ENVIRONMENTAL MOLECULAR SCIENCES LABORATORY

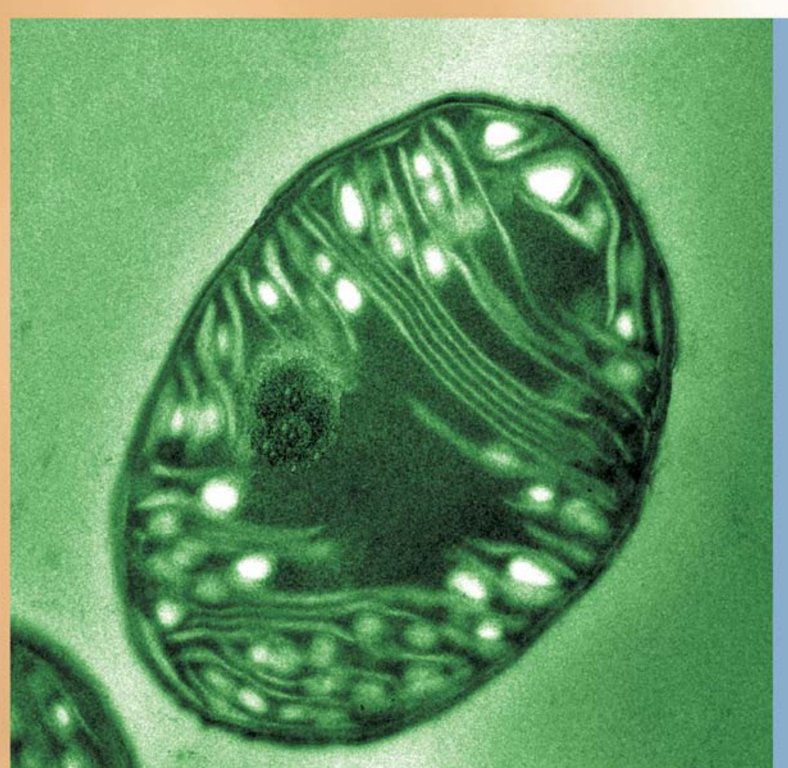
EMSL

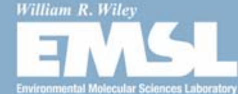
W.R. Wiley Environmental Molecular Sciences Laboratory

Membrane Biology Grand Challenge

2005-2007 Progress Report

In Response to Mission Stretch Goal 1.1.2.1 "Scientific Grand Challenges"





DISCLAIMER

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

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

Printed in the United States of America

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

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.
(8/00)

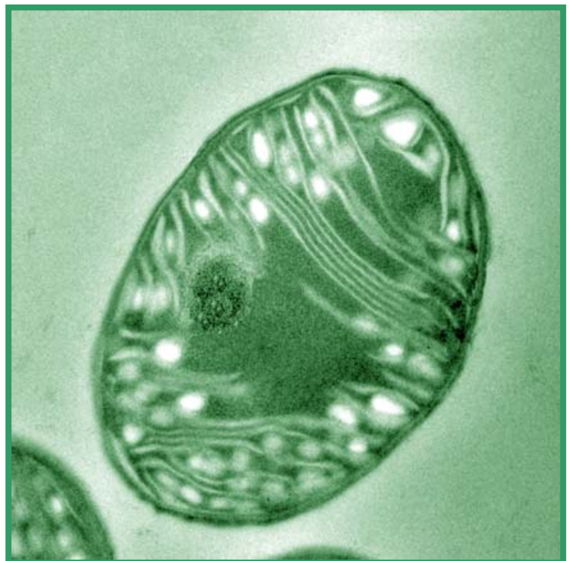
Contents

Why Study Cyanobacteria? .... **ii**
Executive Summary **1**
EMSL Grand Challenges – Development and Perspective **3**
Scientific Aims and Research Plan **5**
Technical Highlights and Accomplishments **9**
Technical Reports **11**
Summary and Future Plans **51**
Appendices and Reference Information **53**
A. Call for Concept Papers *55*
B. Awarded Proposal *57*
C. New Capabilities Resulting in Increased Scientific Impact of EMSL *63*
D. Publications/Presentations/Awards *69*
E. Advisory Committee *75*
F. Advisory Committee Report *77*
G. Participants Contact Information *81*
H. Executive Summary – Biogeochemistry Grand Challenge *85*

Cover Illustration: Micrograph of Cyanothece cell, highlighting thylakoid membrane structure and glycogen storage granules. Courtesy Alice Dohnalkova, PNNL/EMSL.

Why Study Cyanobacteria?

Cyanobacteria are oxygen producing photosynthetic microbes present in numerous ecological niches on the Earth and consume carbon dioxide as part of their natural metabolic process. Accordingly, they provide a potentially powerful biological solution for carbon-neutral energy production and carbon sequestration. The genome of *Cyanothece*, a marine cyanobacterium that executes two separate and incompatible mechanisms of nitrogen fixation in dark and photosynthesis in light, has only recently been sequenced and annotated as part of the EMSL Membrane Biology Grand Challenge project. Many cyanobacteria prevent O_2-mediated damage by compartmentalizing nitrogen fixation and photosynthesis in separate and specialized cells, physically isolating the nitrogenase apparatus, which is very sensitive to O_2, a key product of photosynthesis. In contrast, *Cyanothece* separates nitrogen fixation and photosynthesis temporally in each cell—having to switch its molecular machinery twice daily to accommodate this separation. The presence or absence of light, in part, triggers a cascade of gene synthesis, protein translation, intercellular signaling, and shifts in membrane organization that lead to the functional changeover from photosynthesis to nitrogen fixation, and vice-versa. In tandem with these processes are a host of other time-dependent processes, which the cell switches based on its internal mechanisms as if anticipating the light/dark cycle, rather than reacting to it. This combination of reacting to and anticipating the light-dark cycle is rooted in the genetics of *Cyanothece*, but results in macroscopic changes in the cell—buildup or degradation of energy storage granules, for instance. The link between sensing light, sensing time of day, gene control, membrane organization, protein content, and signaling points to a very complex interplay of physical, chemical, and biological processes which act in concert to enhance the efficiency and hence the likelihood of survival and proliferation of *Cyanothece* in its natural environment. An understanding of these mechanistic processes in an integrated systems context should provide insights into how *Cyanothece* might be optimized for specialized environments and/or industrial purposes. Engineering *Cyanothece* as a viable large-scale component of carbon sequestration or renewable energy production will require detailed understanding of its energy production and cell cycling processes at the systems level.



Cyanothece cell micrograph

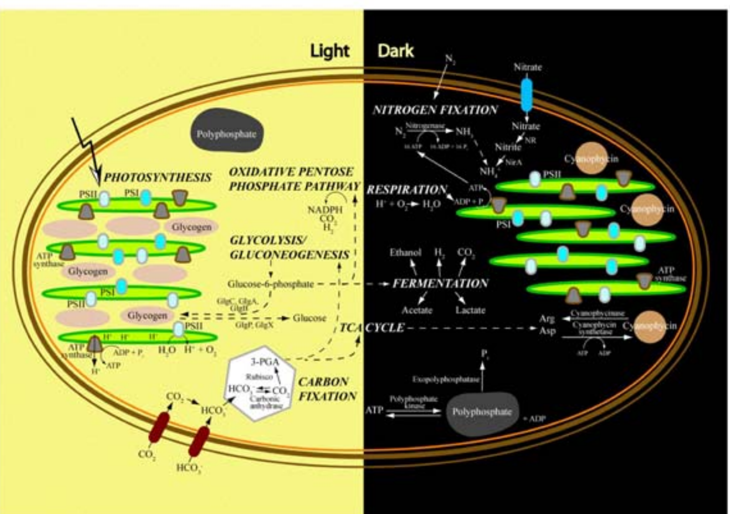


Illustration of *Cyanothece* diurnal cycle

Executive Summary

The W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a U.S. Department of Energy (DOE) national scientific user facility, has undertaken a grand challenge research effort in membrane biology. This effort was undertaken at DOE's request and upon advice from an independent scientific steering committee in early 2003. The grand challenge effort is part of a DOE Mission Stretch Goal incentive to Battelle Memorial Institute for its operation of Pacific Northwest National Laboratory (PNNL), to be formally evaluated after completion of fiscal year 2007 activities. A historical perspective of the development of the EMSL grand challenge concept is given in the following section.

The Membrane Biology Grand Challenge (MBGC) responds to a formal call issued by EMSL in 2004. The awarded project is directed at 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 homeostatis. The model cyanobacterium used in this study is a marine species known as Cyanothece sp ATCC 51142, and was chosen because it exhibits unique abilities to fix carbon via photosynthesis by day (light) and to additionally fix nitrogen by night (dark), using a robust circadian cycle that is temporally regulated by unique cellular biochemistry. The organism possesses a unique system of inner thylakoid membranes that house photosystem and other protein complexes, thus providing a special model system for studying membrane protein processes.

The research team is a multi-institutional group representing academia and private and government research institutions, with expertise in microbiology, biochemistry, proteomics and metabolomics, structural biology, imaging, and computational modeling and bioinformatics. The research team is headed by Professor Himadri Pakrasi of Washington University in St. Louis. Dr. Pakrasi's chief liaison and contact at Pacific Northwest National Laboratory is Dr. David W. Koppenaal. Additional investigators represent Purdue University, St. Louis University, the Danforth Plant Sciences Center, and PNNL/EMSL.

The primary goals of the EMSL grand challenges were to form integrated scientific teams to address complex environmental and biological problems; to demonstrate, enhance and promote EMSL capabilities; to attract new users to the EMSL facility; and to increase the scientific recognition and reputation of EMSL. The MBGC has achieved success in all of these areas by forming a very cohesive, integrated team of scientists that have been very productive in a short 2.5 years. Primary indicators of this success is the publication of over 20 papers to date and the increased interest by the scientific community in the energy-production and carbon sequestration capabilities of cyanobacterial organisms. Specific contributions include the first proteomics-guided genome annotation effort; the development of a new, ultra-fast sequence analysis tool that will great aid microbial genome homology studies; the determination of the first protein structures and types of structures in the Cyanothece model organism; and the development of uniform and reproducible microbial bioreactors that have enabled much of the described research. Additional contributions and details are described in this report.

EMSL Grand Challenges – Development and Perspective

The concept of grand challenge (GC) goals for the enhancement of the impact and use of the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) originated during discussions following an EMSL review in November of 2001. Dr. Ari Patrinos, then Director of the Office of Biological and Environmental Research (OBER) within the U.S. Department of Energy (DOE) Office of Science (SC), first used the grand challenge term to describe the type of integrated, team-science approach he wanted to see EMSL promote in tackling significant biological science problems. He challenged and subsequently directed EMSL management to implement a grand challenge approach as a way of changing the operational paradigm and impact of EMSL research within the DOE user community.

EMSL management responded by formulating and vetting several concepts for such grand challenges over the next 12-16 months. One of the first grand challenges capitalized on well-known Pacific Northwest National Laboratory (PNNL) capabilities in biogeochemistry and microbial science which have relevance to several DOE environmental science theme areas. Drs. Jim Fredrickson and John Zachara of PNNL were chosen to lead this grand challenge, and project planning for this effort was initiated (see Executive Summary for this grand challenge effort in Appendix H). Since this grand challenge was to be led within PNNL, an ancillary effort was undertaken to solicit external (non-PNNL) input for a second grand challenge. The intent was for the second GC endeavor to be focused in a topical biology area and to be led by an external, independent principal investigator. An external steering committee was assembled to recommend to EMSL management candidate biological research problems and to provide advice on a process to solicit best research concepts. This group (Marv Cassmann of the University of California at San Francisco; Len Spicer of Duke University; Steve Wiley of PNNL; Mina Bissell of Lawrence Berkeley National Laboratory; and Mike Knotek, a consultant) met with EMSL management in San Francisco in June 2003. The group recommended a biology grand challenge focused in the general areas of membrane biology and membrane protein dynamics and interactions, and provided specific input on mechanisms and process for solicitation, selection, and implementation of a second grand challenge effort.

