulators. This campaign was initiated in February 2012, and the first data sets appeared online in June 2012. Currently, 10 research groups are participating in the first stage of the challenge, and a multi-institution manuscript is being prepared to be submitted to the journal *Computational Geosciences*. This journal has agreed to publish a series of three comprehensive papers that will be generated by the campaign (one for each stage of the campaign).

1.2.2 Rhizosphere Campaign: Understanding Dynamic Interactions at the Plant-Soil Interface in the Context of Environmental Change

Global climate change may affect plant growth in various ways. Water availability and soil temperature influence plant growth and affect plants' ability to cycle plant nutrients such as carbon and nitrogen. Because plants lack the ability to escape from detrimental environmental changes, they must evolve ways to adapt to such changes. The plant-soil interface (the rhizosphere) is the location where some of these adaptations can be studied. Gaining a mechanistic understanding of the role of subsurface processes (e.g., microbiology geochemistry, root/rhizosphere, soil processes) in the terrestrial carbon cycle is a goal of the climate and environmental sciences division of Biological and Environmental Research (BER). In December 2011, EMSL hosted a small multi-institutional workshop. The resulting recommendation was to create a 12- to 18-month pilot campaign that would point EMSL toward the long-term goals to 1) demonstrate the feasibility of using metagenomic approaches to examine the microbial community of bioenergy crops grown at elevated temperatures and with water deprivation; 2) quantify and examine nutrient partitioning into roots and the exchange of nutrients with the soil microbial community; 3) examine the effect of the plant genotype on the soil microbial community (and vice versa) to mitigate the effects of environmental change; and 4) use the information gained to develop quantitative models of nutrient dynamics within the rhizosphere.

The workshop results suggested a specific short-term campaign objective (12- to 18-months) to compare *A. thaliana* wild-type and SEX-like (SEX=starch excess) mutants. These mutants lack the ability to allocate starch (stored carbon) from their leaves to roots during the dark period. Research has shown that the growth of SEX-like mutants is comparable to that of wild-type (because of normal carbon metabolism during the day) but differs significantly from the wild-type in carbon allocation on a diurnal cycle to the roots. Hence, these plants represent an experimentally tractable system in which to analyze directly the role of plant carbon allocation and root exudation in the control of rhizosphere microbial populations.

This proposal aligned with BER goals on the influence of climate change on subsurface parameters and environmental impacts, and making full use of BER genomic and analytical capabilities. A campaign was initiated in October 2012. This campaign involves participation of many different institutions. Plants are being grown hydroponically in artificial potting mix (e.g., sand mixture) and in soil (University of Minnesota and University of Missouri). The rhizosphere bacterial population arising from inoculation with soil extract will be analyzed using a variety of analytical tools to characterize the live interactions of plant roots with the soil microbial population. Results will include 1) characterization of rhizosphere populations by metagenomic sequencing of 16s rRNA; 2) using microscopic methods to image the rhizosphere using a suite of microscopy methods; 3) using $^{12}C/^{13}C$ labeling methods to examine proteomic and metabolomic components; 4) high throughput sequencing of the transcriptome; and 5) ^{11}C labeling to measure carbon allocation from shoot to roots and follow the metabolic pathway of carbon. The campaign will fully use the resources of EMSL (omics measurements and imaging) and those of collaborating organizations to address the complexity of the plant-soil interface. In addition to EMSL and PNNL, participating institutions include the University of Missouri (G Stacey), Brookhaven National Laboratory (R Ferrieri), Joint Genome Institute (S Tringe), Washington University (H Pakrasi), and the University of Minnesota (M Sadowsky).

1.2.3 Cyanobacteria Campaign: Use of Synthetic Biology to Probe Molecular Machines in Photosynthesis

From 2005 to 2008, Himadri Pakrasi led a multi-institutional team of researchers in a membrane biology grand challenge (MBGC) project focused on obtaining a systems-level understanding of cyanobacterial membrane structure and processes, particularly as they affect and regulate signal transduction, carbon and nitrogen fixation and energy storage, hydrogen production, and metal ion homeostasis with *Cyanothece* sp ATCC 51142 as the model system of study. The MBGC project focused on understanding the diurnal behavior of a group of cyanobacteria at the systems level. In addition to the scientific achievements, the project demonstrated the value of team science (http://www.emsl.pnnl.gov/news/viewArticle.jsp?articleId=277). Inspired by the success of the MBGC project, grand challenge participants discussed concepts for future projects that could leverage the newest capabilities in EMSL at the MBGC close-out workshop in the spring of 2011. This and follow-on discussions led to a proposal "Use of synthetic biology to probe molecular machines in photosynthesis" that was presented to EMSL in May 2012.

The campaign they proposed builds upon the knowledge gained from what is now known as MBGC 1. Cyanobacteria provide distinct advantages over other prospective candidates for bioenergy production because they are readily amenable to genetic modifications, possess high photosynthetic efficiencies, and can be optimized for growth and productivity. The molecular design of the structures that enable these organisms to harvest energy, reducing power and carbon, have been elucidated at sub-nanometer resolution (3−7) and thus set the stage to guide further optimization of these biomolecular machines with the use of synthetic biology approaches. Different methods, such as metabolic engineering, have long been feasible and resulted in a large collection of cyanobacterial mutants with modified light-capture capabilities, photosynthetic efficiencies, and carbon dioxide fixation rates. Decoding the cellular language that controls these specific metabolic operations will provide the framework to elucidate the design principles that drive the production of renewable energy sources in a more efficient way.

The objective of their new proposal is to characterize and improve the molecular machines in photosynthesis that govern bioenergy production in cyanobacteria. Several sets of experiments and analysis are directed toward the goal using sophisticated imaging instrumentation available at EMSL. The profound expertise of the External Project Team in the areas of cyanobacterial and systems biology will be leveraged to attain the objectives of the project goals. The planned experiments for this research campaign focus on the products of a synthetic biology approach targeted in the following important areas of cyanobacterial biology; the goal of each having been designed to provide new approaches to increasing photosynthetic productivity: 1) photosynthetic antenna modification, 2) photosynthetic electron transport and photosystem stoichiometry, and 3) topology of photosynthetic membranes.

Based on the team track record and the proposal evolving from the 2011 Workshop, a campaign was initiated in June 2012. The campaign team has already demonstrated success and productivity in the EMSL MBGC effort and is strongly supportive of collaborative-team science. The principal investigators (PIs) will leverage their current project funding and work closely with EMSL staff to advance the project goals articulated above. The external PIs and their students will visit EMSL frequently to ensure project progress and productivity, and they will attend and participate in an upcoming workshop being held at Washington University in August 2013. Participating institutions include Washington University (Pakrasi), Purdue University (Sherman), and St. Louis University (Aurora).

1.3 Lessons Learned

As suggested by the phased approach to the implementation of campaigns, we expected that the inclusion of collaborative-team campaigns in EMSL user activities would not necessarily be simple or straight forward. First- and second-generation campaigns experienced challenges in the ways in which the campaigns were managed and communication among campaign participants was handled. Lessons learned during the initial campaigns applied as the second generation of campaigns were being formed. These lessons can be roughly categorized as being associated with team formation and function, team function, campaign accountability, and capability access.

Team Formation and Function

- The formation of strong teams takes time and dedication on the part of everyone involved. The initial written white paper and proposal process were effective in forming good proposals (good ideas), but did not necessarily establish the trust required for teams to readily work together; starting with a core of people who have previously worked together may yield more cohesive and viable teams.

- Persistent effort and complete buy-in for all participants is needed to maintain team dedication. Campaign needs may not get the necessary attention, especially when conflicting priorities arise. Unequal dedication to campaign activities can severely damage campaign team morale and limit potential impact; frequent interactions and clear direction are required throughout the duration of a research campaign.

- A team with no funds will normally dissolve. Part of EMSL's function is to provide some "glue" funding to attract and initiate teamwork, but it can't be viewed as a funding agency or provider. One success indicator of a good team is the use leveraged funds to produce important research results and to attract newfound funds or opportunities.

- Because team members may bring data associated with a variety of projects to a campaign, ownership and appropriate handling of campaign-related data may have subtle complexities. It is necessary to clearly communicate data access and use policies and expectations at the beginning of the campaigns.

Team Leadership and Accountability

- Campaigns need a combination of senior visionary leadership and operational leadership. These may be available in the same person, but in some cases the senior visionary leaders do not have the time or skills for operational leadership. It is necessary to identify the organizational requirements and structure the campaign to establish the range of needed leadership to ensure progress.

- EMSL needs a single internal accountable contact to track progress and deal with information, communication, and reporting needs. This person may not necessarily be the project PI, but needs to have the full confidence and support of the PI.

- Although campaigns start with specific objectives, goals naturally change as new scientific directions become evident and challenges arise. However, if targets and objectives require modification, these changes need to be communicated to the scientific team and EMSL management, and fully accepted by both groups.

- When changing demands and priorities of critical campaign participants have the potential to affect progress, this needs to be communicated to EMSL management and appropriate actions initiated.

- EMSL initially attempted to monitor campaign progress on a semiannual progress reporting schedule. Because of the difficulty of keeping a team functioning, team progress and health needs to be assessed at least quarterly.

Instrument Access/Campaign Objectives/EMSL Operation

- EMSL needs to establish a clear process for dealing with conflicts and disputes that may arise during a campaign (e.g., scientific direction, methodology, data access, etc.). It is the responsibility of the EMSL Lead to either mediate disputes or raise them to the level of EMSL leadership.

- Equipment and capability access needed for the success of a campaign may vary with time and are not always easy to anticipate. EMSL needs to establish a process for dealing with access priority questions for campaigns and mitigate issues related to highly subscribed instruments. This can be at least partially handled by including campaigns in the annual resources allocation process, which occurs annually in early summer.

2.0   NMR/Modeling/Catalysis Campaign Report

Research Campaign: An Integrated Multi-Capability Approach for Enhancing Analysis of Catalytic Systems

Team:
External Collaborators: Enrique Iglesia (UC Berkeley), James Kubicki (Penn State University), Ja Hun Kwak (PNNL), and Chuck Peden (PNNL)

EMSL-Supported Team Members: Amity Anderson, Niri Govind, David Hoyt, Jian-Zhi Hu, Jinfeng Lai, Hardeep Mehta, Karl Mueller, Jesse Sears, and Nancy Washton

2.1   Objectives and Approach

In this project, advanced experimental and computational capabilities within EMSL at the PNNL were applied, implemented, and integrated for in-depth, molecular-level studies of catalytic processes that occur on oxide materials under controlled conditions of temperature, pressure, and input gas composition (*in situ* or near *in situ* measurements). Although specific problems in the broad field of catalysis spur novel questions, an integrated and multi-capability infrastructure for studying catalytic systems provides the greatest efficiency in the study of a wide range of problems. Within this research campaign, unique magnetic resonance capabilities (including both *in-situ* and advanced *ex-situ* ultra-high field techniques) were developed and deployed to determine the reaction sites on amorphous aluminosilicates and to determine the sites and coordination of copper within a zeolite. High-level computational analyses to both understand the data and predict the structure and dynamics of catalytic materials and of molecules involved in catalytic reactions on surfaces were conducted in conjunction with the experimental work.

2.2   Overview of Results

The results of this campaign include the development and extension of NMR and EPR capabilities in EMSL and the application of NWChem to interpret the new data types that can be collected. The combination of NMR and EPR expands the range of elements and chemical states to be examined during a catalytic process. The developments and targets are identified here and are described in more detail in appendices to this report.

- Cross-polarization Carr-Purcell Meiboom-Gill (CP-CPMG) MAS NMR was implemented to enable enhanced signal sensitivity. This capability was applied to amorphous silica-alumina catalyst systems using a trifluorosilane (TFS) probe molecule to obtain a quantitative description of silicon and aluminum atomic ordering on the surface of these materials. The measurements were necessarily complemented by the use of NWChem to identify and predict the parameters (e.g., chemical shielding tensors) of the reaction sites on these complex materials.

- New *in situ* CF MAS NMR probes were developed where pressure, gas flow, and temperature could be well controlled. Understanding the details of chemistry of polyoxometalate (POM) catalysts were the target here, and a new probe was developed that allowed measurements up to 300°C. The probe is currently being fine-tuned for use in EMSL user projects.

- *In-situ* EPR spectroscopy approaches that enabled interrogation of samples under controlled atmospheres and temperatures were also developed and applied to study a copper-zeolite catalyst. The experimental EPR developments necessitated advanced *ab initio* computational chemistry analyses for data interpretation. Hyperfine coupling constants (A-tensor) and g-tensors were calculated using new methods implemented within NWChem as a result of this campaign.
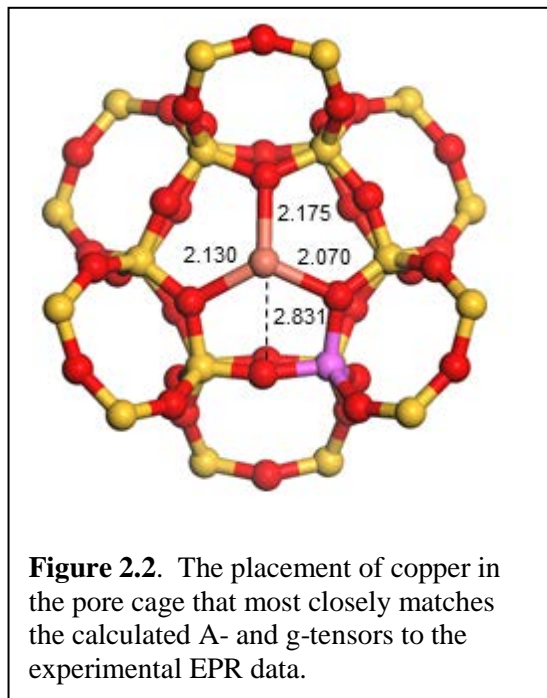
2.3 Results and Impacts

Reactive Site Density on Aluminosilicate Surfaces − The campaign enabled novel studies of reactive site density on the surfaces of aluminosilicate catalyst materials. Bringing together the ultrahigh magnetic field capabilities at EMSL (using the 20 T wide-bore NMR system) with computational chemistry efforts using NWChem, the research team demonstrated the ability to determine the local structures around reactive sites on aluminosilicate surfaces using the TFS probe molecule (Figure 2.1). The results were interpreted using a three-site model for local structure, based on the statistics of expected sites merged with predicted isotropic chemical shifts provided by NWChem. This work took a concerted effort by NMR scientists and computational chemists to bring advanced tools to bear on this problem. The sensitivity and resolution required for the NMR measurements could only be achieved at these high magnetic field strengths and using the advanced CP-CPMG pulse sequences implemented in the NMR lab. The interpretation of the experimental data needed guidance from the chemical shift calculations provided by NWChem. Ultimately, the difference between the statistically predicted distribution of sites and the distribution found through the three-site model can be interpreted in one of two ways. If the reactive sites are distributed randomly on the surface, then