Accordingly in early 2004, EMSL management prepared for the solicitation of research concepts from the broad scientific community. Advertisements for a scientific grand challenge in were placed in *Science* and *ASM News* in April and May (respectively) of 2004 (see Figure 1). Research proposals were requested and 14 proposals were received. Nearly all were responsive and evaluation was conducted using both internal (PNNL) and external review groups.

A Scientific Grand Challenge in Membrane Biology

Call for Concept Papers

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a multiprogram national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington, is seeking concept papers for a Scientific Grand Challenge in the area of membrane biology. Membrane processes are critical to solving highly complex problems related to energy production, carbon sequestration, bioremediation—all mission areas of the U.S. Department of Energy (DOE)—and other problems of national interest in energy and environmental science. The solutions to these problems require an approach that not only involves multidisciplinary teams of experts in the scientific area of choice, but that also require access to a range of cutting-edge instrumentation and computational resources not available at a single institution. This Scientific Grand Challenge will build interdisciplinary "team science" in biology while using multiple capabilities at EMSL to achieve a systems-level understanding of membrane processes.

Background

EMSL houses researchers who perform molecular studies focused on solving the major environmental challenges facing DOE and the nation. The facility contains a broad spectrum of world-class scientific resources that integrate theory, modeling, and simulation with experimentation and enable novel approaches for solving such difficult challenges.

In Fiscal Year 2004, EMSL is implementing a set of Scientific Grand Challenges to address important scientific questions using a collaborative approach. These Scientific Grand Challenges—each expected to last three to five years in duration— will have a set of well-defined goals that must lead to enduring scientific impact and may involve additional capability development. They will focus on critical milestones in the advancement or use of science, will be user driven, will take advantage of the full spectrum of resources and technical expertise available at EMSL, and will support DOE mission areas.

Successful concept papers submitted for EMSL's Scientific Grand Challenges will pose scientific questions that cannot readily be addressed without access for substantial periods of time to the full range of scientific instrumentation, computational resources, and research teams located at EMSL. In addition to the capabilities of the EMSL, these Scientific Grand Challenges will require teaming of scientists from multiple disciplines and use of specialized instrumentation at the participants' home institutions, and possibly use of capabilities located at other DOE user facilities (e.g., Joint Genome Institute, synchrotron light sources). The concept papers must clearly identify a scientific leader who will be responsible for the overall management of the proposed work and who is able to attract some of the best research scientists in the applicable research area.

Research Area of the Call for Concept Papers

One research area of interest to EMSL's Scientific Grand Challenges—addressed by this Call for Concept Papers—is biological membrane processes in cells (e.g., energy transduction, metal/ion regulation and transport, photosynthesis, signal transduction, cellular membrane architecture, dynamics of membrane proteins, and regulation of conformation states of proteins). Understanding membrane processes requires a systems-level analysis of fundamental cellular processes including the genetics, biochemistry, and biophysics of such processes, the characterization of which are particularly well suited to the capabilities of EMSL. Accomplishment of this Scientific Grand Challenge will elucidate how these membrane processes fit into the overall cellular physiology and ecology.

EMSL Signature Characteristics and Capabilities

EMSL's signature characteristics include

- integrated theory, modeling, and simulation with experimentation
- multidisciplinary teams and collaborative modes of operation to solve major scientific problems
- research teams responsible for developing extraordinary tools and methodologies
- scientists who design experimental strategies and operate state-of-the-art instruments
- education and training in the use of sophisticated instrumentation and computational systems and approaches
- a cyber infrastructure that facilitates productive remote interactions
- the charter to deliver capability in a transparent manner and to facilitate user outreach.

EMSL also boasts unparalleled resources and infrastructure in high-performance computing and informatics, nuclear magnetic resonance spectroscopy, multimodal optical spectroscopies, and imaging technologies, as well as advanced mass spectrometry capability for global proteomics—all of which are essential to progress in the study of membrane biology. Supporting these technologies are world-class researchers from PNNL, including chemists, biologists, physicists, and software engineers, as well as instrument designers and builders.

Process for Submitting Concept Papers

EMSL requests submission of a concept paper outlining the Scientific Grand Challenge goal, approach, technical requirements, and expected scientific and technical outcomes. Interested parties should become familiar with EMSL and its capabilities (http://www.emsl.pnl.gov/). Concept papers of no more than five pages must be postmarked by **May 14, 2004**, to the address below. An external review committee will select concepts for further development based upon scientific merit and potential for sustained scientific impact, appropriateness of EMSL resources for investigation of the challenge, and alignment with DOE mission needs in biology. EMSL will then host a workshop(s) for the successful concept team(s) to develop a scientific plan, strategy, and resource requirements for the Scientific Grand Challenge, and this information will be used for final selection of the challenge to be implemented. EMSL staff will provide assistance to the successful team in implementing their Scientific Grand Challenge. To submit concept papers or obtain further information, contact:

J. W. Rogers, Jr., PhD
Director, William R. Wiley Environmental Molecular Sciences Laboratory
P.O. Box 999, MSIN K1-46
Richland, Washington 99352
(509) 376-5328
jw.rogers@pnl.gov

Figure 1. Call for Concept Papers, as published in *Science* and *ASM News (see appendix A)*

The internal review group consisted of Drs. Harvey Bolton, Steven Wiley, George Michaels, and David Koppenaal; the

external group included Drs. Mina Bissell of Lawrence Berkeley National Laboratory; David Galas of Keck Institute; Michael Knotek, a consultant; Marvin Cassman of the University of California at San Francisco; and Professors Len Spicer of Duke University, and James Tiedje of Michigan State. Ranking of proposals was done using overall scientific merit, applicability of EMSL scientific resources, and congruence with DOE mission needs and interests as the primary criteria. Of the 14 proposals, four to five were rated as excellent and selection of a final awardee was difficult. The winning proposal was submitted by Professor Himadri Pakrasi of Washington University in St. Louis (his team included colleagues from Purdue University, St. Louis University, and the Danforth Plant Sciences Center). Their proposal title was titled "Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function in Cyanobacteria." The awardee and other submitters were advised of the call outcome September 2004, and arrangements were made to conduct a scoping workshop in December 2004 at EMSL. A technical workscope and division of task responsibilities were derived at this workshop, serving as the basis for a research subcontract to Washington University (which subsequently subcontracted to other external team members). From the outset, it was realized that a high level of interactive cooperation and coordination would be required to make the GC a success. The scoping workshop provided an excellent forum for the teams to introduce and familiarize themselves with project personnel and the capabilities of EMSL and external groups.

Funding mechanisms for the grand challenge effort presented a challenge in itself for PNNL and the DOE. After considerable negotiation, it was determined to use EMSL User Program operations funds to support activities of the external science team and to use internal discretionary (laboratory directed research & development, or LDRD) funds to support internal activities. The internal and external research programs were then designed to address separate and distinct research projects (as required by funding guidelines), the results of which could then be assimilated and integrated at the top level. The research efforts were scoped at approximately $1M levels for both internal and external teams. Full funding for both teams was awarded in April 2005 with research activities commencing in May 2005. Additional project workshops and team meetings were held, and an external advisory committee was formed. The project timeline (below) provides additional detail for the MBGC effort.

Membrane Biology Grand Challenge Project – Event/Action Timeline

Event/Action	Date
Grand Challenge Steering Committee Meeting	June 30, 2003
First advertisement for the Call in *Science*	April 2, 2004 (May 11 deadline,)
Second advertisement for the Call in *AMS*	May 2004 (June 14 deadline)
Concept Papers sent to external reviewers	June 29, 2004
Concept Papers sent to internal reviewers	July 8, 2004
Internal review scores to external reviewers	July 13, 2004
Deadline for external review	July 23, 2004
Final receipt of external reviews	August 11, 2004
Review of scores	August 20, 2004
Notification to the team members	August 30, 2004
Notification to selected proposal team	August 31, 2004
MBGC project scoping workshop at EMSL	December 2004
MBGC project funding (project start)	April/May 2005
MBGC Advisory Committee review at EMSL	November 2005
MBGC project workshop at Washington University in St. Louis	April 2007
MBGC Advisory Committee review at EMSL	August 2007

Scientific Aims and Research Plan

Membrane processes are critical to solving highly complex problems related to energy production, carbon sequestration, bioremediation, and other problems of national interest in energy and environmental science. Understanding how membrane processes fit into the overall cellular physiology and ecology requires a systems-level analysis of the genetics, biochemistry, and biophysics of membrane components and how molecular machines assemble, function, and disassemble as a function of time. EMSL's signature characteristics and capabilities provide enabling technologies to examine complex biological processes, with new computational and experimental tools to provide an integrated and predictive understanding of how an organism behaves and responds to environmental changes.

Cyanothece is a marine cyanobacterium capable of oxygenic photosynthesis, nitrogen fixation, and heterotrophic growth in the dark. This unicellular organism has evolved an elaborate circadian rhythm to use time to separate the peaks of nitrogen-fixation from the peaks of oxygen production during oxygenic photosynthesis because of the oxygen-sensitive nature of the *nitrogenase* enzyme. The diurnal patterns of nitrogen-fixation in the dark and photosynthesis in the light make *Cyanothece* a unique model organism for studying solar energy harvesting, carbon sequestration, metal acquisition, and hydrogen production.

The objective of this program is to *use systems biology approaches to determine the underlying cell signaling networks which govern the functions of cyanobacterial membranes and their components to accomplish the dramatic diurnal cycling.* The successful implementation of such an approach requires collaborative interactions among multidisciplinary teams of biologists, chemists, physicists, engineers, mathematicians, and computer scientists. The research team is headed by Professor Himadri Pakrasi of Washington University in St. Louis. Dr. Pakrasi's chief liaison and contact at PNNL is Dr. David W. Koppenaal. Additional investigators represent Purdue University, St. Louis University, the Danforth Plant Sciences Center, and PNNL/EMSL (see Figure 2). The integration of these groups brings together expertise in microbiology, biochemistry, proteomics and metabolomics, structural biology, imaging, and computational modeling and bioinformatics to achieve the objective through four primary research focus areas that address key Grand Challenge questions:

1. Proteomics-Assisted Genome Annotation: Are there unique genetic elements that explain the regulation of cell processes that govern diurnal cycling? How do you assign functional annotation to predicted novel proteins with little informative sequence homology?
2. Temporal Regulation of Cell Processes during Diurnal Cycling: How does the transcriptome and metabolome of cyanobacteria respond to environmental changes such as light intensity, CO_2 level, metal concentrations and circadian rhythm?
3. Dynamic Nature of Cell Membranes: How does the cyanobacterial membrane proteome change under similar perturbations, and how well does that relate to the transcriptome? What is the relationship between plasma membrane and thylakoid membrane in cyanobacteria?
4. Regulation of Cell Division during Diurnal Cycling: How does cell age effect nitrogenase activity in the mixed populations of a bioreactor? Can we see signs of asymmetric cell division during cycling and how does this type of cell division correlate with the rate of N_2-fixation?

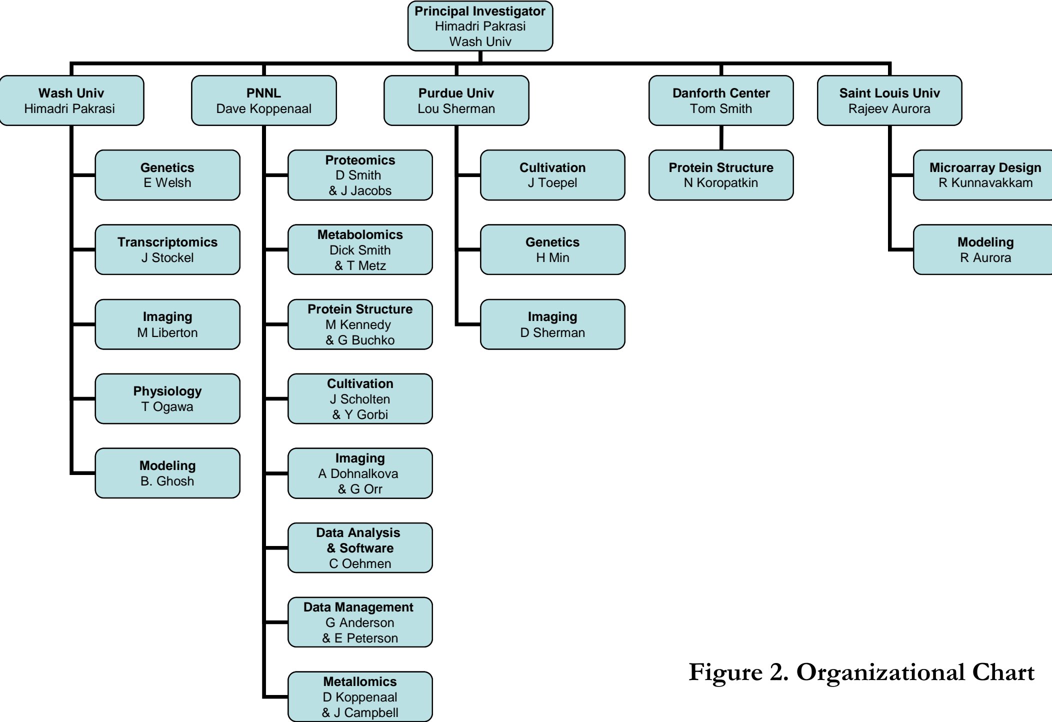


Figure 2. Organizational Chart

Each of these research areas employs collaborators from multiple institutions with diverse areas of expertise (see Figure 3). Scientific leadership for each focus area is provided by the microbiology experts on the team (H. Pakrasi and L. Sherman), and cross-project coordination is administered through regular communication between the team leads at each participating institution (H. Pakrasi at Wash University, D. Koppenaal at PNNL, L. Sherman at Purdue University, T. Smith at Danforth Plant Sciences Center, and R. Aurora at St. Louis University). Individual team members collaborate within and across institutions through means of biweekly meetings, scheduled teleconferences, e-mail correspondence and phone calls. The integration of multiple sources of temporal data collected from a single experimental system demonstrates the additive value of a systems approach to solving complex biological processes more rapidly and in more detail than otherwise possible. Our ultimate goal is to gain fundamental insights into the circadian membrane biology of *Cyanothece,* which will thereby enable further research toward engineering oxygenic photosynthetic microbes with enhanced carbon sequestration and hydrogen production capabilities.

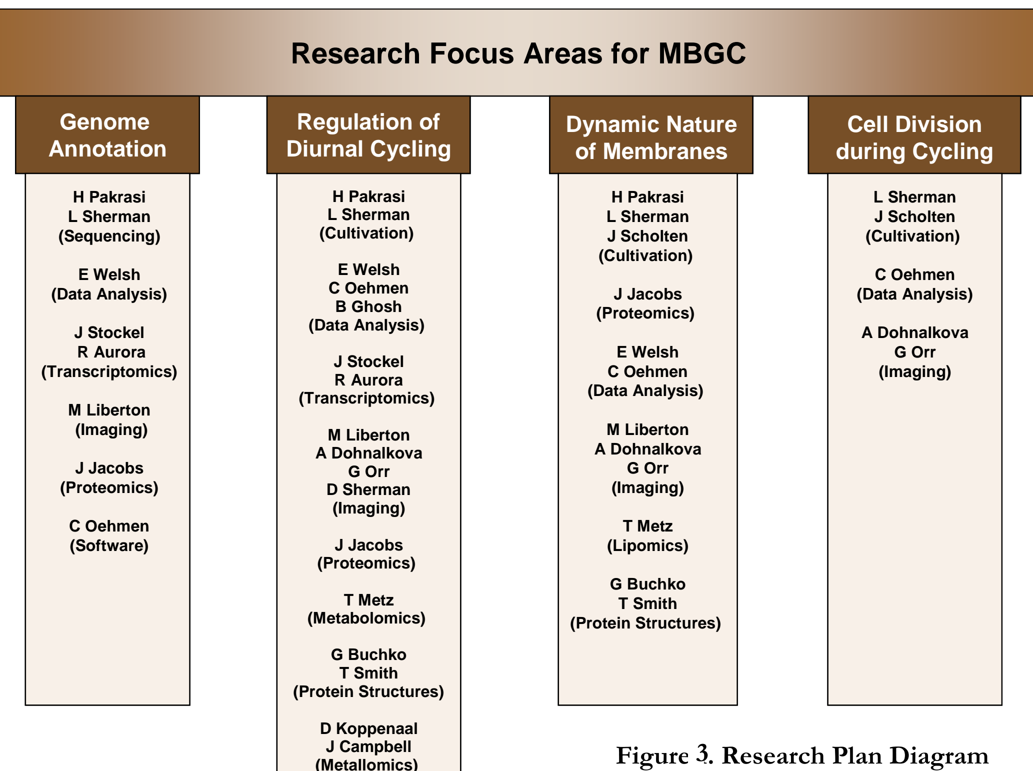


Figure 3. Research Plan Diagram

The cross-cutting technologies employed by the collaborators on the MBGC team span across all of the research focus areas. Therefore, the following technical report topics will form the basis of a series of subproject teams:

H. Pakrasi	*Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function*
L Sherman	*Analyses of the Unicellular, Diazotrophic Cyanobacterium Cyanothece sp ATCC 51142: Genetics, Imaging and Metabolic Rhythms during Growth in a 6-L Bioreactor*
J Scholten	*Ecophysiological Investigation of Cyanobacteria Using Controlled Cultivation*
A Dohnalkova	*Morphological, Functional and Redox Studies of Synechocystis 6803 and Cyanothece 51142 sp. Bacterial Membrane Complexes by Methods of Electron Microscopy*
G Orr	*Dynamic Changes in Molecular Expressions and Interactions along the Circadian Rhythm*
J Jacobs	*Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function: Proteomic, Metabolomic, and Metallomic Characterization*
T Smith	*Structural Analysis of Membrane-Associated Proteins*
G Buchko	*First Protein Structures from Cyanothece 51142*
R Aurora	*Transcriptional Profiling and Computational Modeling*
C Oehmen	*Computational Biology and Bioinformatics Tools for Understanding the Role of Membrane Proteins in Diurnal and Circadian Processes of Prokaryotes*
E Peterson	*Data Assimilation, Visualization, and Mining*

Technical Highlights and Accomplishments

The activities and accomplishments of the MBGC project are deeply rooted in the collaborative environment that encompasses more than 10 research groups at EMSL/PNNL as well as 6 external research groups and involving nearly 30 scientists and students. In the following sections, detailed descriptions are provided of the progress to date and the future research and development plans for each of these groups. Much of the progress has also depended on the unique instrumentation and technologies at EMSL that are accessible to the project team. Large-scale systems biology projects such as this one are becomingly increasingly dependent on the use of expensive, one-of-a kind instruments that are not readily available at individual laboratories, or even many of the academic research institutions. The presence of (a) high-performance mass spectrometers, (b) NMR and EPR instruments, (c) innovative biological imaging instruments, and (d) large-scale computational facilities and expertise at EMSL provides an ideal environment for today's large scale genomics and post-genomics based systems biology projects. The MBGC project, in many ways, provides a testament to this assertion.

Selected technical accomplishments of the MBGC team include:

- Development of a new bioreactor design that provides a basis for uniform cultivation of microbial samples for the entire MBGC team. A standardized platform was developed that allows different teams at different locations to cultivate and grow cyanobacteria samples consistently and uniformly. This enables tailored experiments that can be done much more expediently, allows cross-laboratory comparisons of cultivation conditions, and provides for confirmation and validation of results and observations. Three standardized bioreactors were designed, built, and located at participant institutions (PNNL, Washington University, and Purdue University).

- Proteome assisted genome annotation of *Cyanothece* that has resulted in the submission of a paper to Nature. PNNL/EMSL's premier global proteomics capability, using accurate mass & time tag (AMT) technology, was used to ascertain and confirm existence of many previous hypothetical proteins, significantly improving the completeness and utility of the annotated genome. This is the first and most comprehensive effort of this combined genomics/proteomics approach, and will, we believe, set a new standard for genome annotation endeavors.

- Determination of the first biomolecular structures of *Cyanothece* proteins and the delineation of protein structures for important nutrient transporters in cyanobacteria. Twelve protein structures have been determined by two research groups involved in the MBGC project. Several of these have been membrane-associated proteins responsible for delivering critical nutrients (Fe, Zn, HCO_3^-, and NO_3^-) to the cell.

- Development of ScalaBlast tool for rapid homology studies at multiple genomes level. This tool, a very high-throughput sequence analysis engine, was used to perform high-speed BLAST calculations on the EMSL supercomputer. Over 2 CPU years of data analysis was performed. This capability will be made available to DOE's Joint Genome Institute (JGI) as

part of a new EMSL collaboration. This capability and the collaboration will become immediately useful to the MBGC project as the JGI is sequencing an additional 5-6 Cyanothece strains for comparative genomics study.

- Development of a detailed picture of the ultrastructure of a cyanobacterial cell and its constituent biological membranes. Procedures have been developed to freeze *Cyanothece* cells so that inherent structural relationships of membranes, internal storage granules, and other organelles are preserved. New insights into photosynthetic membrane continuities and granule organization have resulted, using 3-D tomographic techniques to reconstruct entire volumes of the cell. EMSL researcher Alice Dohnalkova won a Diatome U.S. First Place award in August 2007 for parts of this work.

- Network models of cyanobacterial circadian physiology based on global transcriptomics analysis. Two versions of transcriptome microarrays for *Cyanothece* have been designed and produced for the project. Gene expression profiles are used and modeled to better understand enzyme and metabolite pathways, and to deduce or infer the operative biology. Several intriguing insights regarding the diurnal cycle biochemistry of *Cyanothece* are resulting. The models developed provide a solid foundation for future research activities in this systems biology project.

Additional detail on these and other accomplishments are provided in the following detailed project reports. Additionally, public website describing the MBGC teams and their research projects is available at http://mbgc.emsl.pnl.gov. This website serves the need to disseminate information to new researchers or potential collaborators and to release data to the global scientific community after publication.

Technical Reports

Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function

Himadri Pakrasi, Michelle Liberton, Jana Stockel, and Eric Welsh

Cyanothece sp ATCC 51142 is a metabolically versatile organism, with an intriguing ability to store photosynthetically fixed carbon in glycogen granules during the day, and utilize these stored carbons for biosynthesis and energy production during night. Our overall goal is to use systems-level approaches to develop a detailed and predictive physiological model for this bacterium. Specifically, the activities in our research group have been focused on three areas: 1) sequencing and annotation of the genome of *Cyanothece* 51142, 2) transcriptomic and proteomic analyses of the diurnal rhythm in *Cyanothece* 51142, and 3) ultrastructural studies of *Cyanothece* cell membranes during a diurnal cycle.

Sequencing and Annotation of the Genome of *Cyanothece* 51142

Cyanothece 51142 is a unicellular diazotrophic marine cyanobacterium capable of separating the incompatible processes of oxygenic photosynthesis and nitrogen fixation temporally within the same cell, performing photosynthesis during the day and nitrogen fixation at night. The genome of *Cyanothece* 51142 was sequenced and found to contain a unique arrangement of one large circular chromosome, four small plasmids, and one linear chromosome, the first report of such a linear element in a photosynthetic bacterium (see Figure 4). Annotation of the *Cyanothece* genome was aided by the use of high-throughput proteomics data, enabling the reclassification of nearly 25% of the predicted hypothetical proteins. Phylogenetic analysis suggests that nitrogen fixation is an ancient process that arose early in evolution and has subsequently been lost in many cyanobacterial strains.