Figure 2.1. Optimized cluster model derived from surface species of hydrated aluminosilicate surface model for TFS probe molecule bound to a Q3-Si surface site. The components are Si atoms (yellow), Al (magenta), H (white), O (red), C (gray), and F (blue). The data are analyzed using frequency shifts (δ) calculated using functionality in NWChem.

the surface distribution of local structures is different from that predicted by the statistical model. However, the reactivity of surface sites for the TFS probe molecule also could be different based on the local structure, meaning that some sites are more reactive based on the number of aluminum next nearest neighbors to the surface silicon. In either case, an argument can now be made for studying the reactivities of these sites individually in follow-on experimentation and modeling. These results are included in a paper to be submitted to the *Journal of the American Chemical Society* by the end of summer 2013.
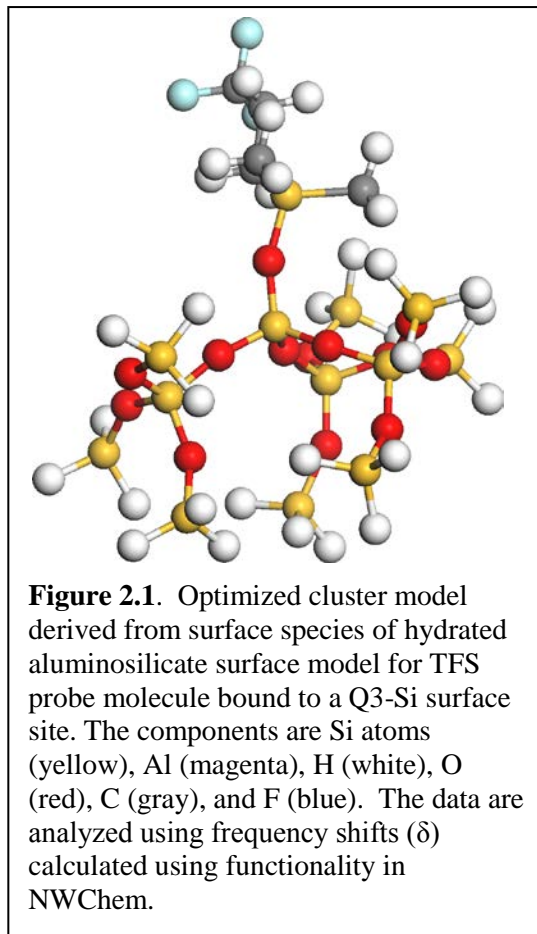
EPR, NWChem, and Copper Siting in Zeolite SSZ-13 − Coupling *ab initio* molecular modeling with experimental EPR was also made possible by the campaign. Although previous releases of NWChem contained algorithms for calculating A-tensors (hyperfine couplings), algorithms for determining g-tensors had not been base-lined and verified against standard compounds. Both the g- and A-tensor calculations use newly developed algorithms that had been tested against benchmark molecules with both light and heavy atoms. However, large copper complexes had not been included in this benchmark set (Figure 2.2). The Cu-SSZ-13 research not only provided important information about the location and ligand coordination of the copper, but it also provided a much-needed test case to compare NWChem output to both experimentally obtained g-tensor values as well as other calculated values available in the literature. Based on these findings, NWChem provides a robust method for calculating EPR parameters. These modules, including paramagnetic NMR, in NWChem's molecular density functional theory (DFT) code were included in the May 2013 release of NWChem 6.3.

Figure 2.2. The placement of copper in the pore cage that most closely matches the calculated A- and g-tensors to the experimental EPR data.

New and Enhanced Capabilities − Both NMR and EPR capabilities were enhanced via development of hardware for *in situ* studies. A CF NMR probe (Figure 2.3) and a CF EPR cell, both capable of temperatures in the range of 25 – 250°C, were developed, tested, and deployed. As part of the NMR system, a new portable cart was developed that allows precise control of the flow of up to four gases into the NMR probe. The EPR cell can also be exposed to a controlled atmosphere, such as $N_2(g)$, $He(g)$, $H_2(g)$, $H_2O(g)$, and $O_2(g)$. This functionality allows control of redox chemistry during heating or cooling cycles.

Figure 2.3. Continuous-flow MAS NMR stator as designed and implemented for 850-MHz NMR.
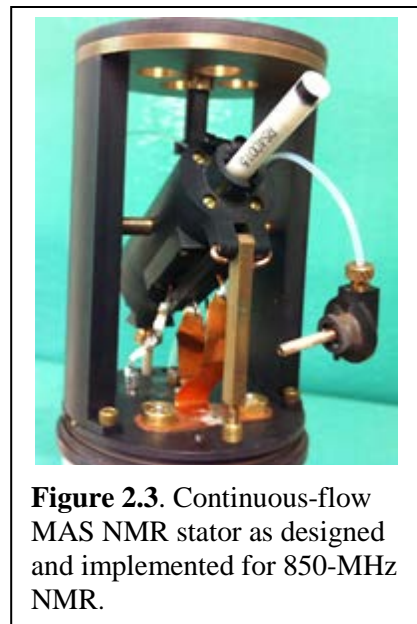
The CP-CPMG pulse sequence development that enabled mapping surface reactivity of a series of amorphous aluminosilicate (AAS) samples is now available to the EMSL user community. This pulse sequence has a wide variety of applications for inorganic and organic materials/systems such as surface-modified oxides, organic matter in soils, carbon-based polymer materials, and samples containing low levels of nuclei of interest. Other EMSL users, particularly in the field of the corrosion of nuclear waste glasses, have used these sequences and are publishing results in the literature. Additional users have transferred these sequences to other EMSL instrumentation (i.e., the 900-MHz solid-state NMR system) and have acquired ^{1}H/^{29}Si CP-CPMG results from model clay samples weathered with a simulated tank waste leachate.

2.4   Summary of Results/Impacts

Capabilities developed or implemented:

- CP-CPMG pulse sequences implemented and applied

- New elevated temperature flow probe established for NMR

- New elevated temperature flow cell developed for EPR

- New codes implemented in NWChem and validated for EPR tensor calculations.

Publications and manuscripts enabled by campaign:

- Andersen, A., E. D. Walter, F. Gao, N. Washton, N. Govind, J. Autschbach, C.H.F. Peden, "Combined electron paramagnetic resonance (EPR) and density functional theory (DFT) characterization of Cu(II)-SSZ-13 for selective catalytic reduction (SCR)," in preparation (2013).

- N. M. Washton, A. Andersen, K. T. Mueller, "Silanization binding sites on aluminosilicate gels determined by solid-state nuclear magnetic resonance and *ab initio* studies," in preparation (2013).

- Mei D, VM Lebarbier, RJ Rousseau, VA Glezakou, KO Albrecht, L Kovarik, MD Flake, and RA Dagle. 2013. "Comparative investigation of benzene steam reforming over spinel supported Rh and Ir catalysts." *ACSCatalysis* 3(6):1133−1143

- Mei D. 2013. "First principles characterization of formate and carboxyl adsorption on the stoichiometric CeO2(111) and CeO2(110) surfaces." *Journal of Energy Chemistry* 22(3):524-532.User proposals enabled by new capabilities:

- Multiple user proposals requesting the use of the CF NMR probe (on the 850 MHz solid-state NMR instrument) and/or CF EPR capability have been submitted as part of the EMSL 2013 call for proposals.

2.5   Follow-On Work

Results from the AAS system are being prepared for publication now that final analyses of site distributions have been accomplished. Further studies of reactive surface area using the TFS probe molecules are ongoing as part of the EMSL user program, focusing on reactions on the surfaces of inorganic systems in the environment and on the corrosion of nuclear waste glasses.

An improved CF NMR probe is under development that will reach 600°C, thereby allowing sample interrogation at temperatures relevant to a wider range of catalytic reactions. The design and development phases are finished as is the majority of the hardware fabrication; the anticipated completion date is September 2013.

Enhanced flow and high-temperature EPR capabilities have been identified as important and await funding opportunities to implement.

3.0   Integrating Biological Data Campaign Report

Research Campaign: Integration of Multi-Omics Technologies to Understand Complex Biology

Team:
External Collaborators: Birgitte Ahring (WSU PI), Junfeng Xue (WSU), Becky Hess (WSU/PNNL), Ron Taylor (PNNL), Lye Meng Markillie (PNNL)

EMSL-Supported Team Members: Bryan Linggi, Galya Orr, Steven Wiley, Ljiljana Pasa-Tolic, Si Wu, Josh Aldrich, Nancy Isern

3.1   Objectives and Approach

This campaign sought an enhanced understanding of the regulation, enzymology, and metabolism associated with the pathways involved in isoprene production by *Bacillus subtilis* by using a "multi-omic" approach. Our central hypotheses for testing were: 1) the formation of isoprene is tightly controlled by unknown regulators, and 2) over-expressing IspS will result in a significant increase in the production of isoprene in *B. subtilis* (future work). The long-term goal of the research is to substantially increase the yield of isoprene and other useful bioproducts by using genetically modified *B. subtilis*. The major aim of this proposal was to use EMSL capabilities to identify unknown activators and repressors of the DXP pathway (Figure 3.1) in *B. subtilis* strain DSM10 by correlating changes in the transcriptomic, proteomic, and metabolomic profiles between the wild-type strain, mutant strains, and environmentally perturbed wild-type samples with changes in isoprene production. The genes or gene products that show significant correlations will then be examined by more targeted approaches to identify which ones are mechanistically linked. Any identified regulators can then be used as targets for future genetic modification.

3.2   Overview of Results

The results of this research campaign include the development of new methods for the analysis of cell metabolites as well as methods to derive biological insight from complex data sets. These methods will continue to be developed as the amount and quality of data gathered from EMSL instruments continues to increase. Deriving biological insight from these large data sets is a challenging yet powerful opportunity that EMSL is uniquely positioned to undertake by expanding the foundations laid by this campaign.

- Our transcriptomic analysis showed that a subset of 213 regulated genes was sufficient to create a predictive model of isoprene production under different conditions and showed correlations at the transcriptional level.
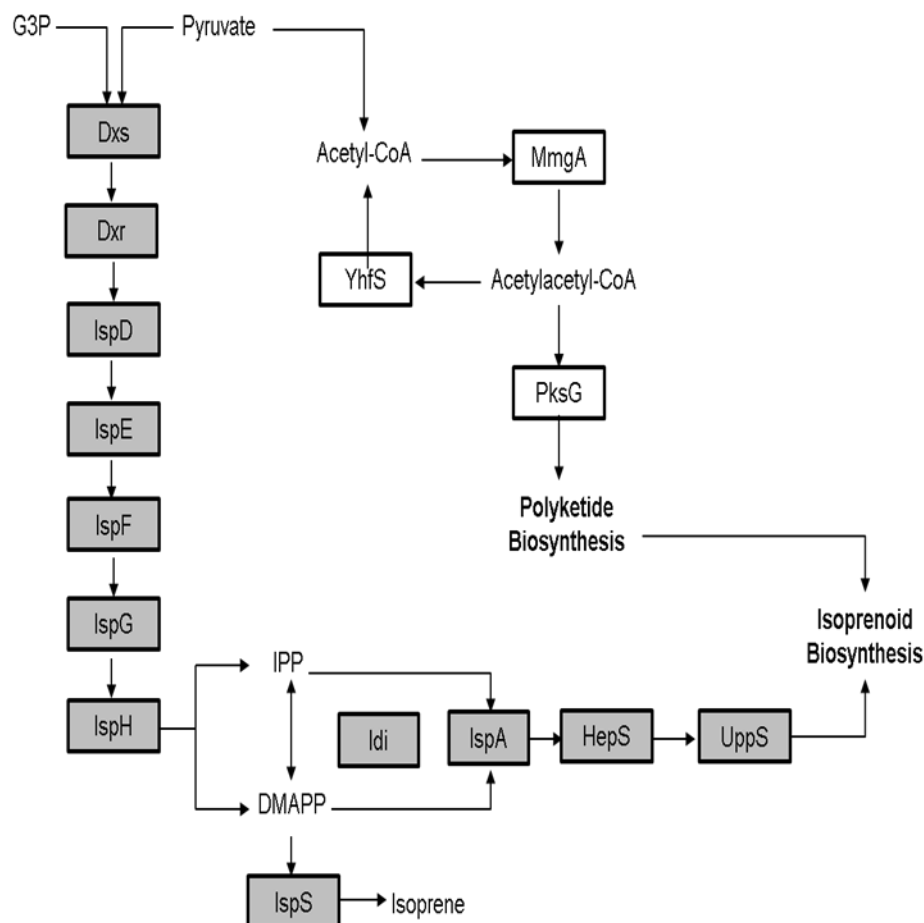
Figure 3.1. Enzymes (encoded by genes referred to in text as "terpenoid genes") examined in this study that are involved in isoprenoid biosynthesis in *Bacillus subtilis*. The 1-deoxy-D-xylulose-5-phosphate (DXP) pathway enzymes (shaded) are: Dxs, 1-Deoxy-D-xylulose-5-phosphate synthase; Dxr, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; IspD, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; IspE, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; IspF, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; IspG, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase; IspH, 1-hydroxy-2-methyl-butenyl 4-diphosphate reductase; Idi (encoded by fni), isopentenyl pyrophosphate isomerase; IspA, farnesyl diphosphate synthase; HepS, heptaprenyl diphosphate synthase component I; UppS, undecaprenyl pyrophosphate synthetase; IspS, putative isoprene synthase. The mevalonate (MVA) pathway enzymes (white) are: MmgA, degradative acetoacetyl-CoA thiolase; YhfS, hydroxymethylglutaryl CoA synthase; PksG, 3-hydroxy-3-methylglutaryl-ACP synthase. Metabolite abbreviations: G3P, glyceraldehyde-3-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate.

- Initial global proteomics profiling did not capture all of the proteins of interest in the investigated samples; further protocol development using iTRAQ analysis was necessary to identify and estimate the abundance of the major metabolic pathway proteins. These data have recently been produced, and we are in the process of analyzing the data sets. Because the protein abundance of all pathway members has been measured, the collective behavior of the pathway members can be analyzed to determine the protein regulation mechanisms that affect isoprene production.

- Our initial investigations of metabolite production by *B. subtilis* revealed that many of the metabolites of interest were at levels that are not detectable by NMR. To detect these less abundant metabolites, we developed a mass spectrometry (MS) approach for detecting isoprene precursors. However, because of the inability to obtain standards for many of the metabolites of interest, which are required for MS methods, we were limited in the number of metabolites that were measurable. This demonstrated the relative strengths and weaknesses of the methods available at EMSL for metabolite detection.