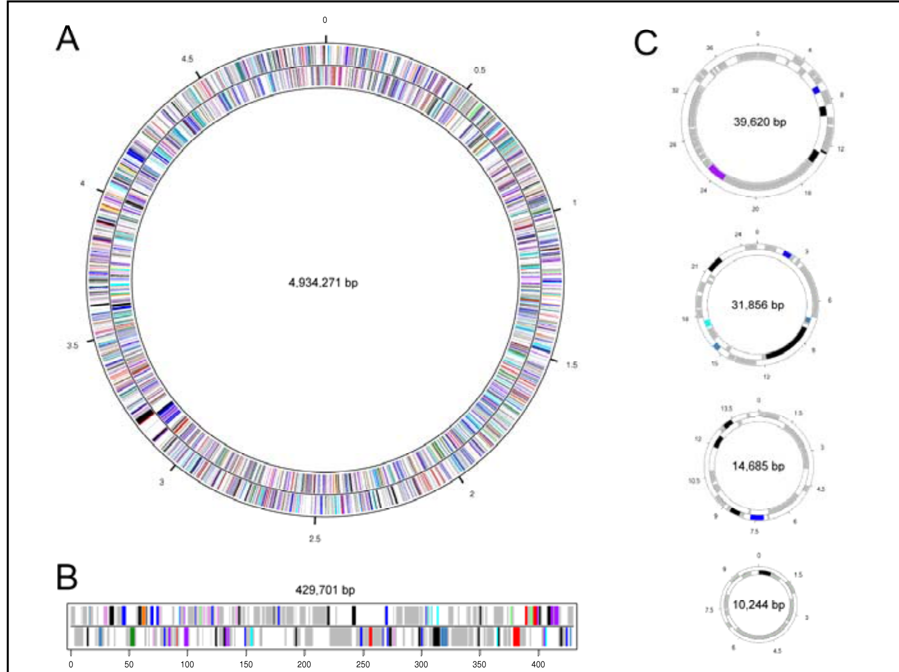


Figure 4. Distribution of gene functions within the Cyanothece genome. Labels are given in megabase pairs for the circular chromosome (a), and kilobase pairs for the linear chromosome (b) and plasmids (c).

Transcriptomic and proteomic analyses of the diurnal rhythm in *Cyanothece* 51142

Unicellular diazotrophic cyanobacteria, such as *Cyanothece* 51142, are challenged with creating and maintaining both aerobic and obligate anaerobic intracellular conditions to fulfill the requirements for oxygenic photosynthesis and nitrogen fixation during a diurnal cycle. Key to deciphering the cellular processes involved in regulating such a dramatic cycle is the identification of gene regulatory events. Global gene expression analysis in *Cyanothece* pinpointed a significant impact of nitrogen fixation on the diurnal cycle of different fundamental pathways and thus provided new systems-level insights. Our study revealed that 27% of all genes in the genome exhibit oscillating expression patterns. Among those genes, 48% are of unknown functions. We observed a remarkable co-regulation of functionally related genes and entire pathways coupled with an increased metabolic diversity and activity during the night (see Figure 5). Furthermore, we found that steady state transcript abundance cycles with a significantly higher transcript accumulation during the dark period. Our analyses suggest that nitrogen fixation, not photosynthesis is determining the major metabolic activities inside the cell and is thus driving the majority of cellular activities.

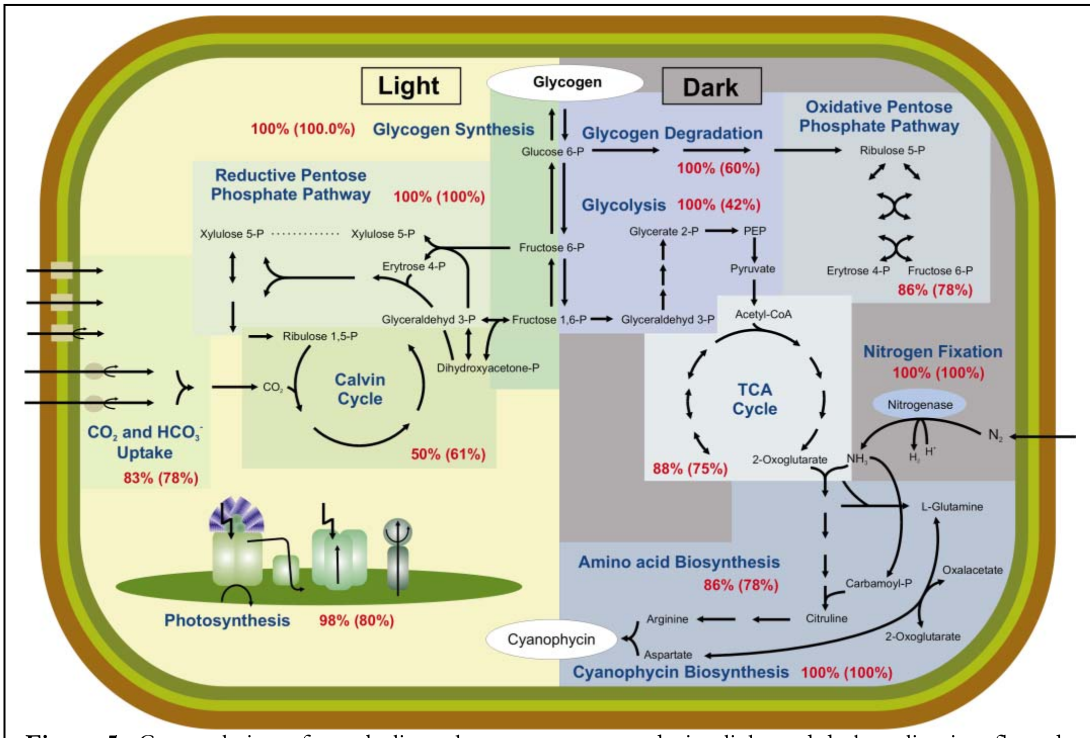


Figure 5. Co-regulation of metabolic pathway components during light and dark cycling is reflected in microarray data.

Parallel to the transcriptomic analysis, we collected samples for global proteomic analysis of *Cyanothece* during a diurnal cycle. These samples were sent to Jon Jacobs and his colleagues at PNNL. Detailed analysis of these samples is in progress and included in the Proteomics' team report.

Ultrastructural studies of *Cyanothece* cell membranes during a diurnal cycle

We have recently completed a detailed tomographic analysis of the membrane architecture in *Cyanothece* 51142 cells. Cyanobacterial cells in general, and *Cyanothece* 51142 in particular, have a

typical gram-negative envelope layer. In addition, they have an elaborate internal thylakoid membrane system. Cyanobacteria are the progenitors of plant chloroplasts. Thylakoids in cyanobacteria and the chloroplasts of algae and plants are the sites where photosynthetic electron transport occurs, an enzymatic reaction fundamental to planetary life. The morphology of thylakoid membranes is intimately related to the physiology of the reactions that occur there, but many details of thylakoid morphology are not well understood, particularly in plant chloroplasts. Cyanobacteria, as the progenitors of chloroplasts, provide a model system to explore the origin of thylakoid morphology. In *Cyanothece* sp. ATCC 51142, a unicellular cyanobacterium, thylakoids display a rudimentary helical organization, a potential evolutionary step to the modern grana and stroma thylakoid arrangement in plant chloroplasts. Our studies have indicated that a defining feature of chloroplast thylakoids is present in a cyanobacterium, and is not a new invention in plants.

Publications, Presentations, and Awards

1. Liberton, M., R.H. Berg, J. Heuser, R. Roth, and H.B. Pakrasi. 2006. "Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803." *Protoplasma*, 227:129-138.
2. Koropatkin, N.M., H.B. Pakrasi, and T.J. Smith. 2006. "Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity." *Proc. Natl. Acad. Sci.* 103:9820-9825.
3. Buchko, G.W., S. Ni, H. Robinson, E.A. Welsh, H.B. Pakrasi, and M.A. Kennedy. 2006. "Characterization of two potentially universal turn motifs that shape the repeated five-residues fold--crystal structure of a lumenal pentapeptide repeat protein from *Cyanothece* 51142." *Protein Science* 15:2579-2595.
4. Buchko, G.W., H. Robinson, S. Ni, H.B. Pakrasi, and M.A. Kennedy. 2006. "Cloning, expression, crystallization, and preliminary crystallographic analysis of a pentapeptide repeat protein (Rfr23) from the bacterium *Cyanothece* sp. 51142." *Acta Crystallgr.* 62:1251–1254.
5. Liberton, M. and H.B. Pakrasi. 2007. "Membrane systems in cyanobacteria." In the *Cyanobacteria: Molecular Biology, Genomics and Evolution* (Herrero, A. and Flores, E., eds.), pp. 271-287, Horizon Scientific Press, Norwich, UK.
6. Koropatkin, N.M., D.W. Koppenaal, H.B. Pakrasi, and T.J. Smith. 2007. "The Structure of a Cyanobacterial Bicarbonate Transport Protein, CmpA." *J. Biol. Chem.* 282:2606-2614.
7. Aurora, R., Y. Hihara, and H.B. Pakrasi. 2007. "A Network of Genes Regulated by Light in Cyanobacteria." *OMICS*, 11:166-185.
8. Wei, B., A.M. Randich, M. Bhattacharyya-Pakrasi, H.B. Pakrasi, and T.J. Smith. 2007. "Possible Regulatory Role for the Histidine-Rich Loop in the Zinc Transport Protein, ZnuA." *Biochemistry* 46:8734-8743.
9. Welsh, E.A., M.L. Liberton, J. Stockel, H. Sato, T. Loh, C. Wang, A. Wollam, R.S. Fulton, S.W. Clifton, J.M. Jacobs, R. Aurora, L.A. Sherman, R.D. Smith, R.K. Wilson, and H.B. Pakrasi. 2007. "Proteomics Assisted Genome Sequence of Cyanothece 51142, a Unicellular N_2-fixing Cyanobacterium." Presented by Eric Welsh at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, Wisconsin on June 6, 2007.
10. Welsh, E.A., M. Liberton, J. Stöckel, H. Sato, T. Loh, C. Wang, A. Wollam, R.S. Fulton, S.W. Clifton, J.M. Jacobs, R. Aurora, L.A.. Sherman, R.D. Smith, R.K. Wilson, and H.B. Pakrasi. 2007. "Genome sequence of *Cyanothece* 51142 shines new light on the circadian life style of a cyanobacterium." Submitted *Nature*.

Analyses of the Unicellular, Diazotrophic Cyanobacterium *Cyanothece* sp ATCC 51142: Genetics, Imaging and Metabolic Rhythms during Growth in a 6-L Bioreactor

Louis Sherman, Hongtao Min, and Jörg Toepel

The main goals of this project are to study *Cyanothece* throughout its diurnal cycle from many different perspectives and to obtain the annotated genome. The ultimate objective is to develop a complete metabolic and structural model to determine how cellular components, especially membranes, operate as a function of time throughout the day/night cycle. Our component of the overall program involves three projects. The first is genetics, so that we can target specific genes in order to knockout their ability to produce their protein products. In this way, we can test very precisely the type of predictions made in the overall metabolic model. A second objective has been to use the modern techniques of high-pressure freezing and 3-D tomography in order to determine the precise structure of the cell and cellular components. The third objective has been to determine appropriate conditions for growth of *Cyanothece* in a large bioreactor such that similar growth conditions can be achieved at Purdue, Washington University, and PNNL.

The basic premise for the project was established in this lab some years ago. At that time, we determined that this cyanobacterium could grow beautifully under N_2-fixing conditions and that it separated the oxygen-sensitive nitrogenase from oxygen evolving photosynthesis via temporal regulation. Thus, when cells were grown under N_2-fixing conditions, nitrogen fixation took place exclusively at night and photosynthesis exclusively in the dark. However, the metabolic rhythms persisted for many days and it was clear that the entire process was controlled by the circadian clock. At that time, all of the work was done one gene or one protein at a time and, although we developed an overall idea of the process, our overall knowledge was rather limited. Therefore, the ability to study virtually all components of the cell using the more modern high throughput techniques was both daunting and exhilarating. Therefore, one major objective of this project has been to obtain transcriptomics, proteomics and metabolomics data as a function of time over the daily life cycle of *Cyanothece*. Our involvement has been to lead the research in areas such as genetics and cultivation and then to provide material to various groups at EMSL for analysis by a variety of techniques. We have performed the transcriptomics ourselves using Agilent microarrays developed by Rajeev Aurora (see his technical report for more information). This project demands the talents and techniques of many different types of scientists and can only be successful in collaboration with scientists at EMSL.

Genetics

We quickly worked out procedures for transformation using techniques such as electroporation and $CaCl_2$ treatment. We used a broad-host-range plasmid, pRL1383a, and obtained high frequencies of transformation depending upon the exact nature of growth. The efficiency of transformation depended upon growth conditions and, the faster the growth, the higher the efficiency of transformation. We then determined that the same growth rate dependency was true for a fraction of DNA that we first thought was plasmid. This is where the genome sequence data became very important. The indication that *Cyanothece* contained a linear chromosome provided an important clue to our situation. Thus, we hypothesized that the higher growth rates generated a higher copy

number of the linear chromosome and led to a higher rate of transformation. We have shown that this correlation is correct.

Our second goal in genetics was homologous recombination that would be used to construct individual gene knockout mutants so that we could target specific genes for mutagenesis. However, we have not yet been able to construct such knockouts, in part, because there is both non-homologous and site-specific recombination occurring in *Cyanothece* that competes with the homologous recombination. From the sequence, we know that there is a site specific recombinase that targets the DNA into specific sites and overwhelms the targeted mutagenesis into specific genes. We are now developing approaches that can overcome this problem. However, the genome sequences also provide us with some very valuable information so that we may be able to use an RNAi-like system to target specific genes and to produce what is called knock-down mutations. We have also begun introducing GFP markers into strains of *Cyanothece*.

Imaging of Dynamic Membrane Structures

One objective has been to identify the best procedures for the study of the 3-dimensional structure of *Cyanothece*. We have worked with high pressure freeze (HPF) techniques and freeze substitution to develop fixation procedures that stabilize membranes and internal storage granules. Meeting both goals was difficult, but we have done so and provided this information to other labs within the project. We have now used these procedures to obtain high-resolution tomographic images of *Cyanothece* and are working to develop a model of glycogen granule structure in relation to the photosynthetic membrane. We are also interested in the relationship of the photosynthetic and cytoplasmic membranes and other particles that interact with one or both membrane systems. We have found that the glycogen granules are not necessarily discreet entities, but represent beads on a string (see Figure 6). Therefore, depending upon the projection, the granules come together and all the glycogen granules between two photosynthetic membranes are in contact with each other.

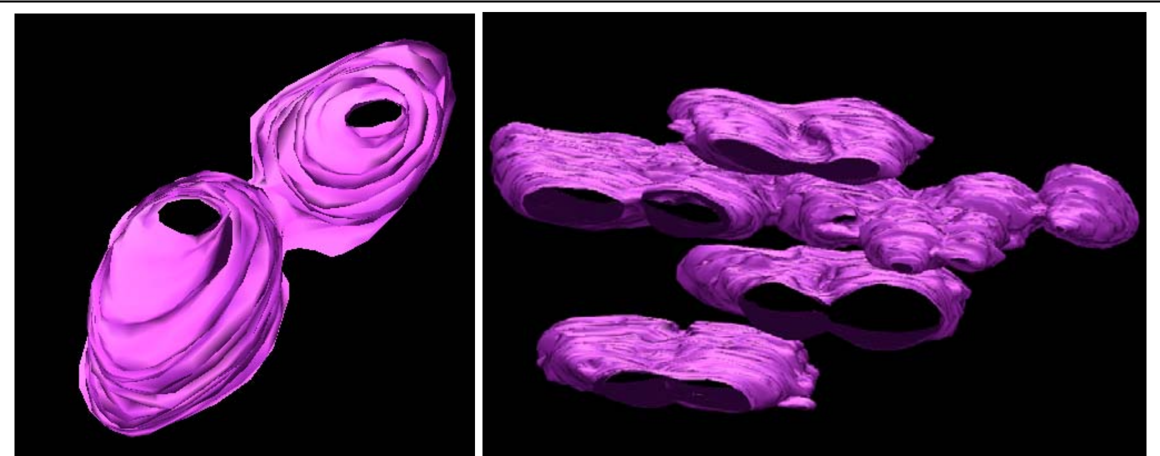


Figure 6. 3-D model of glycogen granules. It is evident that the external morphology of the granules is different at L4 (left) and at D4 (right). The granules are smoother in the light as they are being produced and much rougher in the dark when they are being degraded. L4 = 4 hrs into light cycle; D4 = 4 hrs into dark cycle.

These images have also given us a good idea of how the glycogen granules form or are degraded in the light and dark, respectively. We have also imaged a number of small particles near the interface of the photosynthetic and cytoplasmic membranes and we are interested to know if these are involved in assembly. We are now completing the modeling of these structures so that we have a good representation of these components within a part of the cell.

Cultivation and Regulation of Cell Processes during Diurnal Cycling

One of our objectives was to develop a bioreactor suitable for consistent growth of a cyanobacterium-like *Cyanothece*. The bioreactor was developed in conjunction with Dr. Pakrasi and Dr. Gorby and has proven to be well conceived. The resulting vessel is good for cell growth and for metabolic analyses and provides us with the type of information we need to investigate the metabolic rhythms. We have found that cells can be grown with a high degree of reproducibility under nitrogen-fixing, light-dark conditions including with periods of continuous light or continuous dark. Such cells are very well synchronized and can be used for all of the high throughput experiments discussed above.

We have concentrated specifically on physiological and transcriptional analysis of cells growing under light-dark conditions with periods of continuous light. The specific question that we have asked is: how does the regulation of cells growing under a subjective dark condition (in the presence of light) compare to true dark? We have used a variety of different primers, including pH and dissolved oxygen, which can be measured in the bioreactor, along with photosynthesis, respiration, nitrogen fixation and glycogen concentration measurements. We determined that glycogen is not degraded as quickly in continuous light and that glycogen storage remains high under these conditions (see Figure 7). Once cells are returned to the dark, there is a huge N_2-fixation peak; in addition, respiration is very high and glycogen is quickly degraded. We have done a thorough transcriptional analysis of both light/dark and continuous light, that follows on the work done by Jana Stockel in Dr. Pakrasi's laboratory. We were able to show that our results compare very nicely to those done under light/dark conditions at Washington University, and we also obtained

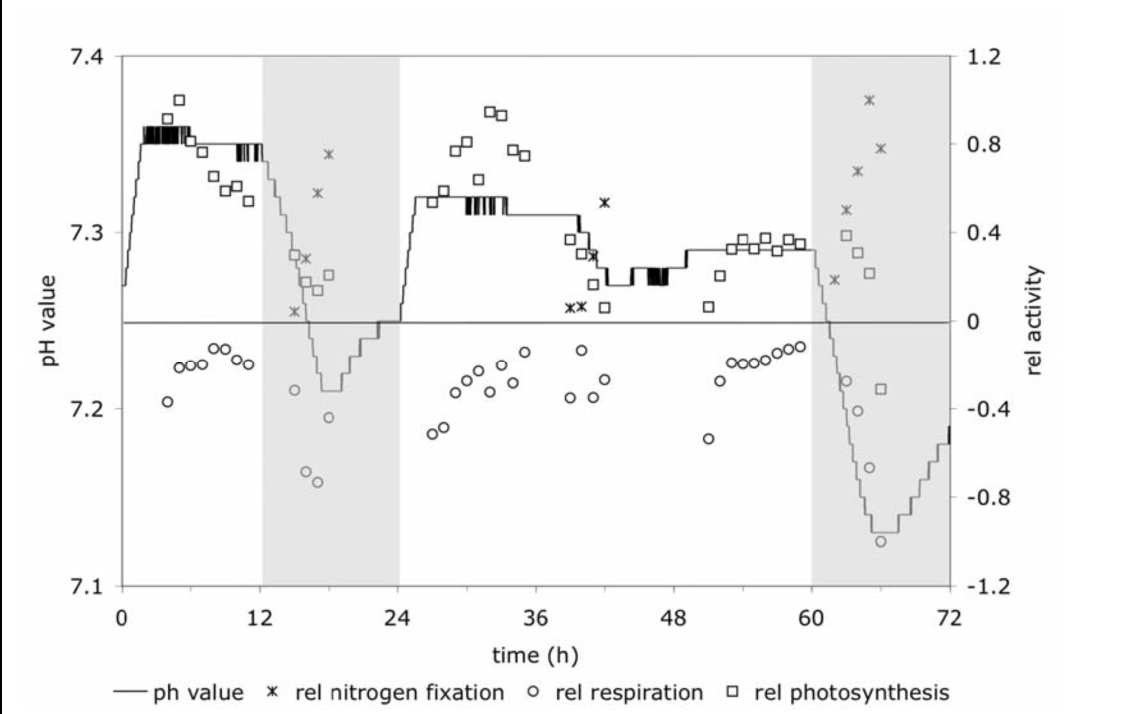


Figure 7. *Cyanothece* response to a light-dark to light-light to light-dark transfer in the absence of nitrogen. *Cyanothece* showed the highest nitrogen fixation activity at D6, combined with high respiration rate and a reduced photosynthetic rate. Under continuous light, the nitrogen fixing activity was reduced by 50% and the respiration was not enhanced, but both processes still peaked at D6.

significant data on the transcription during continuous light (LL) versus light/dark cycling (LD). In particular, we found that the gene encoding the protein involved in glycogen degradation is not induced as much as in the subjective dark (second light phase of LL) as in the regular dark providing a reason for the high accumulation of glycogen in the subjective dark. We are now in the process of providing such material to the metabolomics and proteomics group to obtain a thorough understanding of these phenomena.

Publications, Presentation, and Awards

1. Sherman, L.A. 2005 "The unicellular, diazotrophic cyanobacterium *Cyanothece*--genomic sequence and metabolic rhythms." Marine Cyanobacteria: Evolution, Function and Genomes. Stockholm, Sweden, August 24-27, 2005.
2. Welsh, E.A., L.A. Sherman, and H.B. Pakrasi. 2006. "Genome Sequence Analysis of *Cyanothece* sp ATCC 51142, a Unicellular Nitrogen-fixing Cyanobacterium." 1st Annual Finishing the Genome in the Future Meeting, Santa Fe, New Mexico, May 4-5, 2006.
3. Min, H-T. and L.A. Sherman. 2006. "Genetic Transformation and Mutagenesis in the Unicellular Diazotrophic Cyanobacterium *Cyanothece* sp ATCC 51142." 12th International Symposium on Phototrophic Prokaryotes, Pau, France, August 27 - September 1, 2006.
4. Welsh, E.A., M. Liberton, J. Stockel, H.B. Pakrasi, H. Min, J. Toepel, and L.A. Sherman. 2007. "Genomics and Systems Biology analysis of *Cyanothece* sp ATCC 51142, a unicellular, diazotrophic cyanobacterium with robust metabolic rhythms." Plenary lecture, 15th International Nitrogen Fixation Congress, Cape Town, South Africa, January 21-28, 2007.
5. Min, H. and L.A. Sherman. 2007. "Genomics and Genetic Transformation in the unicellular, diazotrophic cyanobacterium, *Cyanothece* sp ATCC 51142." 15th International Nitrogen Fixation Congress, Cape Town, South Africa, January 21-28, 2007.
6. Min, H. and L.A. Sherman. 2007. "Genetic transformation and recombination in the unicellular, diazotrophic cyanobacterium *Cyanothece* sp ATCC 51142." 9th Cyanobacterial Molecular Biology Workshop, Delavan, Wisconsin, June 7-10, 2007.
7. Toepel, J. and L.A. Sherman. 2007. "Metabolic regulation of the cyanobacterium *Cyanothece* sp. ATCC 51142 when grown in light/dark or in continuous light." 9th Cyanobacterial Molecular Biology Workshop, Delavan, Wisconsin, June 7-10, 2007.
8. Shukla, V.B., N. Jiang, and L.A. Sherman. 2007. "Fermentation ability of the cyanobacterium *Cyanothece*." 9th Cyanobacterial Molecular Biology Workshop, Delavan, Wisconsin, June 7-10, 2007.
9. Welsh, E.A., many collaborators and L.A. Sherman. 2007. "Complete genome Sequence of *Cyanothece* sp ATCC 51142." DOE Joint Initiative Meeting "Finishing Genomes in the Future." Santa Fe, New Mexico, June 16-20, 2007.
10. Toepel, J. and L.A. Sherman. 2007. "Metabolic regulation of the cyanobacterium *Cyanothece* sp ATCC 51142 when grown in light-dark or continuous light." IV European Phycological Conference, Oviedo, Spain, July 22-26, 2007.