3.3   Results and Impacts

3.3.1   Transcriptomics

Although the regulation of several pathway genes has been described, there is a paucity of information regarding system-level regulation and control of the pathway. To address these limitations, we examined *B. subtilis* grown under multiple conditions and determined the relationship between altered isoprene production and gene expression patterns (Figure 3.2). This global analysis identified several classes of genes that were coregulated following different perturbations (Figure 3.2, clusters 1, 2, and 3). Surprisingly, perturbations that produced high and low amounts of isoprene were similar at the level of gene expression changes (Figure 3.2 cluster A). Additionally, we found that with respect to the amount of isoprene produced, terpenoid genes fall into two distinct subsets with opposing correlations. 1) The group whose expression levels positively correlated with isoprene production included dxs, which is responsible for the commitment step in the pathway, ispD, and two genes that participate in the mevalonate pathway, yhfS and pksG. 2) The subset of terpenoid genes that inversely correlated with isoprene production included ispH, ispF, hepS, uppS, ispE, and dxr. A genome-wide partial least squares regression (PLSR) model was created to identify other genes or pathways that contribute to isoprene production. These analyses showed that a subset of 213 regulated genes was sufficient to create a predictive model of isoprene production under different conditions and showed correlations at the transcriptional level. We conclude that gene expression levels alone are sufficiently informative about the metabolic state of a cell that produces increased isoprene and can be used to build a model that accurately predicts production of this secondary metabolite across many simulated environmental conditions.
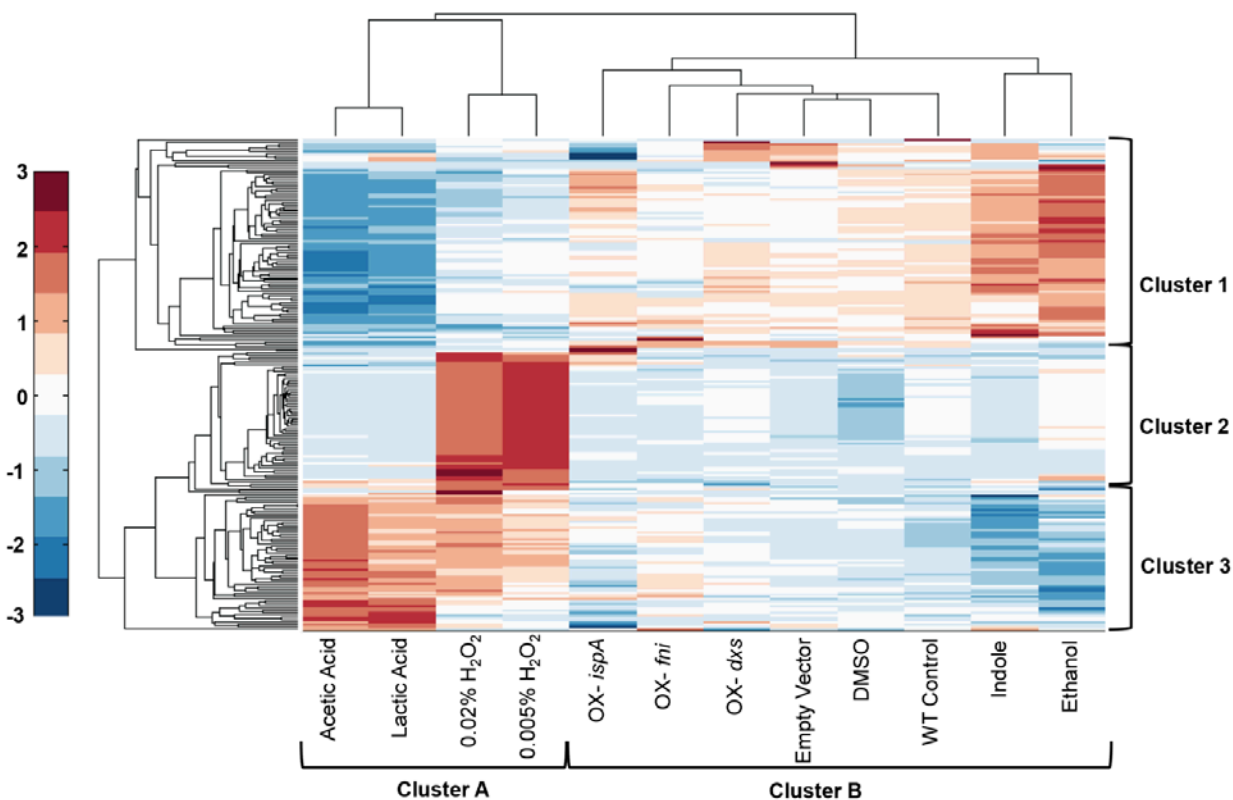
Figure 3.2. . **The terpenoid genes cluster into three main groups with respect to coexpression and into two main groups with respect to perturbations Two-dimensional hierarchical clustering of the terpenoid genes against isoprene production in the investigated environmental and genetic perturbations in *B. subtilis***

3.3.2 Proteomics

Quantitative protein measurements can provide insight into the post-translational regulatory mechanisms, such as regulation of translational efficiency or protein turnover rates. Protein abundance was measured using Accurate Mass and Time (AMT)-based label-free quantitation with tandem mass spectrometers (MS/MS) equipped with a linear ion trap quadrapole mass filter (LTQ Orbitrap) (Figure 3.3). Initial global proteomics profiling did not capture all of the proteins of interest in the investigated samples; further protocol development using iTRAQ analysis was necessary to capture the abundance of these proteins. These data have recently been produced, and we are in the process of analyzing the data sets. Because the protein abundance of all pathway members has been measured, their collective behavior can be analyzed to determine the protein regulation mechanisms that are in place and that affect isoprene production.

Initial hierarchical clustering analysis of the proteome data has revealed a pattern of protein inhibition in the *IspA* overexpression condition (Figure 3.4). The relationship of these proteins to the DXP pathway is currently being investigated to determine the significance of this initial observation.
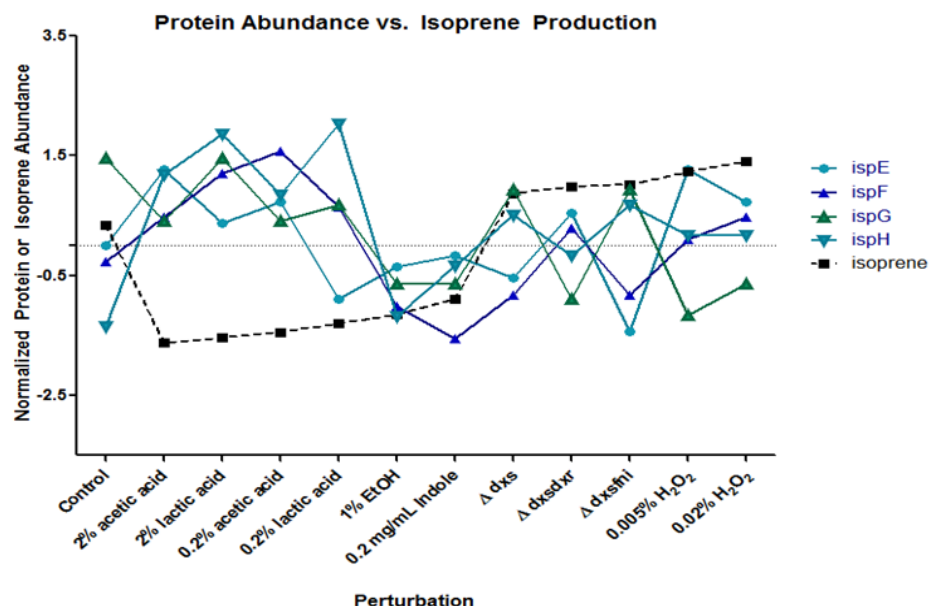
Figure 3.3. Normalized protein abundance compared to normalized isoprene production (z-scores). The isoprene concentrations in the headspace were measured by gas chromatography–mass spectrometry (GC-MS). Perturbation descriptions which that begin with "Δ" are representative of genetic perturbations in which the listed gene or genes were overexpressed; EtOH, ethanol**.**

Figure 3.4. Heat map and clustering of protein expression changes in each condition. Red indicates increase in expression and green indicates decrease in expression. Expression differences are

calculated for each condition compared to the control and are significant with p < .05. Colored clusters on the left side indicate groups that are highly abundant in the OX-IspA condition (red) and low (blue) compared to the OX-DXS.
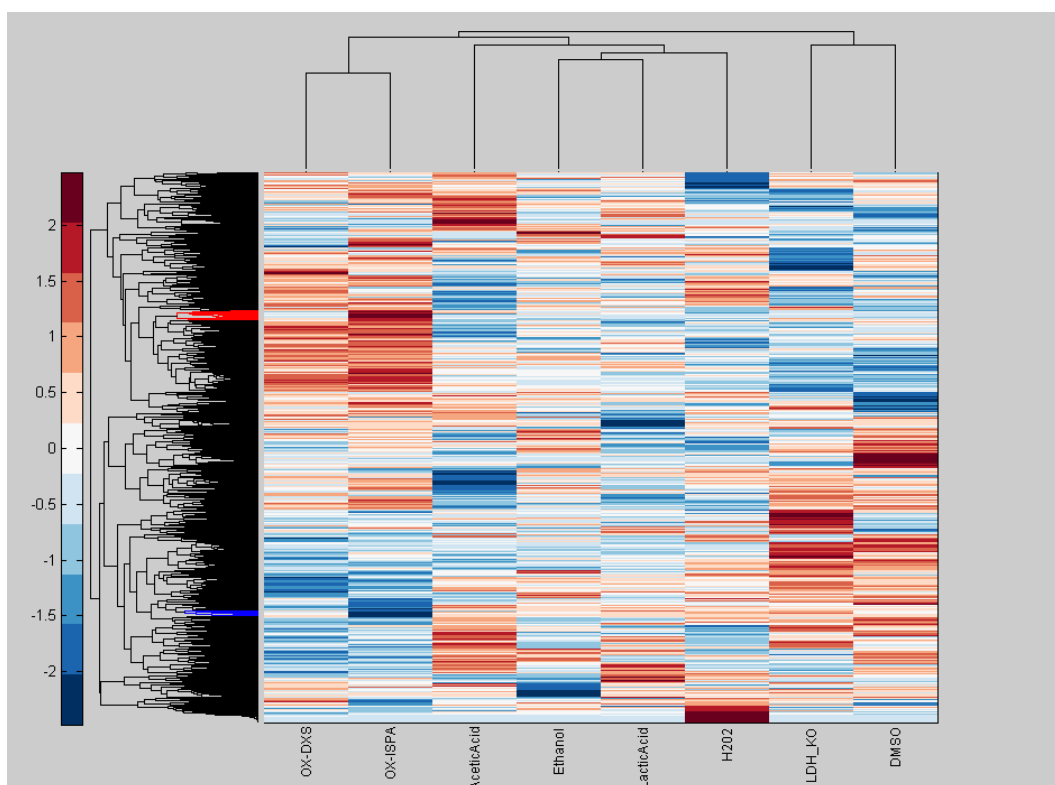
3.3.3   Metabolomics

Preliminary studies with one-dimensional (1-D) NMR were used to determine the concentrations of primary carbon metabolism in the different conditions tested (Figure 3.5). From these data, we were able to determine that there is a correlation between lactate production and isoprene production. Correlating lactate, a known overflow metabolite, with isoprene production lends further support to the hypothesis that the role of isoprene in the cell is to serve as a balance between carbon catabolism and anabolism (Shirk et al., 2002).

Bacillus subtilis produces higher concentrations of isoprene in comparison to other prokaryotic organisms tested. This suggests that this organism has a high metabolic flux into the DXP pathway and is tolerant of toxic intermediate metabolites, such as dimethylallyl diphosphate (DMAPP) (Figure 3.1). To investigate these possibilities, we needed to develop a method that was sufficiently sensitive in its limit of detection to quantitatively measure the level of metabolites in the pathway. Preliminary studies with 1-D NMR showed that this technique lacked the sensitivity to detect DMAPP. Thus, we developed an approach based on HPLC-MS-MS. Using this method, we detected significant levels of DMAPP, geranyl pyrophosphate (GPP), and farnesyl diphosphate (FPP) in cell lysates (Figure 3.6). The ability to identify and measure these metabolites will be invaluable for building metabolic models and for assigning appropriate levels of flux to individual reactions.
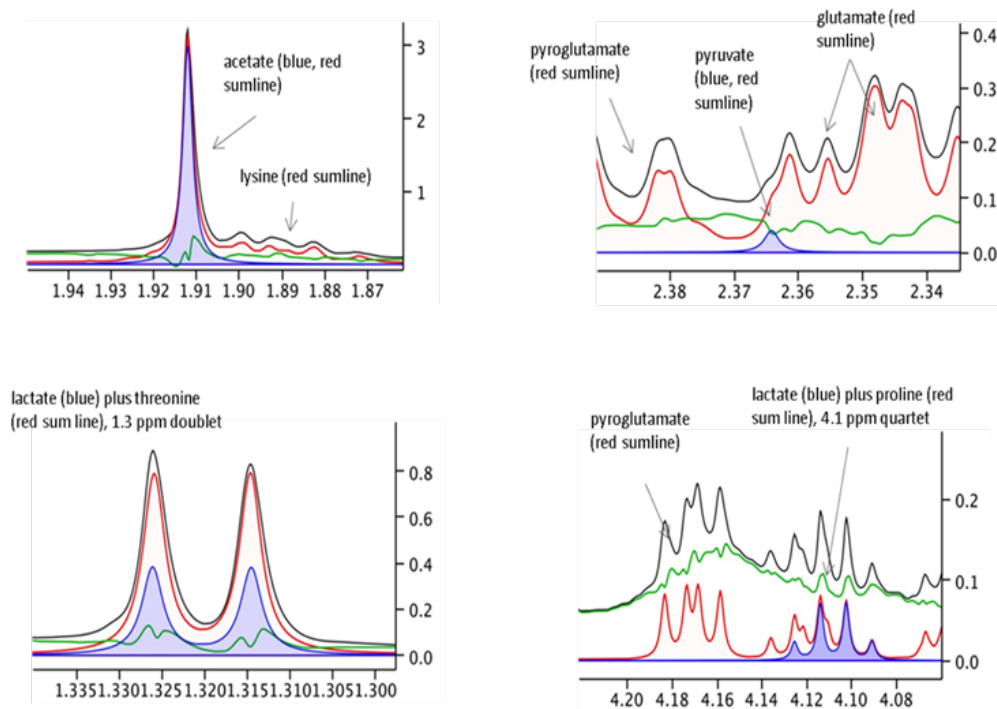
Figure 3.5. NMR spectra of lactate and acetate, and pyruvate in B. subtilis DSM10 grown in **luria broth** (LB) medium at 30°C.
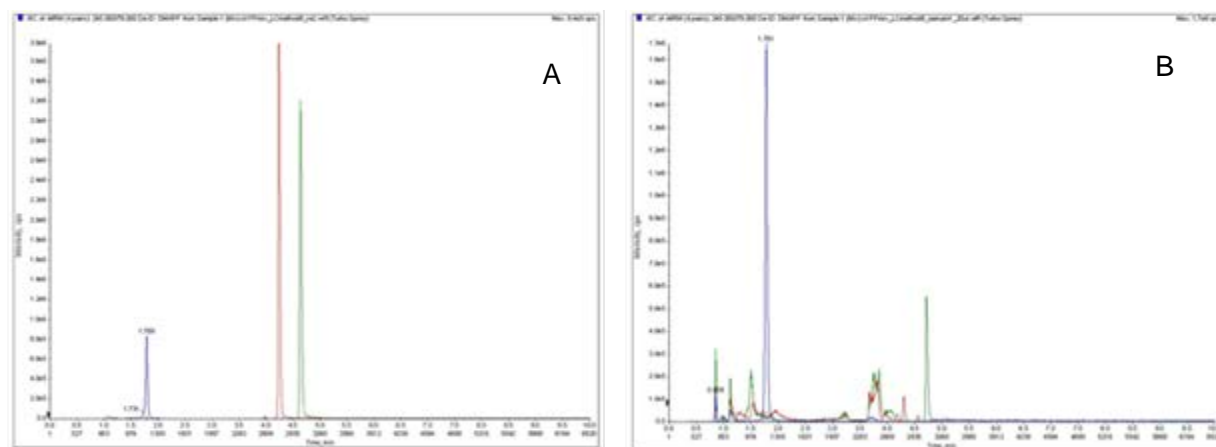

Figure 3.6. HPLC-MS-MS measurements of intermediate metabolites in the DXP pathway. Blue, dimethylallyl diphosphate; red, geranyl diphosphate; green, farnesyl diphosphate. Panel A, standards for each metabolite measured; Panel B, metabolites measured in bacterial extracts.