Ecophysiological Investigation of Cyanobacteria Using Controlled Cultivation

Johannes Scholten, Tom Metz, Eric Hill, and Yuri Gorby

This project targets the development of a two-stage photobiological hydrogen production process (also know as "biophotolysis") from water and sunlight using the cyanobacterium *Cyanothece* 51142 (Figure 8). We will initially focus on the first stage of the process, the production of biomass with a high content of carbohydrates (glycogen) under non-limiting nitrogen conditions, which is the fundamental requirement of the process. Once the maximum photosynthetic capacity is known we will focus on the second stage of the process, the production of H_2 from glycogen during O_2-limited dark fermentation. We are using highly monitored and controlled batch cultivation technology, developed in the laboratory, combined with advanced analytical technologies to optimize both processes. This capability not only will provide a more complete understanding of the physiological response of cells to changes in environmental conditions, it will provide high quality, reproducible samples that will optimize biological information obtained from advanced analytical technologies. We will compare the metabolomics results from the continuous culture system with the light/dark cycling (LD) data collected from batch cultured cells. Eventually, we hope to use the genome-scale metabolic model for *Cyanothece* 51142 from Rajeev Aurora to make predictions that will allow us to come up with a rigorous experimental design to optimize glycogen accumulation during photosynthesis and H_2-production during O_2-limited dark fermentation.

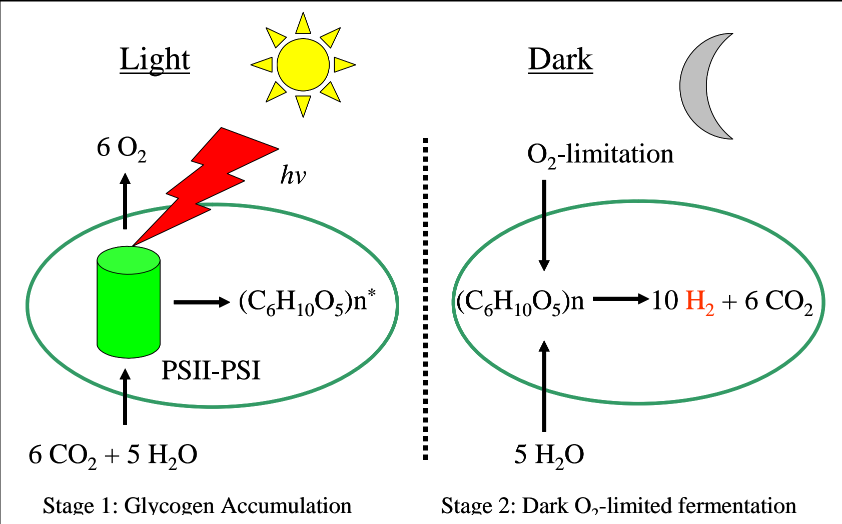


Figure 8. Indirect biophotolysis process with glycogen accumulation and dark fermentation stages.

Bioreactor Development for Continuous Cultivation

Efforts in the FY 2005 and FY 2006 resulted in the design of novel photo-bioreactors (Figure 9) with a high aspect ratio (tall and skinny) to optimize light penetration during growth of photosynthetic organisms. The photo-bioreactors developed at our laboratory were installed at Washington University in St Louis and Purdue University. Staff members at both universities were trained by our Microbial Cell Dynamics

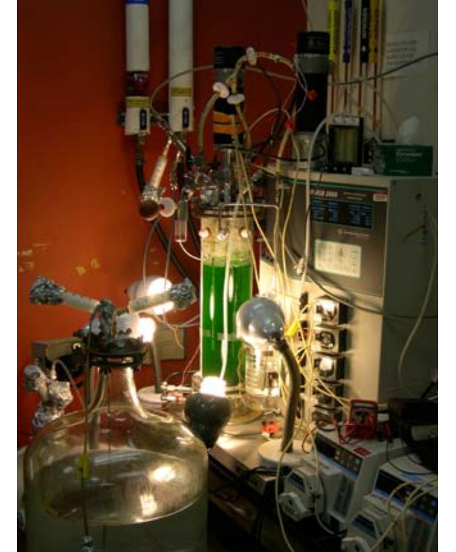


Figure 9. Photobioreactor developed at PNNL.

Laboratory personnel in operating the reactors. Furthermore we provided technical assistance to the university researchers to help them keep their reactors running properly. Other achievements we made in FY 2005 were: 1) purified antibiotic resistant mutants of *Synechocystis* 6803, 2) provided chemostat grown samples of *Synechocystis* 6803 to Galya Orr to optimize laser spectroscopy methods, 3) quantified the changes in pigment expression (chlorophyll vs. phycobilin) in *Synechocystis* 6803 during diurnal carbon dioxide fluctuations and their impact on the growth of *Synechocystis* 6803. As a result we proposed that all controlled growth experiments with cyanobacteria have to be done with a constant source of carbon dioxide. Controlling the carbon dioxide source will allow better interpretation of other diurnal cycle experiments, and 5) operated the first photovoltaic fuel cell to produce electricity from *Synechocystis* 6803.

Regulation of Cell Processes during Continuous Cultivation

In FY 2007, *Cyanothece* 51142 was grown in a bioreactor under continuous cultivation in the light (LL) and dark (DD). Before these experiments were started the ASP2 medium was analyzed and tested to see if it needed adjustment for the use in continuous cultivation. The following adjustments resulted in a better growth rate and yield: 1) increasing the phosphate concentration to 2 mM and 2) lowering the calcium concentration to 0.5 mM. For the LL cultivations the carbon source was CO_2 (400 ppm) and the light intensity (45 μEinsteins m^{-2} sec^{-1}). The DD cultivation was performed with glycerol (10 mM) as the carbon- and energy source, and O_2 as the electron acceptor (20% dissolved oxygen). Nitrate was added as an N-source for since N_2 can not be fixed by the organism during aerobic conditions. A number of growth rates were tested to determine the maximum growth rate of the organism under LL and DD conditions. The estimated maximum growth rate was for LL conditions, 0.021 h^{-1}, and for DD conditions 0.0054-0.0057 h^{-1}. Our results suggest that the *Cyanothece* 51142 cells became limited for CO_2 during our light cultivations but not for light. During the cultivation cell samples were taken for metabolomics analysis. Also samples were taken to determine optical density, dry weight, CO_2 consumption and O_2 production. These parameters will be used to establish the specific consumption/ production rates (qCO_2, $qO_{2qbiomass}$, etc.) and macrochemical balance. Initially global metabolomics has been done so far only on the LL samples. The LL samples were compared with 1) samples obtained during different growth rates (LL versus LL) and 2) pooled LD samples obtained from the diurnal rhythms experiment performed by the Himadri group. It seems that in both cases there are observable differences in the LC-MS chromatograms suggesting that growth rate (LL versus LL) and cultivation conditions (LL versus LD) may result in different metabolite profiles or even a whole metabolome (Figure 10). Currently more metabolomic measurements are done to confirm our initial results.

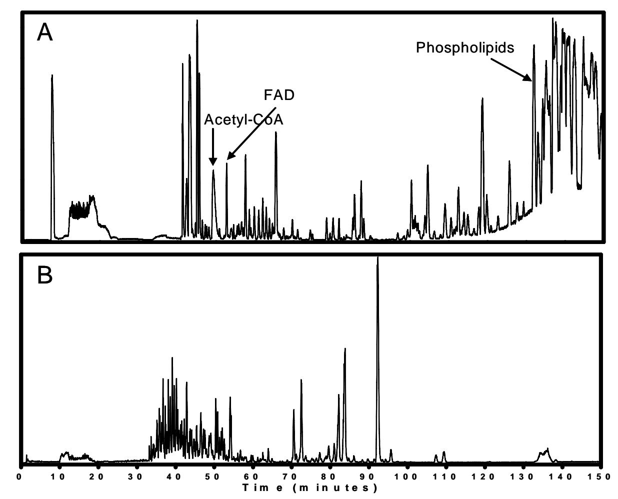


Figure 10. Comparison of global metabolomic profiles between continuous light (A) and light/dark (B) cultivation conditions.

Regulation of Cell Division during Continuous Cultivation

Cyanothece 51142 undergoes a complex circadian rhythm, while carrying out two antagonistic pathways: photosynthesis and nitrogen fixation. In FY 2008, we will develop and perform a stable isotope-labeling experiment that will label *Cyanothece* 51142 cells cultured in the presence of ^{13}C-labeled CO_2 as the sole carbon source during photosynthesis and ^{15}N-labeled N_2 as the sole nitrogen source during nitrogen fixation. The overall objective of this project is to cultivate isotopically labeled cells under controlled and monitored cultivation conditions. These cells will be used for microscopic, proteomic and metabolomics analyses to identify and characterize dynamic changes in morphology, protein expression and metabolite production that are important to comprehend the mechanisms involved in the circadian rhythm of *Cyanothece*. Microscopic analyses will be performed by Alice Dohnalkova and Galya Orr. The proteomic and metabolomics measurements will be carried out by Jon Jacobs, Garry Buchko, and Tom Metz, and data analysis will be performed by Chris Oehmen.

Publications, Presentations, and Awards

1. Presentation given at Washington University during the EMSL MBGC workshop, St. Louis, Missouri on April 12-13, 2007, entitled "Continuous light and dark metabolism of Cyanothece strain ATCC 51142."
2. Presentation given (Invited Speaker) at Joint Genomics: GTL Contractor-Grantee Workshop V and Inter-Agency Conference on Metabolic Engineering 2007, Bethesda, Maryland on February 12, 2007, entitled "Controlled Cultivation in System Biology."

Morphological, Functional and Redox Studies of *Synechocystis* 6803 and *Cyanothece* 51142 Bacterial Membrane Complexes by Methods of Electron Microscopy

Alice Dohnalkova, William Harvey, Eric A. Hill, and Christina Bilskis

Understanding the biogenesis of the photosynthetic membranes based on reactions that occur in the cyanobacteria during the circadian cycle would have, besides great scientific importance, huge implications on development of alternative solar energy technologies, and engineering technologies directly addressing environmental concerns based on carbon storage and nitrogen fixation as novel and economical means of repairing our environment. We are using novel methods of electron microscopy such as TEM tomography and cryostage as a tool for investigations of morphological and immunocytochemical dynamic changes that occur in *Cyanothece* during the circadian rhythm. These newly developed state-of-the-art capabilities and expertise will expand the W.R. Wiley Environmental Molecular Sciences Laboratory's (EMSL) capacity to provide support to user's programmatic work in accordance with the U.S. Department of Energy (DOE) mission.

One of the main Membrane Biology Grand Challenge (MBGC) goals is to assemble a blueprint of reactions that occur in the organism during a circadian cycle involving photosynthesis, respiration, carbon storage, and nitrogen fixation. In order to interpret functional information obtained by genome, proteome and metabolome studies, a thorough understanding of the structure and dynamics of the cell is crucial. Although cyanobacteria have been widely studied in the past, a comprehensive model of the cell morphology is still lacking. We apply novel methods based on

high-resolution electron microscopy such as three-dimensional reconstruction of images acquired by electron tomography, and morphological studies of frozen-hydrated cells observed at a TEM cryostage to address this concern. These newly acquired EMSL capabilities directly support research projects within the DOE programmatic scope, and broadly benefit the research community.

Membrane-Metal Interactions

In the first year we focused on characterization of Manganese transport and storage in *Synechocystis* at conditions simulating natural fluctuations in Manganese bioavailability and different redox states of Manganese substrates. Manganese was selected for its essential role as a catalytic element for photosystem II electron-proton coupling during water oxidation process where hydrogen is extracted from water and bound to a $Mn_4Ca_1Cl_x$ cluster. We identified a process of passive adsorption to cells outer membrane mediated by external polymeric substances (EPS), and metabolically mediated intracellular Manganese uptake associated with plasma and thylakoid membranes (see Figure 11). This work presents implication for better understanding of metal interactions with cyanobacteria in terms of metal precipitation, solubilization and cycling in the environment.