3.3.4   Data Integration and Metabolic Modeling

We have integrated the metabolomics data into the existing, highly curated metabolic model for *B. subtilis* (iBsu1103V2) and conducted growth simulations that match experimental conditions. The simulation results are consistent with experimental results, indicating that we have constrained the solution space for the metabolic model relative to our specific strain and specific experimental conditions. We have also conducted *in silico* experiments to determine appropriate gene targets to modify the isoprenoid pathway and are currently constructing mutants to determine the accuracy of these predictions.

In an effort to further improve the accuracy of the metabolic model, work is underway to construct a genome-scale transcription regulatory network model for *B. subtilis* DSM10. This will be integrated into the metabolic model using the probabilistic regulation of metabolism (PROM) algorithm within the DOE Knowledge Base (Kbase) Environment. This work is part of the beta testing phase of the modeling tools that have been built as part of the Kbase. Our integrated model will be used to predict genetic perturbations that will enhance isoprene production while maintaining robust biomass production and will serve as a test bed for additional Kbase modeling tool development.

3.4   Summary of Results/Impacts

Integration of a large number of diverse omics datasets had not been previously attempted in EMSL. The campaign highlighted both advantages and challenges (both technical and logistical) of generating and meaningfully integrating such data. The campaign produced biological insights and helped develop capabilities that would not have been available without it. The development of methods to integrate the multi-omics datasets have allowed the identification of several previously unknown pathways that correlate with isoprene production. Furthermore, the global analysis has highlighted the overall cellular

changes that occur in response to local perturbations and some of the physiological responses that must be considered during metabolic engineering approaches.

Capabilities developed or implemented:

- HPLC-MS/MS to detect DXP pathway intermediates

- Detection of low abundance proteins using iTRAQ proteomic approaches.

- Metabolic modeling and implementation of Kbase modeling tools.

Publications and manuscripts enabled by the campaign:

- Coregulation of terpenoid pathway genes and the predictive capacity of transcriptomics profiles related to isoprene production in *B. subtilis*. Hess, B. M., Xue, J., Markillie, L. M., Taylor, R. C., Wiley, H. S., Ahring, B. K., & Linggi, B. (2013). *PLoS ONE*. doi:10.1371/journal.pone.0066104

- Characterization of metabolic profiles using ^{1}H NMR spectroscopy and HPLC-MS/MS facilitates the strain improvement to overproduce isoprene in *Bacillus subtilis* (in preparation). Xue, J, Isern, N, Hess, B. M, Linggi, B., Wiley, H. S., & Ahring, B. K.

- Overexpression of the geranyltransferase IspA in *Bacillus subtilis* decreases genetic competence. (in preparation). Becky M. Hess, Junfeng Xue, Josh Aldrich, Si Wu, Lilijana Pasa-Tolic, Birgitte Ahring, H. Steven Wiley, & Bryan Linggi

- Regulation of sporulation pathways in *Bacillus subtilis* overexpressing IspA (in preparation). Becky M. Hess, Junfeng Xue, Josh Aldrich, Si Wu, Lilijana Pasa-Tolic, Birgitte Ahring, H. Steven Wiley, & Bryan Linggi.

3.5   Follow-On Work

Follow-on work includes the following:

- EMSL User Proposal, Christopher Henry, and Birgitte Ahring

- Engineering "plug and play" synthetic modules for microbial expression of high value products

- Transcriptomics modeling to enable rational target selection for genetic perturbations

- Collaboration with Argonne National Laboratory (Christopher Henry) to enhance the capabilities of the DOE Knowledge Base metabolic modeling tools

- Integrating multiple data streams into a computational framework to simulate phenotype changes from in silico perturbations

- Dynamic modeling of transcriptional regulation.

Appendix A

Call for Campaign Proposals – First Generation Campaigns

Appendix A

Call for Campaign Proposals – First Generation Campaigns

http://www.emsl.pnl.gov/access/research_call/

RESEARCH CAMPAIGNS – AN OPPORTUNITY FOR HIGHLY INNOVATIVE MULTIDISCIPLINARY SCIENCE

DATA INTEGRATION IN BIOLOGY & INTERFACIAL SCIENCE

EMSL is a national user facility that provides integrated experimental and computational resources for discovery and technological innovation. Its primary research focus is environmental molecular sciences to support the needs of the Department of Energy, but it also supports research in the biological, chemical, physical, and computational sciences relevant to other governmental agencies and the nation.

Although much of the user-proposed research conducted in EMSL uses multiple instruments and technical capabilities, we believe that there are significant opportunities to truly advance science and speed scientific progress by appropriately designed experimental and computational research efforts and advanced data integration.

The purpose of this call is to request concept papers on how to exploit the multi-technologies and data integration capability of EMSL to significantly advance our understanding of a specific scientific problem in either biology or the interfacial sciences.

BIOLOGY

In the biological sciences, EMSL has world-class capabilities in quantitative proteomics, meta-proteomics, and studies of post-translational modifications (e.g., phosphorylation, methylation, etc.). EMSL also has notable biologically oriented strengths in electron and light microscopy, NMR spectroscopy, and high performance computing. Recently, EMSL has added capabilities in deep sequencing-based transcriptional profiling and metabolomics. *We are particularly interested in* concept papers in which proteomics data is integrated with transcriptomics data to answer fundamental questions regarding the dynamics of cellular networks and how the flow of energy, matter, and information is regulated.

INTEGRATED INTERFACIAL SCIENCE

EMSL has advanced capabilities in *in situ* spectroscopy and imaging at interfaces including non-linear optical techniques, NMR, scanning probe microscopy, x-ray photoelectron spectroscopy, and transmission electron microscopy. Recent high performance computational advances within EMSL's NWChem computational chemistry software are focused on molecular interfaces specifically predicting the short-range molecular interaction of ions in solution, the structure and dynamics of molecules on surfaces, and the excited states of materials and surface adsorbates. *We are particularly interested in* concept papers that link computational capabilities with dynamic measurements at molecular interfaces to address important questions related to catalysis, atmospheric chemistry, geochemistry, biogeochemistry, emission sequestration, and energy production, transformation and storage.

SUCCESSFUL PROPOSALS

Successful proposals will pose scientific questions that cannot be readily addressed without access to integrated research capabilities. EMSL will provide not only access to research instrumentation and sample preparation capabilities, but also to the computational and informatics resources and expertise needed to integrate and understand the resulting data. Other capabilities within EMSL can also be requested to be included as part of a Research Campaign. Ideas involving EMSL capabilities and those of the Spallation Neutron Source at Oak Ridge National Laboratory are also welcome. Financial support for investigators is not included as part of this research call.

The Proposal Process: Interested investigators should submit a two-page concept paper outlining an important scientific question and how data from the specified integrated EMSL capabilities can address that particular question. EMSL staff will work with selected applicants to expand the initial concepts into five-page proposals. These expanded proposals will address the specific resources to be requested for the research campaigns, which can be up to two years in duration. The expanded proposals will be evaluated using a peer review process.

The two-page concept papers should be submitted to Don Baer by June 1, 2010.

Overall Schedule:

Submission date for concept papers: June 1

Successful applicants will be notified by: July 20

Expanded proposals will be due by: Aug. 27

Research Campaign start date: Oct. 1

Additional information:

Don Baer, EMSL Science Lead for Interfacial Chemistry (don.baer@pnl.gov)

Nancy Hess, EMSL Science Lead for Geochemistry (nancy.hess@pnl.gov)

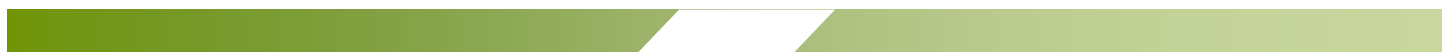Steven Wiley, EMSL Science Lead for Biology (steven.wiley@pnl.gov)

Bert DeJong, NWChem and High Performance Computing (wibe.dejong@pnl.gov)

Capability information can be found at http://www.emsl.pnl.gov/capabilities/.

Appendix B

Pore-Scale Modeling Challenge and Workshop

Appendix B

Pore-Scale Modeling Challenge and Workshop

http://www.emsl.pnnl.gov/news/conferences/pore_scale.jsp

EMSL

BOOKMARK US

PORE-SCALE MODELING CHALLENGE & WORKSHOP

Pore-scale Modeling Challenge & Workshop

EMSL at Pacific Northwest National Laboratory in Richland, Wash.

EMSL hosted a two-day pore-scale modeling workshop on Aug. 9-10, 2011. The group's challenge was to assess predictive models of multiphase flow at the pore scale. Validating models at the pore scale are necessary to predict contaminant and colloid transport, multiphase fluid displacement, mineralizing reactions, bioremediation, and deep geologic sequestration of carbon dioxide.

Modeling groups of interested theorists and experimentalists from around the U.S. came together to define and develop a set of pore-scale experimental studies to calibrate numerical models. Nine invited speakers and 40 participants engaged in presentations and discussions which culminated in a consensus by the community on a series of experiments for flow and transport in micro models.
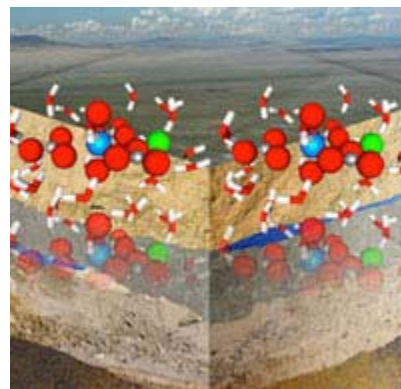
Invited speakers and participants gave oral presentations of both experimental and simulation work, scanning the landscape of current challenges. Two break-out groups were able to outline a set of experiments for both displacement and reactive mixing to be conducted at EMSL. The next steps include a full development of the experimental plan which will be circulated to participants for comment and input.

The workshop also featured invited speakers discussing the current state of knowledge and concepts for initial benchmarking experiments and provided a forum to refine pore-scale experiments and define rules for a benchmarking challenge. Topics on experimental systems included: diffusion in pore networks; multiphase systems, such as oil-water, water-air, mineral precipitation, and microbial growth; and colloidal transport.

Invited speakers who attended:

- Michael Celia, Princeton University
- Markus Hilpert, Johns Hopkins University
- Peter Lichtner, Los Alamos National Laboratory
- Laura Pyrak-Nolte, Purdue University
- Timothy Scheibe, PNNL
- Alexandre Tartakovsky, PNNL
- Charles Werth, University of Illinois
- Dorthe Wildenschild, Oregon State University

EMSL is a Department of Energy national scientific user facility located at PNNL.

For information regarding EMSL's Pore-scale Modeling Challenge/Workshop, contact: Nancy Hess nancy.hess@pnl.gov
Mart Oostrom mart.oostrom@pnl.gov

Appendix C

Detailed Description of Important Results from Research Campaign − An Integrated Multi-Capability Approach for Enhancing Analysis of Catalytic Systems

Appendix C

Detailed Description of Important Results from Research Campaign − An Integrated Multi-Capability Approach for Enhancing Analysis of Catalytic Systems

C.1 Novel Studies of Reaction Sites on Aluminosilicates

The development of new catalyst systems can build upon the molecular-level structure of known catalyst materials through a description of the reactive sites present on the surface of these materials. In pursuit of this goal, new surface-sensitive capabilities in solid-state nuclear magnetic resonance (NMR) were developed, deployed, and coupled with computational chemistry studies that predict measured NMR parameters. Cross-polarization Carr-Purcell-Meiboom-Gill (CP-CPMG) pulse sequences under magic-angle spinning (MAS) conditions were programmed and deployed on EMSL instrumentation, and they provided high-resolution NMR data with increased sensitivity that are amenable to modeling based on results from quantum chemical calculations. This method for investigating reactive hydroxyl species on the surface of oxide materials was used to probe local ordering of the Si and Al species on amorphous aluminosilicate (AAS) and zeolite catalyst surfaces. In parallel, computational chemistry using NWChem was used to identify and predict the spectral parameters of the reaction sites on these complex materials.

To determine the number of reactive sites and to study the local environment of surface reaction sites, a trifluorosilane probe molecule (TFS) is reacted quantitatively with complex oxide surfaces. This probe molecule is known to react selectively at lone hydroxyl species (predominantly Q3 silanol species) on the surfaces. The M-region (corresponding to the ^{29}Si from the TFS molecule itself that is covalently bound to the surfaces) of the ^{1}H-^{29}Si CP-CPMG spectra of TFS-treated silica gel, a set of AASs and a mesoporous γ-Al$_2$O$_3$ are shown in Figure C.1. Samples A−F are AASs that have increasing amounts of Al$_2$O$_3$, ranging from ca 5.5 to 50 weight percent. The center of mass of the spikelet pattern shifts toward lower frequencies as Al$_2$O$_3$ weight percent increases, from ca 12 ppm for silica gel to 7 ppm for mesoporous alumina. In addition, the spikelet pattern clearly shows the subtle

Figure C.1. M-region ^{1}H-^{29}Si CP-CPMG spectra of TFS treated silica gel, a series of AAS materials with increasing Al$_2$O$_3$ content and mesoporous alumina.

intensity changes of each spike as a function of Al$_2$O$_3$ weight percent. However, the spikes are not representative of separate isotropic peaks, but rather clarify small changes occurring in the distribution of
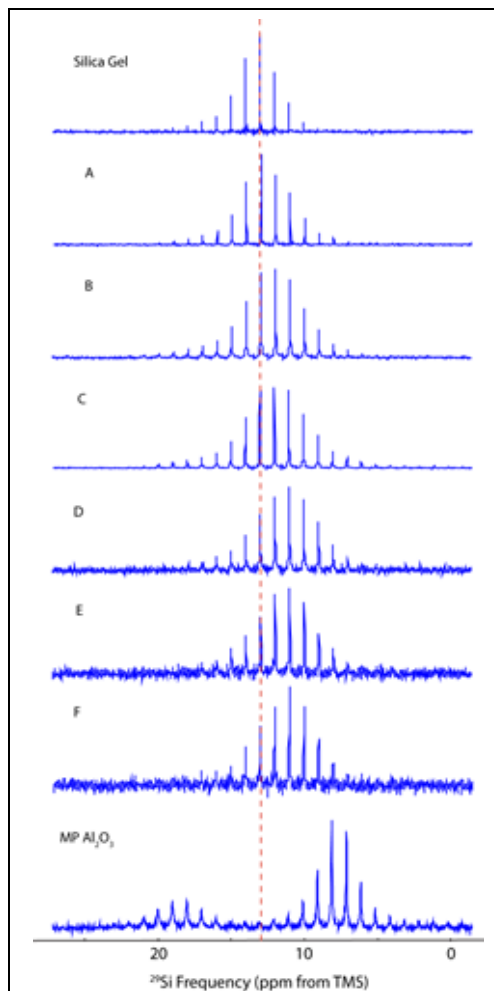
isotropic peaks within the single broad resonance that encompasses all of the chemical species in the M-region. *Ab initio* chemical shift tensor calculations were performed to map the surface species onto the spikelet pattern. Table C.1 contains the calculated isotropic chemical shift values (δiso) for the ^{29}Si-[TFS] covalently attached to model oxide clusters (Figure C.2). The model oxide clusters contain 0-3 Al neighbors to the primary Q3 species Si, or contain a primary Al^{IV} Q species.

The inability to assign discrete resonances within the spikelet pattern to specific Q-species necessitated modeling the data using a three-site model (Figure C.3) based on the probability of occurrence for each Q-species,

$$P_{n,m} = \frac{n!}{m!(n-m)!}\rho^m(1-\rho)^{(n-m)} \tag{C.1}$$

where *n* is the Q*n* species, m is the number of Al neighbors, and ρ is the Al^{IV}/Si ratio from bulk analysis. The most probable Q*n* sites were Q3-0Al, Q3-1Al, and Q3-2Al. The modeling was accomplished using inverse matrix analysis in the Mathematica™ software package. Results indicate that the percentage of the Q*n*-mAl within these systems differs from the values obtained from Equation C.1.

Table C.1. Isotropic ^{29}Si δ (in ppm) for TFS probe covalently bonded via an O atom to each type of surface species of AAS surfaces.

Surface Species	Basis Set[a]			
	6-31G**	6-311G**	6-311++G(2p,2d)	6-311++G(3pd,3df)
Q3-3Si	14.4	18.1	14.4	13.2
Q3-2Si1Al	20.3	18.0	13.6	12.2
Q3-1Si2Al	18.4	13.9	10.3	9.3
Q3-3Al	16.2	12.0	8.1	7.2
Q2-2Si	19.6	15.2	11.7	10.5
Q2-1Si1Al	17.7	13.9	10.0	9.0
Q2-2Al	17.2	11.9	8.5	7.1
Al$_{IV}$-3Si	11.1	9.0	5.6	4.8

(a) Reference σTMS = 412.1, 340.3, 328.4, 323.7 ppm for the GIAO B3LYP-DFT calculations with the 6-31G**, 6-311G**, 6-311++G(2p,2d), and 6-311++G(3pd,3df), respectively.

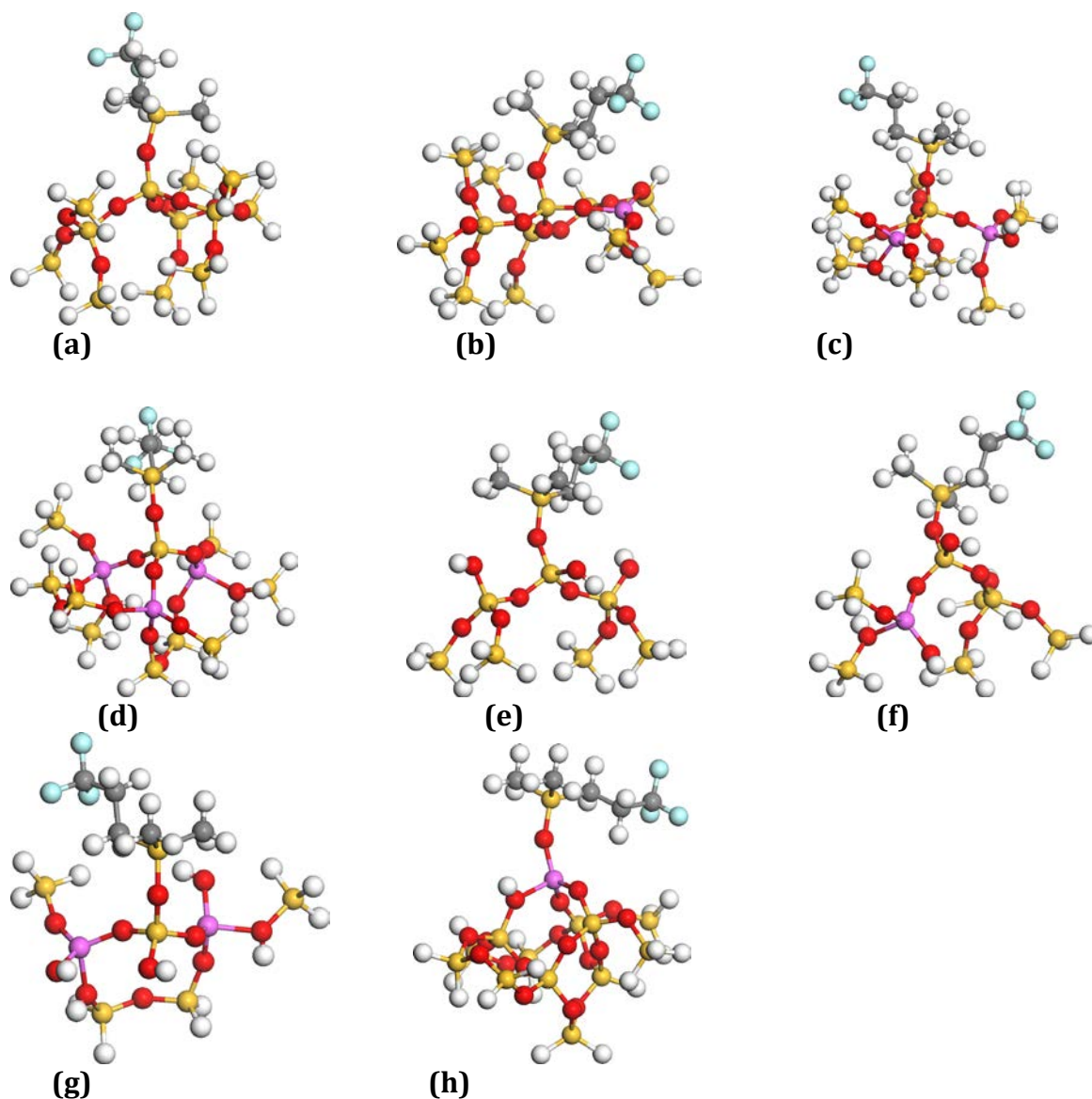**Figure C.2**. Representative optimized cluster models derived from surface species of hydrated amorphous aluminosilicate surface models with bonded TFS probe molecule. (a) Q3-3Si, (b) Q3-1Al, (c) Q3-2Al, (d) Q3-3Al, (e) Q2-2Si, (f) Q2-1Al, (g) Q2-2Al, (h) Al$_{IV}$-3Si. Yellow balls are Si atoms, magenta balls are Al atoms, white balls are H atoms, red balls are O atoms, gray balls are C atoms, and blue balls are F atoms.
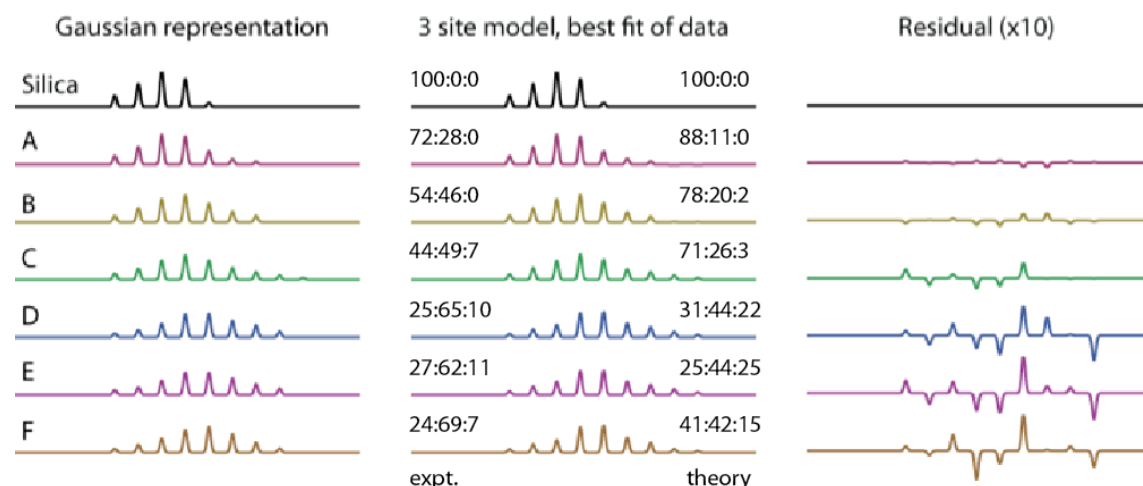
Figure C.3. Three-site model fitting based on experimental data (Gaussian representation) and best fit of the data to three discrete sites, Q3-0Al:Q3-1Al:Q3-2Al; numerical values represent the three-site best fit model (left) and the theoretical populations, based on measured Si/Al^{IV} ratios and Equation (C.1) (right).

C.2 In Situ NMR Using a Continuous-Flow MAS NMR Capability

Industry uses catalysts to more rapidly and efficiently carry out chemical reactions for energy production or storage, industrial pollution mitigation efforts, and vehicle emission management. Fundamental studies of a catalyst's molecular structure, and its impact on chemical reactivity, are critical for improving or designing new, more effective catalysts. The catalysis team involved in this campaign is studying the reaction mechanisms and details for chemistry carried out by a class of catalysts, called polyoxometalates or POMs, supported on the surfaces of metal oxides. POMs are a versatile catalyst that can be tuned to have different reaction properties as a function of chemical composition. The team is using an integrated approach combining experiments and computational modeling to unravel molecular-level details of important chemical reactions carried out using POMs.

The University of California (UC)-Berkeley and PNNL scientists are building these POMs, which have been studied in EMSL using NMR spectroscopy. POMs have been studied using other spectroscopic techniques, but NMR provides improved molecular information related to the POM structures as well as the molecular structures of catalytic intermediates. During this campaign, researchers developed a modified *in situ* continuous-flow (CF) MAS probe, based on earlier EMSL designs, to further studies of POMs and related materials. The new probe is shown in Figure C.4. Past concerns with the inhomogeneous penetration of the reactants into the catalyst bed were addressed by using a new Venturi tube inlet and a frit with multiple holes on the outlet side of the rotor. Through the development of this probe, a number of interesting catalyst systems may be studied, all of which are catalytically active at relatively moderate temperatures (250 – 300 K). However, to make the CF-MAS *in situ* experiment useful for a wide range of catalyst reaction systems, temperatures as high as 400°C will be required. Implementation of this extended feature, as well as the design of probes to be used at higher magnetic field strengths, makes these tools ready for use with a variety of catalytic systems under truly operational conditions.
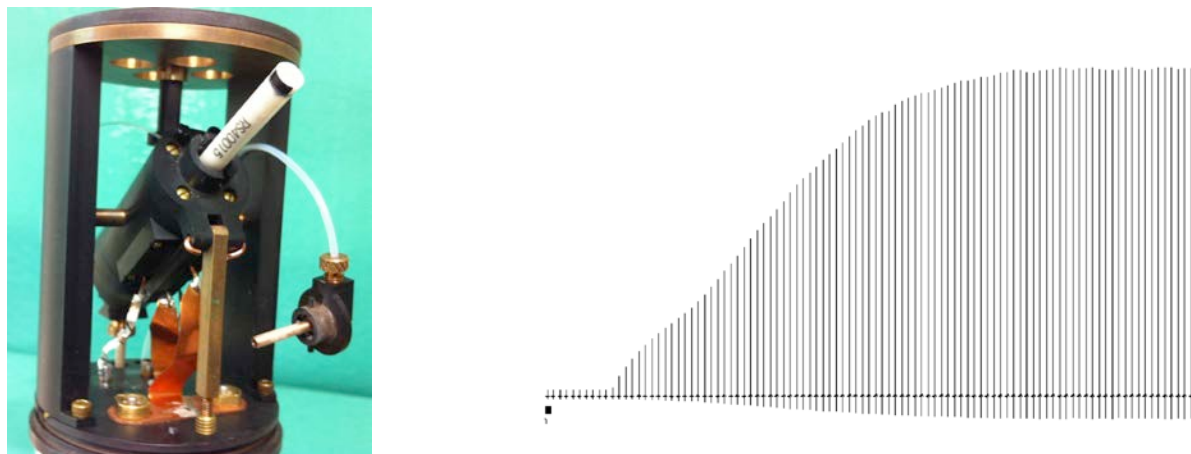
Figure C.4. Continuous-flow NMR stator (left); ^{1}H MAS of 2-butanol under flow conditions as a function of uptake time (right).