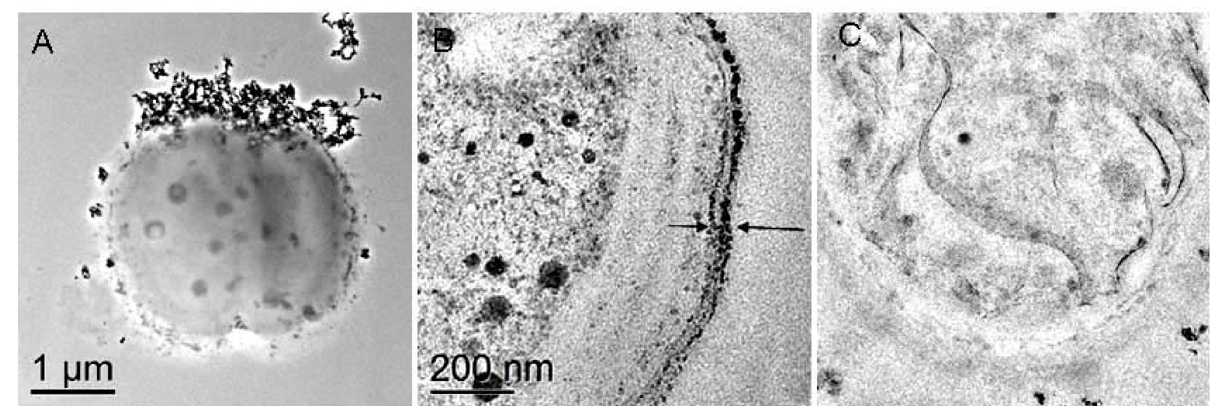


Figure 11. TEM images of *Synechocystis* exposed to 1mM $MnCl_2$ for 48 hours demonstrate massive Mn adsorption to the outer membrane associated with EPS (A), the internal Mn uptake by outer and cytoplasmic membrane (B), and a very thin Mn layer deposition along thylakoid membranes (C).

Imaging of Dynamic Membrane Structures

In 2006, EMSL acquired a devoted biological TEM with the high tilt stage allowing 3D reconstruction by TEM tomography, and a cryostage for observation of frozen-hydrated samples in their utmost natural state. By leveraging EMSL capability development funding, we were able to develop a new state-of-the-art capability using programmatic work of MBGC. Our two areas of interest were the 3D architecture and integrity of thylakoid membranes system in *Cyanothece*, and the spatial distribution of storage granules within a cell. In collaboration with the Sherman group and Michelle Liberton in Pakrasi's laboratory, we learned and established methodologies for reproductive sample preparation, and acquired archival samples during a 24-hour cycle for morphological changes analyses. By acquisition of tilt series of consecutive sections of a *Cyanothece* cell, we combined serial tomograms into one total volume of a cell, and by modeling in 3D visualized the organization of cellular components (see Figure 12). The TEM tomography brought literally a new dimension to our morphological studies, with interesting findings on *Cyanothece* cellular

architecture such as compartmentalization (using specialized membrane compartments for specific function such as glycogen granules production), with implications to its physiology. We will continue to follow cellular changes during the light and dark cycle, as well as to focus on sub-cellular dynamic features at the high resolution during the final year.

Correlation of Membrane Events with Cell Process Regulation

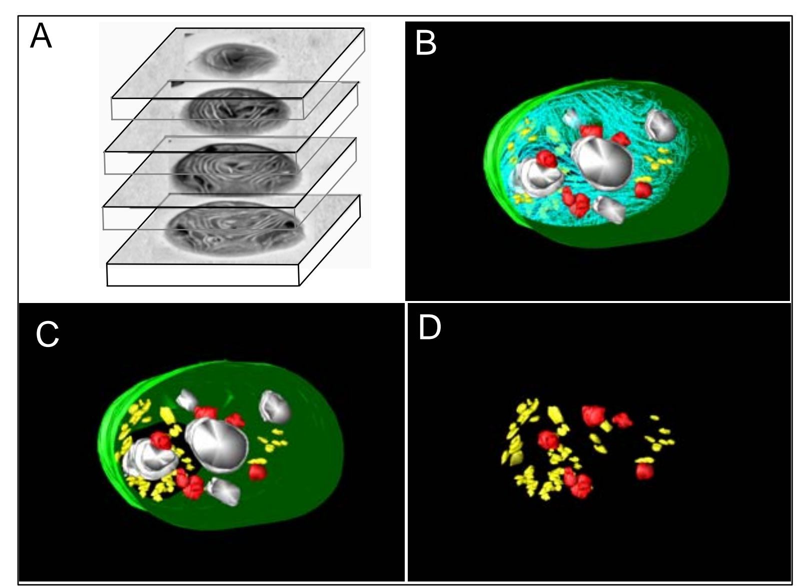


Figure 12. A scheme of alignment of consecutive tomograms for reconstruction of a whole cell (A). 3D model based on segmented features of interest provide a spatial idea of their localization with excellent spatial resolution. Green=outer membrane, blue=thylakoid membranes, grey= poly-phosphate granules, yellow=glycogen granules, red= carboxysomes.

In FY 2008, besides the 3D studies of dynamic morphological changes of selected cellular components, we would like to correlate oscillating cellular events with the proteomics data. We plan a correlative microscopy experiment based on immunocytochemical localization of CmpABC transporter assembly in *Cyanothece* plasma membrane during the cycle. CmpABC is a high-affinity bicarbonate transporter that plays an important role in intracellular carbon sequestration. We plan to conclude our research on cell division with implications on thylakoid membranes biogenesis (in collaboration with Lou Sherman group and PNNL's J Sholten).

Publications, Presentations, and Awards

1. Dohnalkova, A., C. Bilskis, and D.W. Kennedy. 2006. "TEM Study of manganese biosorption by cyanobacterium *Synechocystis* 6803." *Microsc. Microanal.* 12(Supp 2):444-5.
2. Dohnalkova, A., R. Mendoza, C. Gassman, and C. Bilskis. 2007. "Creating 3D reconstruction of cyanobacterium *Cyanothece* sp. by alignment of serial TEM tomograms from consecutive plastic sections." *Microsc. Microanal.* 13(Supp 2):1338-9.
3. Dohnalkova, A., C. Bilskis, and D.W. Kennedy. 2006. "TEM Study of manganese biosorption by cyanobaterium *Synechocystis* 6803." 64th Annual Meeting of Microscopy Society of America, Chicago, Illinois.
4. Dohnalkova, A., R. Mendoza, D. Panther, and G. Orr. 2006. "Morphological changes in cyanobacterium *Cyanothece* 51142 sp. during circadian rhythm visualized by electron tomography." The 4th International congress on Electron Tomography, San Diego, California.
5. Dohnalkova, A., W. Harvey, C. Gassman, E.A. Hill, and C. Bilskis. 2007. "Creating 3D reconstruction of cyanobacterium *Cyanothece* sp. by alignment of serial TEM tomograms from consecutive plastic sections." 65th Annual meeting of Microscopy Society of America, Ft Lauderdale, Florida.

6. Alice Dohnalkova received The Diatome US First Place award for "Creating 3D reconstruction of cyanobacterium *Cyanothece* sp. by alignment of serial TEM tomograms from consecutive plastic sections." The award was presented to Alice by the president of the Microscopy Society of America at the Annual Meeting in Ft. Lauderdale, Florida, August 2007.

Dynamic Changes in Molecular Expressions and Interactions along the Circadian Rhythm

Galya Orr and David Panther

Cyanobacteria undergo a complex circadian rhythm that must originate, in part, from dynamic changes in protein expressions, and in the assembly and disassembly of protein complexes and cellular structures along the cycle. The ultimate goal of this project is to elucidate mechanisms that underlie the unique ability of specific cyanobacteria to go through a complex circadian rhythm, while executing two antagonistic pathways: photosynthesis and nitrogen fixation. This ability must originate, at least in part, from the dynamic changes in the molecular interactions, and assembly and disassembly of complexes along the circadian rhythm. To achieve the goal, we have been developing and applying fluorescence techniques, including fluorescence resonance energy transfer (FRET), to detect dynamic changes in protein expression and report on molecular interactions and protein complex formation in the intact cyanobacterial cell.

Our effort has been focused on the high affinity CmpABC bicarbonate transporter, which belongs to the bacterial ATP Binding Cassette transporter family, and plays an important role in the cyanobacterial CO_2 concentrating mechanism (CCM). The transporter is thought to be assembled in the cytoplasmic membrane from four different subunits (from Koropatkin, et al., 2007). Carbon uptake is minimal in the dark and is significantly increased in the light when photosynthesis and carbon fixation take place. The cyclic uptake of carbon has been thought to result from the increase in the expression and/or activation of proteins in the CCM, including the CmpABC transporter. However, no information is available about the changes in the protein expression levels of each of the transporter's subunits and their assembly in the cytoplasmic membrane along the light-dark cycles. While an increase in the number of transcripts or expression of some of the subunits has been reported under low carbon concentration or intense light, no information about these processes along the circadian rhythm is available.

Imaging of Dynamic Membrane Protein Expression

By developing quantitative fluorescence techniques and reagents specific to *Cyanothece* 51142 we have been able to investigate dynamic changes in the expression patterns of the CmpABC transporter's subunits and their assembly in the cytoplasmic membrane along the circadian rhythm. Using the antibody against CmpA, the method for introducing antibodies into the periplasm, and the isolation of two different fluorophores for multi-color fluorescence imaging in cyanobacteria have been established, as demonstrated in Figure 13. The simultaneous emissions of CmpA antibodies tagged either with Alexa 488 or Alexa 546 are captured and clearly isolated from the emission of the endogenous phycobilisome or chlorophyll. The established immuno-fluorescence and multi-channel imaging allows us to investigate the spatial and temporal expression patterns of several proteins relative to each other, and detect the formation of protein complex and molecular interaction dynamics using FRET.

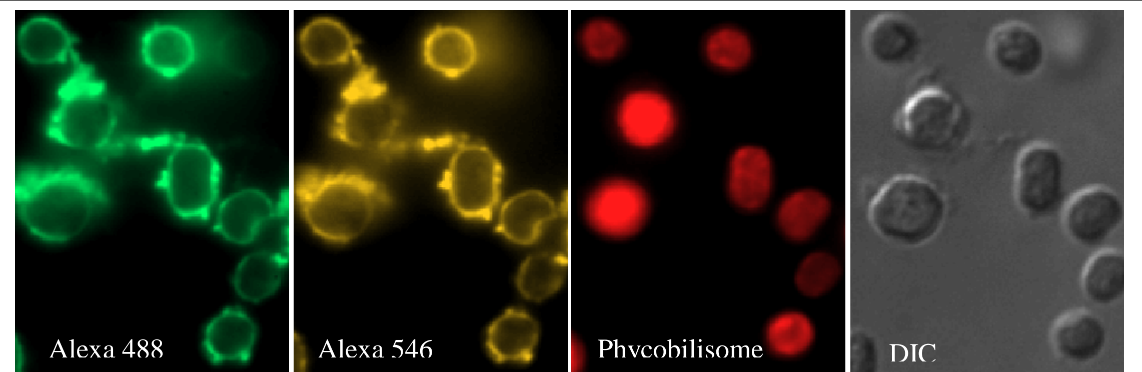


Figure 13. Multi-channel, immuno-fluorescence imaging for investigating spatial and temporal expression patterns of CmpA in intact *Cyanothece*.

Tagging each antibody (or its respective Fab fragment) with 1-2 dye molecules enables the estimation of the number of subunits within a fluorescent spot using single-molecule fluorescence techniques. The series of images in Figure 14 was taken from a cell stained with the antibody against CmpA. By quantifying the fluorescence intensity that is expected from one antibody, we have been estimating the number of subunits within each fluorescent spot. This information, taken at different time points along the circadian rhythm, allows us to determine changes in the cytoplasmic membrane expression of the protein along the cycle.