Loading of reactant molecules onto a POM catalyst material is demonstrated in the spectra of Figure C.4. The ^{1}H MAS NMR spectrum of 2-butanol is shown under flow conditions while being taken up onto a sample, demonstrating the flow capabilities and sensitivity of this probe operating at 850 MHz (for ^{1}H). The next step in these studies will involve watching the dehydration reaction of butanol on these materials at elevated operating temperatures.

C.3 Determination of Copper Siting and Ligand Coordination within Zeolite SSZ-13

One of the greatest challenges to interrogating catalytic systems is doing so under *in situ* experimental conditions. As part of this campaign, researchers within EMSL developed new *in situ* methods for obtaining EPR spectra of a copper-zeolite catalyst (Cu-SSZ-13), which is of interest as a selective catalytic reduction (SCR) material. Controlled dehydration/rehydration treatments under specific atmospheric conditions (e.g., N_2 (g), H_2 (g), etc.) were conducted to determine the effect of the reactions on the electronic state and local structure of the copper (Figure C.5). Currently, the structure of Cu-SSZ-13 is under debate, and, more specifically, the placement, hydration, and ligand association of the copper ions are largely unknown within the SSZ-13 framework (Figure C.6).

The research team at EMSL was successful in obtaining clearly resolved EPR spectra as a function of temperature, hydration state, and redox gas presence. However, these data do not directly provide structural information, but rather provide g- and A-tensors (hyperfine couplings). These EPR parameters are related to the structure and may be calculated via *ab initio* methods now available in the NWChem package. The g- and A-tensors were calculated for 11 Cu-SSZ-13 structures including explicit water, hydroxide, and O_2 (Table C.2 and Figure C.7 and Figure C.8). Based on these data, structure one (1) most closely corresponds to the experimental EPR data and also corresponds to the structure suggested by Lobo based on x-ray diffraction analyses (one of the competing structures debated within the SCR community).
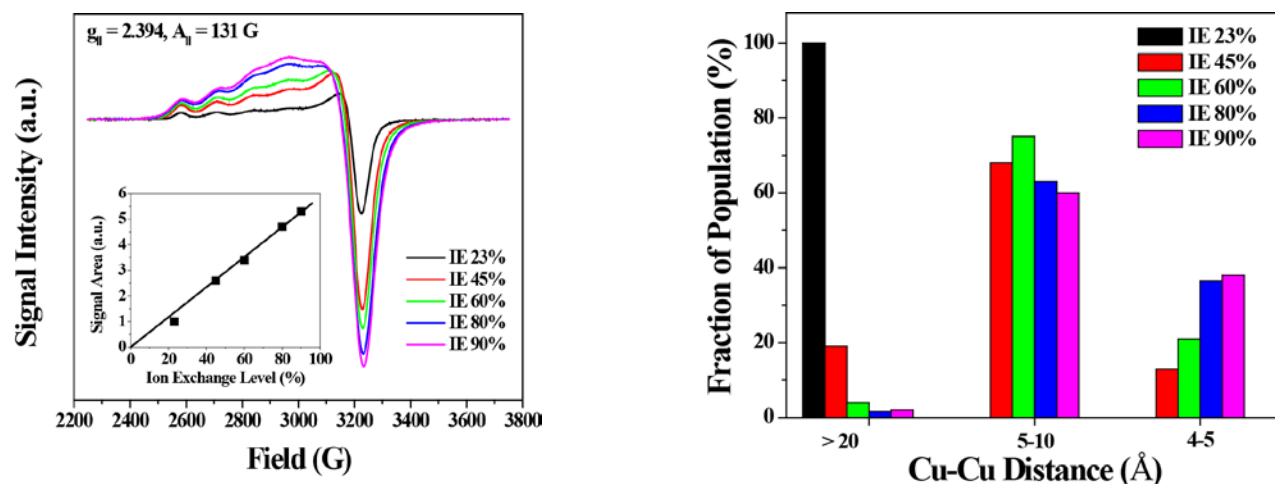
Figure C.5. EPR spectra of hydrated Cu-SSZ-13 samples measured at 155 K. The insert displays integrated signal areas versus ion exchange levels. Cu-Cu nearest neighbor distance estimations based on the line broadening of the EPR spectra shown (left).
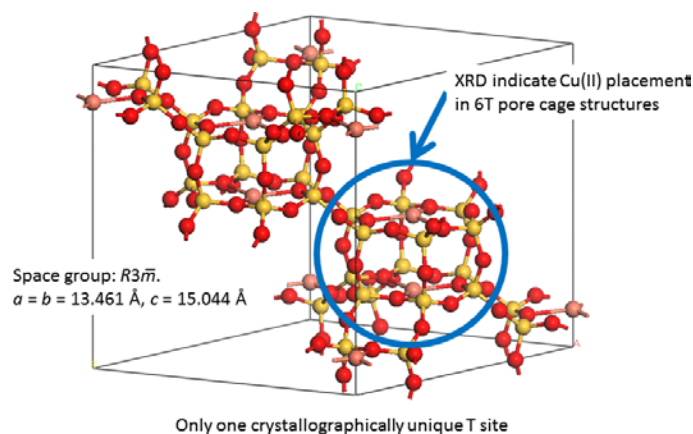
Figure C.6. Cu^{2+}-SSZ-13 (CHA) crystal structure. Si atoms are in gold, O atoms are in red, and atoms are in pink.

Table C.2. g-tensor principal components for the cluster structures 1−4 and A-tensor principal components for the Cu^{2+} center.

Structure	g-tensor			hyperfine (A) tensor (MHz)		
	g_{xx}	g_{yy}	g_{zz}	$|A_{xx}|$	$|A_{yy}|$	$|A_{zz}|$
1	2.079	2.116	2.485	94	131	354
2	1.684	2.153	2.470	109	314	357
3						
4	2.058	2.084	2.248	55	116	501
5	2.067	2.074	2.227	33	40	589
6	2.063	2.081	2.265	13	44	470
7	2.067	2.071	2.236	28	45	547
7'[(a)]	2.060	2.086	2.247	117	132	370

Table C.2. (contd)

Structure	g-tensor			hyperfine (A) tensor (MHz)		
	g_{xx}	g_{yy}	g_{zz}	$|A_{xx}|$	$|A_{yy}|$	$|A_{zz}|$
8	2.096	2.100	2.269	144	147	511
9	2.062	2.113	2.262	142	236	240
10	2.032	2.036	2.131	102	150	710
11	2.061	2.095	2.218	40	99	146
11'[b]	2.154	2.218	2.622	124	270	471

(a) Water ligand removed, structure unoptimized.

(b) O_2 ligand removed, structure unoptimized.
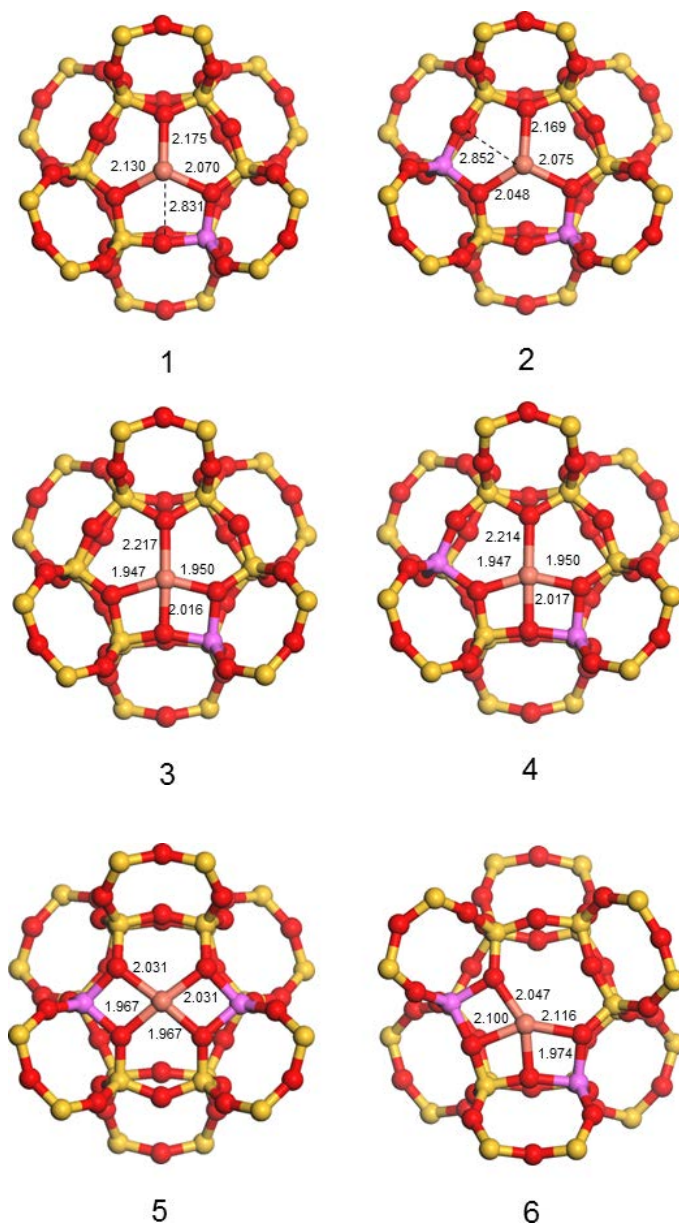

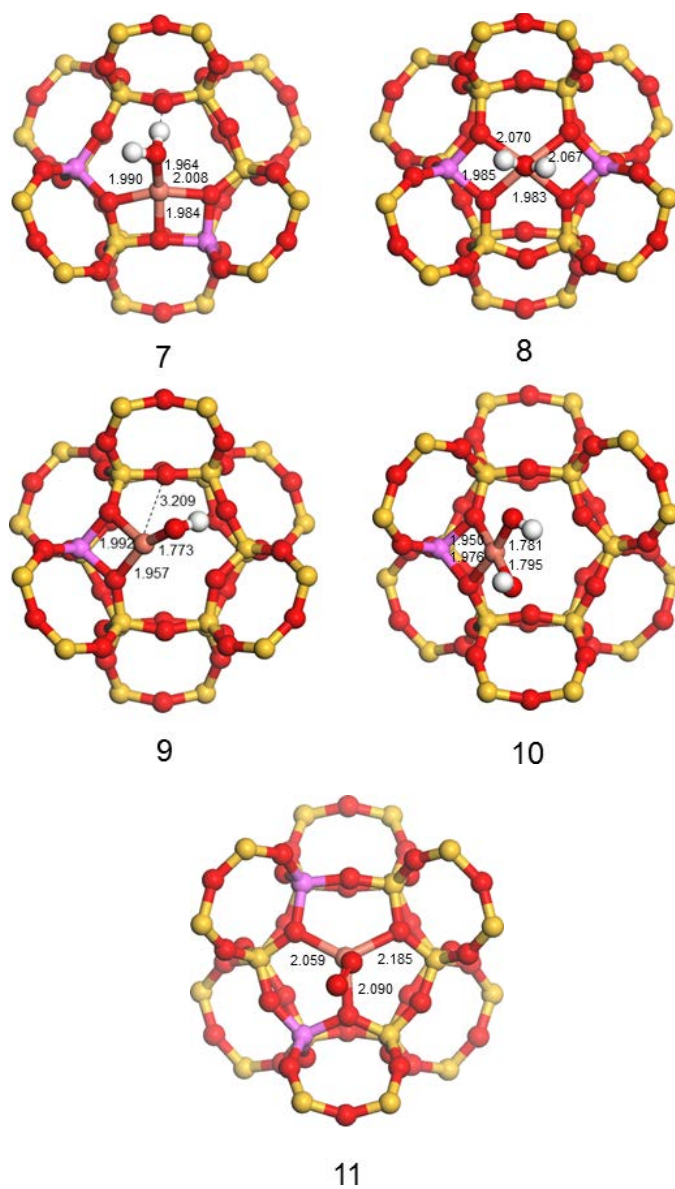**Figure C.7**. Cu-SSZ-13 structures used for calculating EPR g- and A-tensors.

Figure C.8. Cu-SSZ-13 structures used for calculating EPR g- and A-tensors.

Appendix D

Detailed Description of Selected Results from Research Campaign − Integration of Multi-Omics Technologies to Understand Complex Biology

D.1 Significance

Isoprene (2-metyl-1,3-butadiene) is a key industrial commodity because it is the monomer used for synthetic rubber production. The global demand for synthetic rubber continues to rise and is expected to reach 13.4 million metric tons by 2015 (Vocus, 2010). Under standard conditions (0°C, 100 kPa), isoprene is a colorless liquid and is primarily produced using traditional thermal cracking techniques from naphtha oil. Synthetic isoprene has advantages over naturally produced rubber (e.g., from plant sources), because it is higher in purity and in uniformity. The bulk of synthetically produced isoprene (> 90%) is used to produce polyisoprene rubber, butyl rubber, and styrenic copolymers (styrene-isoprene-styrene thermoplastic elastomer block copolymers) (Withers and Keasling, 2007, Klaus Weissermel, 2003, Kuzma et al., 1995). Methods for producing bioisoprene as a source of renewable fuels and synthetic adhesives are being investigated to meet product demand and reduce the environmental impact of current production methods that involve petroleum cracking. To that end, methods for large-scale production of bioisoprene from a microbial host are being explored as a cleaner source of raw material. Microbes naturally produce isoprene through the 1-deoxyxylulose-D-5-phosphate (DXP) pathway. Isoprene production through microbial conversion of biomass could potentially become a substitute for petroleum production.

Despite the large volume of research dedicated to understanding the isoprene biosynthesis pathway, many gaps still exist in our knowledge. Why is isoprene formed? How is the formation of isoprene regulated in bacteria? The proposed research will address this void by identifying unknown regulators of the DXP pathway. Furthermore, the work will focus on identifying the last enzyme in the pathway, isoprene synthase, which is still unknown.

D.2 Model System

Bacillus species have been used for many years in microbial fermentation for the production of high-value enzymes, and much is known about the conditions required to support industrial-level production from these species (Schallmey et al., 2004). *Bacillus* species are gram positive, spore-forming bacteria and have the potential to be engineered as a production strain for isoprene because of several critical features (Xue and Ahring, 2011). First, in comparison to other bacterial

species tested, including gram negative (*Escherichia coli*) and gram positive (*Micrococcus luteus*) bacteria, the *Bacillus* species produce the highest concentration of isoprene on a per cell basis. Of the *Bacillus* species, the *Bacillus subtilis* DSM10 strain (also referred to as the 6051 strain) produces the highest concentration of isoprene and does so at a rate that is 17 times faster than that of *E. coli* (up to 13 nmol/g of cells per hour (Kuzma et al., 1995)). For this reason, *B. subtilis* DSM10 was chosen as our model organism. The DXP pathway in *B. subtilis* is believed to be responsible for the production of isoprene in the organism (Withers and Keasling, 2007).

The ability of *B. subtilis* to produce isoprene at higher concentrations and at higher rates in comparison to other bacteria in a wild-type strain makes this bacterium a good candidate for metabolic engineering of a production strain, because it is likely that *B. subtilis* has a higher tolerance for metabolites that have a known toxicity, such as DMAPP and IPP. In addition, previous work by Zhao et. al. demonstrated that the 1-Deoxy-D-xylulose-5-phosphate synthase (Dxs) and 1-deoxy-D-xylulose-5-phosphate reductoisomerase (Dxr) enzymes in the native *B. subtilis* are more efficient in enhancing isoprene production in microbes than the Dxs and Dxr enzymes native to *E. coli* (Zhao et al., 2011). Another significant feature of *B. subtilis* is its adaptability to stressful environmental conditions such as osmotic, pH, and temperature stresses. This organism also is nonpathogenic and has been classified as GRAS (generally regarded as safe) by the Food and Drug Administration (Romero et al., 2007), alleviating regulatory restrictions on uses of dried biomass following fermentation activities.

D.3   Objectives and Approach

Successful metabolic engineering requires genetic modification of an organism to induce a specific phenotype (Vemuri and Aristidou, 2005). To identify genes that directly or indirectly affect the phenotype of interest, such as isoprene production, the system must be perturbed so that phenotype changes can be correlated with gene expression changes. Choosing different conditions that effectively perturb the pathway of interest is especially critical when attempting to identify distant genes that affect the pathway through precursor redistribution, global regulatory control, or indirect kinetic control (Alper et al., 2005, Keasling, 2010). Random mutagenesis coupled with a rational screening method has been used for decades as the technique of choice for improving wild-type strains. However, this approach is impractical for identifying IspS because the assay for measuring its activity is both indirect and time consuming. Measuring isoprene production is, by definition, an indirect measurement of enzymatic activity. In addition, the only effective measurement technique for isoprene is sampling the headspace of a culture using gas chromatography coupled to mass spectrometry (GC-MS), which takes approximately 20 minutes per sample. *B. subtilis* has a total of 4,100 genes, meaning 8200 mutants could potentially be produced using a random mutagenesis method. Measuring isoprene production from all potential mutants would require a total run time of 2,733 hours (~114 days), not including instrument maintenance time, culture preparation, or manipulation. An additional drawback of random mutagenesis is an inability to screen essential genes. If *IspS* is an essential gene, the mutant could never be created using this traditional approach (Cunningham and Wells, 1987, Kobayashi et al., 2003). Therefore, alternative methods and capabilities must be applied to resolving these unanswered questions.

D.4   Results

D.4.1   Perturbations and Isoprene Production

The rate of isoprene production in *B.s subtilis* is dependent on growth conditions and nutrient availability (Shirk et al., 2002, Sivy et al., 2002). To identify the changes in gene expression that correspond to altered isoprene production, we screened for conditions that either increase or decrease isoprene production in *B. subtilis* strain DSM10. We tested 22

environmental conditions and seven genetic perturbations. After the medium was supplemented with chemical stressors or nutrients, the production of isoprene in the culture head space was measured as previously described by members of the research team (Xue and Ahring, 2011). Because isoprene production levels are correlated to both total cell number and phase of growth (Xue and Ahring, 2011, Sivy et al., 2002), supplements were added at mid log phase (corresponding to an OD_{600} of 0.5), and isoprene levels were normalized to culture density (Figure 7).

We identified eight conditions that altered the rate of isoprene production in wild-type *B. subtilis*. Of these, addition of acetic acid and lactic acid had the most dramatic effect on isoprene production and reduced it to nearly undetectable levels. Addition of ethanol, which induces chemical stress, or indole, which functions as a signaling molecule in bacteria and is required for amino acid biosynthesis (Yanofsky, 2007, Gollnick et al., 2005), reduced isoprene production, albeit to a lesser extent than either of the acids. Addition of dimethyl sulfoxide (DMSO), a solvent that negatively affects the enzymatic activity of NAD+ synthetase and alters the available NAD/NADH pool (Nakano et al., 1998, Yang et al., 2004), caused a moderate but reproducibly lower isoprene production. Hydrogen peroxide (H_2O_2), which induces oxidative stress, was the strongest inducer of isoprene production. In plants, it has been proposed that H_2O_2-induced isoprene production evolved as a mechanism to react with and mitigate the damaging effects of reactive oxygen species (Velikova et al., 2008, Loreto and Velikova, 2001, Sharkey and Yeh, 2001, Loreto et al., 2001). In total, we identified seven media supplements that altered the production of isoprene in *B. subtilis*: 2% acetic acid, 2% lactic acid, 0.2 mg/mL indole, 70 mM DMSO, 1% ethanol, 0.005%, and 0.02% H_2O_2.

Genetic perturbations also were tested to determine if overexpression of selected enzymes in the DXP pathway would influence isoprene production. Production of isoprene proceeds through the condensation of pyruvate and glyceraldehyde-3-phosphate by the enzyme Dxs into a metabolic cascade ultimately producing either isoprene or larger terpenoid compounds (Figure D.1) (Wagner et al., 1999, Sivy et al., 2002, Kunst et al., 1997, Kanehisa et al., 2012). Several enzymes in this pathway could serve as rate-limiting steps in the production of isoprene. Specifically, metabolic flux through the DXP pathway appears to be heavily dependent on the activity of Dxs, Idi (encoded by *fni*), and IspA enzymes (Julsing et al., 2007, Broun and Somerville, 2001, Kim and Keasling, 2001, Kajiwara et al., 1997, Martin et al., 2003). The regulation of these enzymes in bacteria, however, has yet to be fully described. Because of their dispersed locations in the genome, these genes are not part of a single operon, which suggests a more complicated regulatory strategy for these genes and for the terpenoid pathway as a whole (Kanehisa and Goto, 2000, Kanehisa et al., 2012, Kunst et al., 1997).
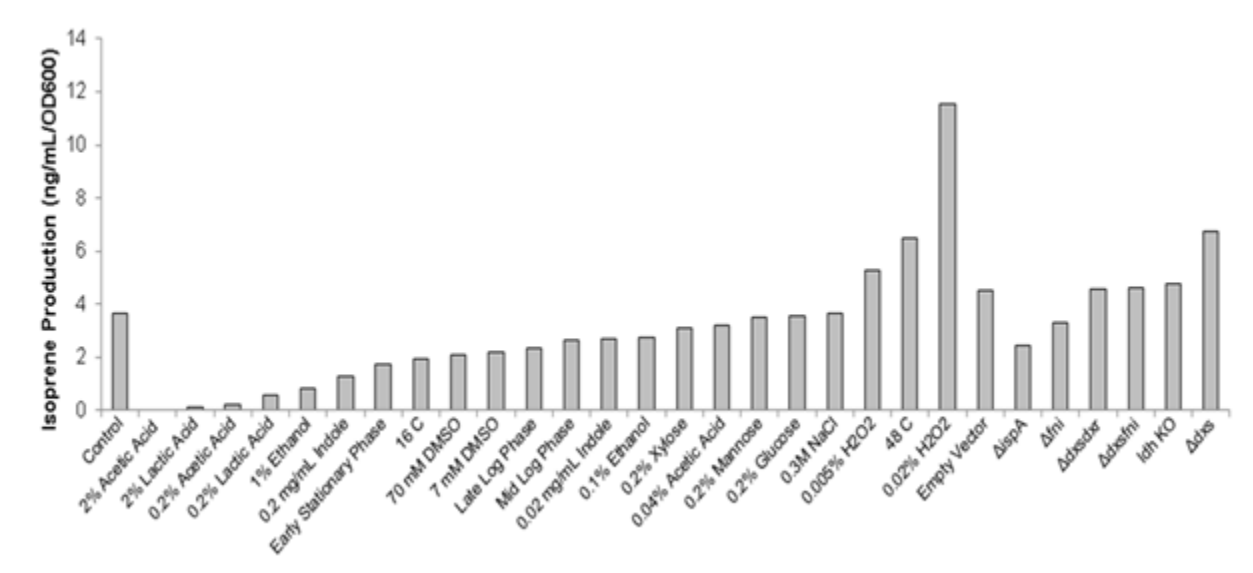
Figure D.1. Isoprene production is modulated by environmental and genetic perturbations in B. subtilis. The isoprene concentrations in the headspace were determined by GC-MS and the OD600 values of the cultures were measured using a spectrophotometer. "KO" indicates the listed gene was deleted (knock out) whereas "Δ" represents genetic perturbations in which the listed gene was overexpressed.

D.5 Transcriptomics

To determine whether the expression levels of genes that modulate isoprene production are coordinately regulated, we isolated ribonucleic acid (RNA) from *B. subtilis* cultures that had been perturbed and assessed the cellular transcription profile using RNA-seq. We were particularly interested in the expression profile of genes involved directly in the DXP pathway, as well as *mmgA*, *yhfS*, and *pksG,* which are reported to belong to a "dead end" mevalonate pathway in *B. subtilis* (Kanehisa and Goto, 2000, Kanehisa et al., 2012). The *pksG* gene also is known to play a function role in production of polyketide biosynthesis, which has recently been linked to isoprenoid biosynthesis (Figure D.2) (Calderone et al., 2006). We collectively refer to these 14 genes as the terpenoid genes.

Hierarchical clustering was used to identify the terpenoid pathway genes that are coexpressed under the 12 perturbations (Figure D.2). The 14 genes split into three main clusters that were identified. Cluster 3 contains the genes *dxs*, *yhfS*, *pksG*, *ispD,* and *mmgA*. These genes on average demonstrate similarly low levels of expression in the first six perturbations (as ordered on the heat map, discussed below) and moderate to high expression in the remaining six conditions. Cluster 2 contains the genes *ispH*, *uppS*, *hepS*, *ispE*, *dxr*, and *ispF*. These genes present the opposite pattern of expression compared to the cluster 1, being highly expressed in the first six conditions compared to the last condition. Cluster 1 contains three genes (*fni, ispG,* and *ispA*) that in comparison to the other genes have dissimilar expression patterns.
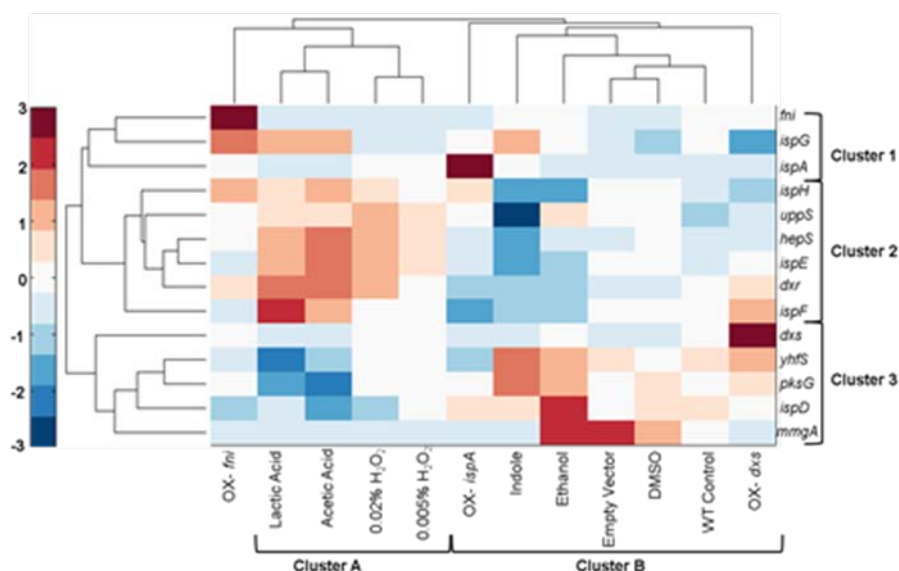
Figure D.2. The terpenoid genes cluster into three main groups with respect to coexpression and into two main groups with respect to perturbations. Two-dimensional hierarchical clustering of the terpenoid genes against isoprene production in the investigated environmental and genetic perturbations in *B. subtilis*. "OX" refers to overexpression of the listed gene.

From the clustering data, we inferred that rather than a simple linear relationship between groups of genes that would predict isoprene production, a more accurate prediction could be made by examining the relative importance of different gene expression levels on isoprene production. A genome-wide partial least squares regression (PLSR) analysis was thus performed to identify other genes or pathways that contribute to isoprene production (Karin and Kevin, 2012). This analysis showed that a subset of 213 regulated genes was sufficient to create a predictive model of isoprene production under different conditions and showed correlations at the transcriptional level.

Our initial PLSR model was created with 12 perturbations constituting the training set. This model was then tested against RNA profiles from an independent experiment of 19 perturbations that were not present in the training set and resulted in a significantly positive prediction of isoprene production (prediction fit of $R^2=0.69$). A model that was created after randomizing the isoprene levels yielded a poor model ($R^2= 0.25$). Because the extremely high level of isoprene produced in the H_2O_2 perturbations is likely to dominate the predictions of the model, a reduced training set was then used that did not include the H_2O_2 perturbations. The final model was built with this reduced training set of 10 perturbations, using three latent variables (LVs), which together captured 99.4% of the variance in the isoprene levels. Weights were used to develop matrices in the model, and the LVs were used to drive weights in the matrices. When the model was applied to the test set, the model predicted the isoprene production accurately with prediction fit of $R^2 = 0.64$ (Figure 9). The level of isoprene produced in the H_2O_2 perturbations, even though it was not in the training data, was accurately predicted to be the highest isoprene producer; however, the prediction underestimated the actual amount produced. This demonstrates that the set of 213 genes includes the necessary information to predict the relative amount of isoprene produce in novel cellular perturbations.

The need for *B. subtilis* to balance its needs for carbon through central metabolism as well as secondary metabolites requires complex regulation of enzyme expression and activity. To use bacteria such as *B. subtilis* to produce secondary metabolites such as isoprene requires understanding how this balance is maintained and how it can be shifted with the desired effect. Our study begins to address this balance as it relates to isoprene production and shows that measurements of gene expression, showing correlation at the transcriptional level, can give insight into the production of isoprene across

many simulated environmental conditions. We conclude that gene expression levels alone are sufficiently informative about the metabolic state of a cell that produces increased isoprene and can be used to build a model that accurately predicts production of this secondary metabolite.
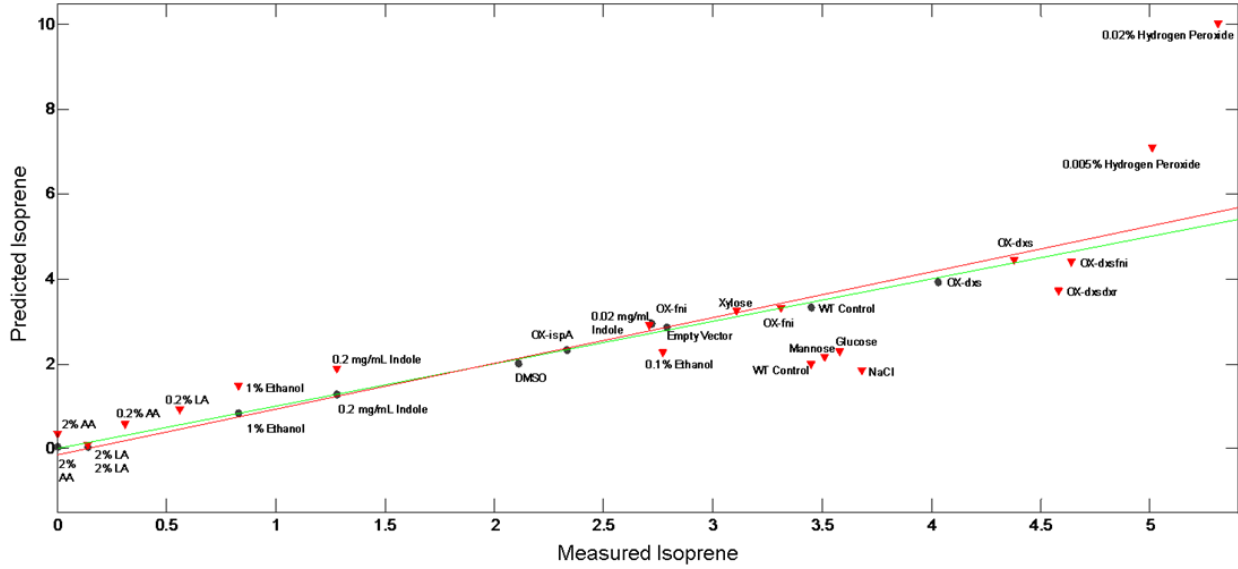

Figure D.3. Isoprene production induced by perturbations can be predicted by a PLSR model based on a reduced transcriptome. The model was created using 213 genes and trained using 10 perturbations with cross validation. Closed circles are representative of training set values; the red triangles are the test set conditions. The green line is the fit for the training data set. The red line is the fit for the testing data set. The R^2 value for prediction of the test set is 0.64. Perturbation abbreviations: AA, acetic acid; LA, lactic acid.

D.6 Modeling Translational Efficiency

Our group developed a novel approach that combines transcriptomics profiling and quantitative proteomics to determine regulatory mechanisms that control protein levels. The concept is based on steady-state growth conditions in which, like metabolites, the rate of production of a particular protein is equivalent to its rate of loss (e.g., by protein turnover or export). Therefore,

$$[Prot]_x K_{tx} = [mRNA]_x T_{eff x}$$

(D.1)

where $[Prot]_x$ is the steady-state concentration of protein x, $[mRNA]_x$ is the concentration of its corresponding message, K_{tx} is the protein-specific turnover rate constant, and $T_{eff\, x}$ is the translational efficiency of the particular mRNA. Rearrangement yields

$$\frac{[Prot]_x}{[mRNA]_x} = \frac{T_{eff}}{K_{tx}}$$

(D.2)

Thus, the steady state ratio of an individual protein to its corresponding mRNA is simply the ratio of the translational efficiency of its message to the turnover rate constant of the protein. Note that both T_{eff} and K_t are "lumped" constants that include a variety of different processes. For example, T_{eff} includes ribosomal binding and codon-dependent translation

rates. Similarly, K_t would include the rate of protein degradation and export from the cell (if relevant). Because both T_{eff} and K_t are generally gene product-specific, the relationship between the levels of proteins and their cognate messages should be different between various genes. Importantly, if the protein:mRNA ratio of specific genes remains constant at steady state following an environmental change, this implies that regulation of those genes is at the transcriptional rather than post-transcriptional level. We intend to use this approach in our data interpretation and analysis work to identify significant variations and correlations between the proteome and the transcriptome.

D.7 References

ALPER, H., MIYAOKU, K. & STEPHANOPOULOS, G. 2005. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol,* 23**,** 612-6.

BROUN, P. & SOMERVILLE, C. 2001. Progress in plant metabolic engineering. *Proceedings of the National Academy of Sciences,* 98**,** 8925-8927.

CALDERONE, C. T., KOWTONIUK, W. E., KELLEHER, N. L., WALSH, C. T. & DORRESTEIN, P. C. 2006. Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis. Proceedings of the National Academy of Sciences,* 103**,** 8977-8982.

CUNNINGHAM, B. C. & WELLS, J. A. 1987. Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure. *Protein Engineering,* 1**,** 319-325.

GOLLNICK, P., BABITZKE, P., ANTSON, A. & YANOFSKY, C. 2005. Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis. Annual Review of Genetics.* Palo Alto: Annual Reviews.

JULSING, M., RIJPKEMA, M., WOERDENBAG, H., QUAX, W. & KAYSER, O. 2007. Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis. Applied Microbiology and Biotechnology,* 75**,** 1377-1384.

KAJIWARA, S., FRASER, P. D., KONDO, K. & MISAWA, N. 1997. Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli. Biochem. J.,* 324**,** 421-426.

KANEHISA, M. & GOTO, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research,* 28**,** 27-30.

KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. & TANABE, M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research,* 40**,** D109-D114.

KARIN, J. J. & KEVIN, A. J. 2012. Modeling the latent dimensions of multivariate signaling datasets. *Physical Biology,* 9**,** 045004.

KEASLING, J. D. 2010. Manufacturing molecules through metabolic engineering. *Science,* 330**,** 1355-8.

KIM, S.-W. & KEASLING, J. D. 2001. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnology and Bioengineering,* 72**,** 408-415.

KLAUS WEISSERMEL, H.-J. A. 2003. *Industrial organic chemistry, Fourth edition*, WILEY-VCH.

KOBAYASHI, K., EHRLICH, S. D., ALBERTINI, A., AMATI, G., ANDERSEN, K. K., ARNAUD, M., ASAI, K., ASHIKAGA, S., AYMERICH, S., BESSIERES, P., BOLAND, F., BRIGNELL, S. C., BRON, S., BUNAI, K., CHAPUIS, J., CHRISTIANSEN, L. C., DANCHIN, A., DEBARBOUILLE, M., DERVYN, E., DEUERLING, E., DEVINE, K., DEVINE, S. K., DREESEN, O., ERRINGTON, J., FILLINGER, S., FOSTER, S. J., FUJITA, Y., GALIZZI, A., GARDAN, R., ESCHEVINS, C., FUKUSHIMA, T., HAGA, K., HARWOOD, C. R., HECKER, M., HOSOYA, D., HULLO, M. F., KAKESHITA, H., KARAMATA, D., KASAHARA, Y., KAWAMURA, F., KOGA, K., KOSKI, P., KUWANA, R., IMAMURA, D., ISHIMARU, M., ISHIKAWA, S., ISHIO, I., LE COQ, D., MASSON, A., MAUEL, C., MEIMA, R., MELLADO, R. P., MOIR, A., MORIYA, S., NAGAKAWA, E., NANAMIYA, H., NAKAI, S., NYGAARD, P., OGURA, M., OHANAN, T., O'REILLY, M., O'ROURKE, M., PRAGAI, Z., POOLEY, H. M., RAPOPORT, G., RAWLINS, J. P., RIVAS, L. A., RIVOLTA, C., SADAIE, A., SADAIE, Y., SARVAS, M., SATO, T., SAXILD, H. H., SCANLAN, E., SCHUMANN, W., SEEGERS, J. F., SEKIGUCHI, J., SEKOWSKA, A., SEROR, S. J., SIMON, M., STRAGIER, P., STUDER, R., TAKAMATSU, H., TANAKA, T., TAKEUCHI, M., THOMAIDES, H. B., VAGNER, V., VAN DIJL, J. M., WATABE, K., WIPAT, A., YAMAMOTO, H., YAMAMOTO, M., YAMAMOTO, Y., YAMANE, K., YATA,

K., YOSHIDA, K., YOSHIKAWA, H., ZUBER, U. & OGASAWARA, N. 2003. Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A,* 100**,** 4678-83.

KUNST, F., OGASAWARA, N., MOSZER, I., ALBERTINI, A. M., ALLONI, G., AZEVEDO, V., BERTERO, M. G., BESSIERES, P., BOLOTIN, A., BORCHERT, S., BORRISS, R., BOURSIER, L., BRANS, A., BRAUN, M., BRIGNELL, S. C., BRON, S., BROUILLET, S., BRUSCHI, C. V., CALDWELL, B., CAPUANO, V., CARTER, N. M., CHOI, S. K., CODANI, J. J., CONNERTON, I. F., CUMMINGS, N. J., DANIEL, R. A., DENIZOT, F., DEVINE, K. M., DUSTERHOFT, A., EHRLICH, S. D., EMMERSON, P. T., ENTIAN, K. D., ERRINGTON, J., FABRET, C., FERRARI, E., FOULGER, D., FRITZ, C., FUJITA, M., FUJITA, Y., FUMA, S., GALIZZI, A., GALLERON, N., GHIM, S. Y., GLASER, P., GOFFEAU, A., GOLIGHTLY, E. J., GRANDI, G., GUISEPPI, G., GUY, B. J., HAGA, K., HAIECH, J., HARWOOD, C. R., HENAUT, A., HILBERT, H., HOLSAPPEL, S., HOSONO, S., HULLO, M. F., ITAYA, M., JONES, L., JORIS, B., KARAMATA, D., KASAHARA, Y., KLAERR-BLANCHARD, M., KLEIN, C., KOBAYASHI, Y., KOETTER, P., KONINGSTEIN, G., KROGH, S., KUMANO, M., KURITA, K., LAPIDUS, A., LARDINOIS, S., LAUBER, J., LAZAREVIC, V., LEE, S. M., LEVINE, A., LIU, H., MASUDA, S., MAUEL, C., MEDIGUE, C., MEDINA, N., MELLADO, R. P., MIZUNO, M., MOESTL, D., NAKAI, S., NOBACK, M., NOONE, D., O'REILLY, M., OGAWA, K., OGIWARA, A., OUDEGA, B., PARK, S. H., PARRO, V., POHL, T. M., PORTETELLE, D., PORWOLLIK, S., PRESCOTT, A. M., PRESECAN, E., PUJIC, P., PURNELLE, B., et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature,* 390**,** 249-256.

KUZMA, J., NEMECEK-MARSHALL, M., POLLOCK, W. H. & FALL, R. 1995. Bacteria produce the volatile hydrocarbon isoprene. *Curr Microbiol,* 30**,** 97-103.

LORETO, F., MANNOZZI, M., MARIS, C., NASCETTI, P., FERRANTI, F. & PASQUALINI, S. 2001. Ozone quenching properties of isoprene and its antioxidant role in leaves. *Plant Physiology,* 126**,** 993-1000.

LORETO, F. & VELIKOVA, V. 2001. Isoprene Produced by Leaves Protects the Photosynthetic Apparatus against Ozone Damage, Quenches Ozone Products, and Reduces Lipid Peroxidation of Cellular Membranes. *Plant Physiology,* 127**,** 1781-1787.

MARTIN, V. J. J., PITERA, D. J., WITHERS, S. T., NEWMAN, J. D. & KEASLING, J. D. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotech,* 21**,** 796-802.

NAKANO, M. M., HOFFMANN, T., ZHU, Y. & JAHN, D. 1998. Nitrogen and Oxygen Regulation of *Bacillus subtilis* nasDEF Encoding NADH-Dependent Nitrite Reductase by TnrA and ResDE. *Journal of Bacteriology,* 180**,** 5344-5350.

ROMERO, S., MERINO, E., BOLIVAR, F., GOSSET, G. & MARTINEZ, A. 2007. Metabolic engineering of *Bacillus subtilis* for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. *Appl Environ Microbiol,* 73**,** 5190-8.

SCHALLMEY, M., SINGH, A. & WARD, O. P. 2004. Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol,* 50**,** 1-17.

SHARKEY, T. D. & YEH, S. S. 2001. Isoprene emission from plants. *Annual Review of Plant Physiology and Plant Molecular Biology,* 52**,** 407-436.

SHIRK, M. C., WAGNER, W. P. & FALL, R. 2002. Isoprene Formation in *Bacillus subtilis*: A Barometer of Central Carbon Assimilation in a Bioreactor? *Biotechnology Progress,* 18**,** 1109-1115.

SIVY, T. L., SHIRK, M. C. & FALL, R. 2002. Isoprene synthase activity parallels fluctuations of isoprene release during growth of *Bacillus subtilis*. *Biochemical and Biophysical Research Communications,* 294**,** 71-75.

VELIKOVA, V., FARES, S. & LORETO, F. 2008. Isoprene and nitric oxide reduce damages in leaves exposed to oxidative stress. *Plant, Cell & Environment,* 31**,** 1882-1894.

VEMURI, G. N. & ARISTIDOU, A. A. 2005. Metabolic Engineering in the -omics Era: Elucidating and Modulating Regulatory Networks. *Microbiology and Molecular Biology Reviews,* 69**,** 197-216.

VOCUS. 2010. *Global Synthetic Rubber Demand to Reach 13.4 Million Metric Tons by 2015, According to a New Report by Global Industry Analysts, Inc.* [Online]. San Francisco: San Francisco Chronicle. Available: http://www.sfgate.com/business/article/Global-Synthetic-Rubber-Demand-to-Reach-13-4-2474902.php [Accessed 11/14/2012.

WAGNER, W. P., HELMIG, D. & FALL, R. 1999. Isoprene Biosynthesis in *Bacillus subtilis* via the Methylerythritol Phosphate Pathway. *Journal of Natural Products,* 63**,** 37-40.

WITHERS, S. & KEASLING, J. 2007. Biosynthesis and engineering of isoprenoid small molecules. *Applied Microbiology and Biotechnology,* 73**,** 980-990.

XUE, J. & AHRING, B. K. 2011. Enhancing Isoprene Production by Genetic Modification of the 1-Deoxy-d-Xylulose-5-Phosphate Pathway in Bacillus subtilis. *Applied and Environmental Microbiology,* 77**,** 2399-2405.

YANG, Z. W., TENDIAN, S. W., CARSON, W. M., BROUILLETTE, W. J., DELUCAS, L. J. & BROUILLETTE, C. G. 2004. Dimethyl sulfoxide at 2.5% (v/v) alters the structural cooperativity and unfolding mechanism of dimeric bacterial NAD+ synthetase. *Protein Science,* 13**,** 830-841.

YANOFSKY, C. 2007. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA,* 13**,** 1141-1154.

ZHAO, Y., YANG, J., QIN, B., LI, Y., SUN, Y., SU, S. & XIAN, M. 2011. Biosynthesis of isoprene in Escherichia coli via methylerythritol phosphate (MEP) pathway. *Appl Microbiol Biotechnol,* 90**,** 1915-22.