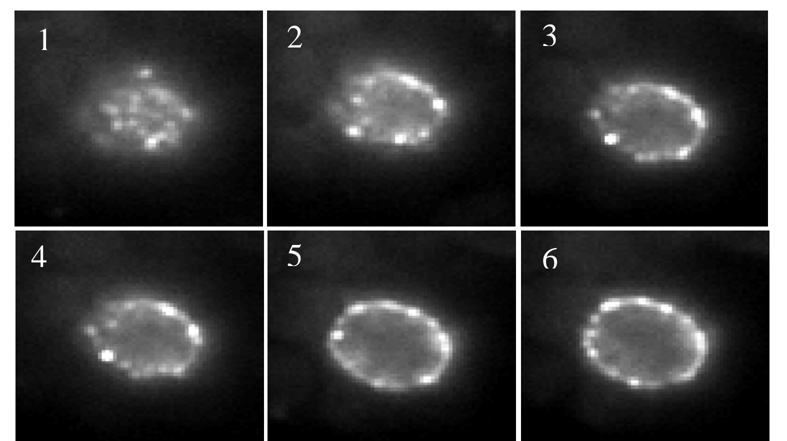


Figure 14. Detecting and quantifying CmpA expression and subunit insertion in the cytoplasmic membrane of the intact *Cyanothece* at different time points of the diurnal cycle (1-6).

Guided by the annotated genome from Dr. Pakrasi's lab, we have identified two amino-acid sequences, corresponding to the periplasmic domain of the CmpB subunit, which were used to raise a specific polyclonal antibody against the subunit. Together with the antibody against the periplasmic domain of CmpA, the new antibody has enabled the detection and quantification of fluorescence co-localization as shown in Figure 15. The cell in A, which was taken at L1, shows clusters of co-localized CmpA and CmpB (yellow), in addition to unpaired CmpA subunits (green). The cell in B, which was taken at L4, shows many more co-localized clusters (yellow), with few unpaired CmpB subunits (red). These observations indicate a dynamic and differential change in the expression pattern of the two subunits and their assembly along the cycle. Clusters of co-localized CmpA and CmpB (yellow), in addition to unpaired CmpA subunits (green). The cell in B, which was taken at L4, shows many more co-localized clusters (yellow), with few unpaired CmpB subunits (red). These observations indicate a dynamic and differential change in the expression pattern of the two subunits and their assembly along the cycle.

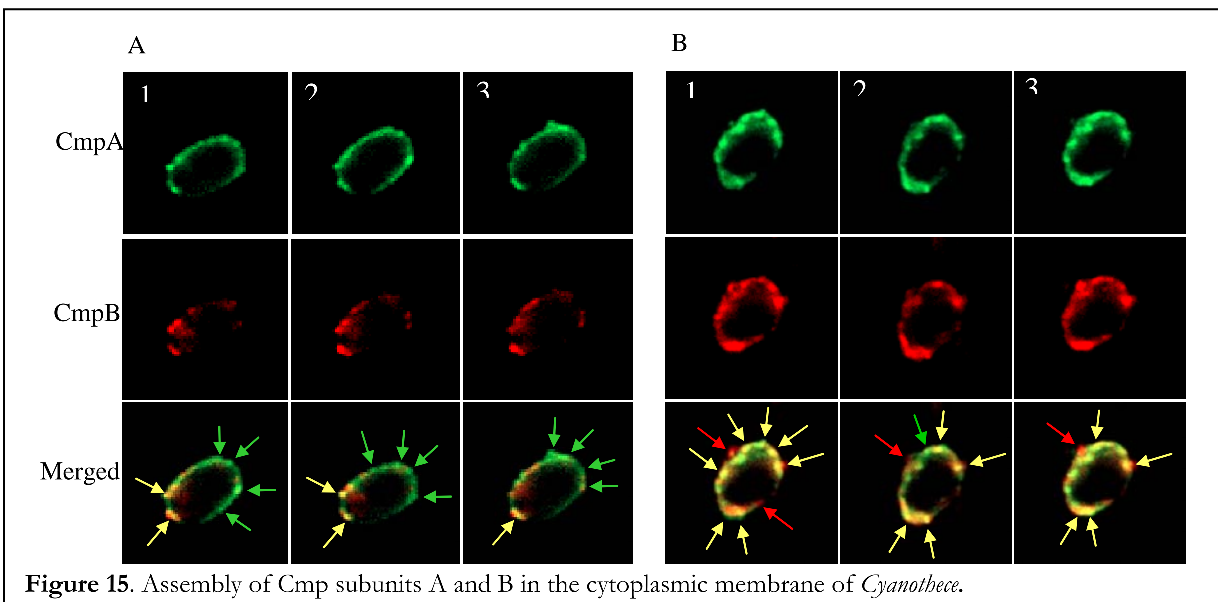


Figure 15. Assembly of Cmp subunits A and B in the cytoplasmic membrane of *Cyanothece*.

Imaging of Dynamic Membrane Structures during Diurnal Cycling

Using the FACSAria to measure changes in forward and side scatters every two hours during 48 hours sessions, we were able to identify rhythmic changes in *Cyanothece* cell size and complexity, respectively. By screening ~50,000 cells at each time point along the circadian rhythm, we have provided a statistically powerful data set that confirmed and strengthened parallel observations made at Washington University using microscopy. We found that cell size increased significantly during the light period, reaching maximum around 4-6 hours in the light and minimum at the beginning of the dark period (Figure 16-left) as measured in fixed cells. The time frame was slightly different when measured in live cells, where maximal and minimal sizes were reached about two hours later (Figure 16-right). Rhythmic changes have been also observed in cell complexity using side scatter (SSC) values (data not shown). The above observations support parallel studies that have been done at PNNL and at Georgetown University using different approaches.

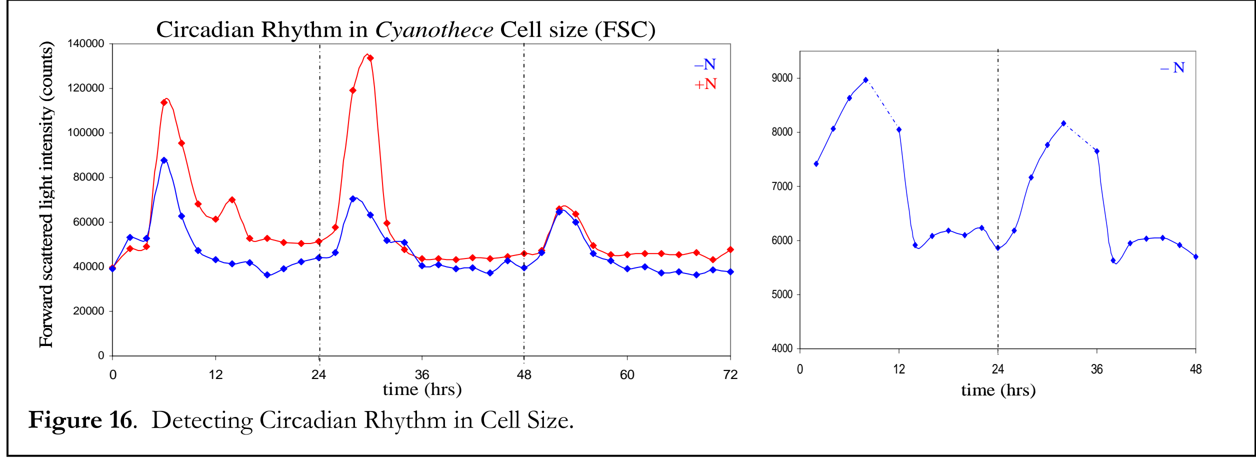


Figure 16. Detecting Circadian Rhythm in Cell Size.

Using the FACSAria, we have also quantified the changes in the phycobilisome and chlorophyll fluorescence within individual cells. We found dynamic changes in fluorescence intensities along the circadian rhythm, with distinct sub-populations of cells showing different degrees of fluorescence

intensities. Rhythmic changes in phycobilisome emission are observed along the cycle, using the FACSAria at 633 nm laser excitation and 660/20 nm emission (Figure 17). A consistent pattern is observed, showing one peak during the light and another, smaller, peak during the dark periods. A similar approach was used to quantify the expression of chlorophyll along the cycle, which reflected one peak during the light cycle but no peak during the dark cycle (data not shown).

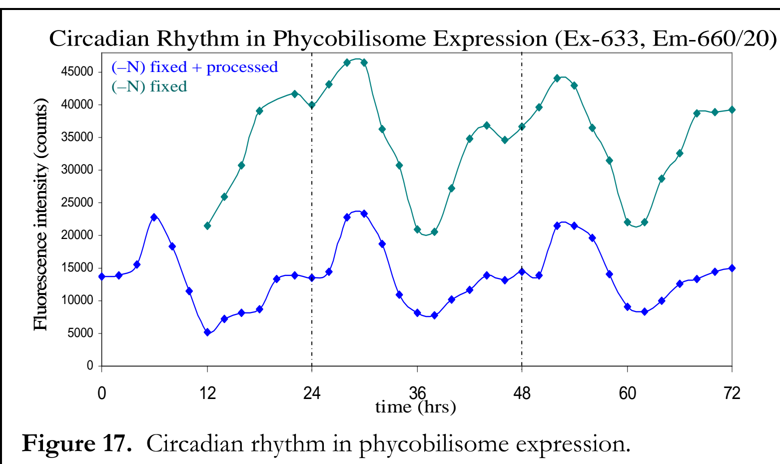


Figure 17. Circadian rhythm in phycobilisome expression.

Future directions

Using the established immuno-fluorescence and imaging techniques we will continue to characterize and quantify the changes in the protein expression and membrane insertion of the two subunits and their assembly/disassembly along the cycle.

We are producing Fab fragments, which are smaller than the whole immunoglobulin and therefore are more likely to support FRET. Using the Fab fragments, we will apply quantitative FRET imaging to detect and quantify molecular interactions between the two subunits along the cycle. While fluorescence co-localization suggests the assembly of the subunits within a cluster, FRET can unambiguously report on their molecular interactions.

Additional antibodies specific for the transporter subunits C and D will be developed and the method for their introduction into the cytoplasm will be developed. Using these antibodies we will investigate the expression pattern of these subunits and their assembly with subunits A and B along the cycle to gain a complete understanding of the cyclic formation/function of the transporter.

Publications, Presentations, and Awards

1. Dohnalkova, A., R Mendoza, D.J. Panther, and G. Orr. 2006. "Morphological Changes in Cyanobacterium Cyanothece 51142 during the Circadian Rhythm Visualized by Electron Tomography." Presented at the 4th International Congress on Electron Tomography, San Diego, California on November 5, 2006.
2. Oehmen, C.S., J.E. McDermott, J. Stockel, E.A. Welsh, G.W. Buchko, J.M. Jacobs, T. Metz, A. Dohnalkova, G. Orr, E.S. Peterson, and H.B. Pakrasi. 2007. "Systems approach to understanding molecular mechanisms in the light-dark cycle of Cyanothece." Presented at the 9th Cyanobacteria Molecular Biology Workshop, Delevan, Wisconsin on July 7, 2007.
3. Orr, G., D. Panther, A. Dohnalkova, J. Jacobs, M. Liberton, L. Sherman, H. Pakrasi. "Circadian Rhythm in the assembly of the CmpABC transporter in the cytoplasmic membrane of Cyanothece 51142." Manuscript in preparation.

Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function: Proteomic, Metabolomic, and Metallomic Characterization

Jon M. Jacobs, Thomas O. Metz, Jie Ding, Christina M. Sorensen, Qibin Zhang, James A. Campbell, George Hager, Charles Barinaga, David G. Camp II, David W. Koppenaal, and Richard D. Smith

The annotation of the genome and comprehensive analysis of metabolic pathways in *Cyanothece* can provide tremendous information about how this photosynthetic organism temporally regulates nitrogen fixation and carbon fixation at opposite points of a diurnal cycle. This project is directed towards developing approaches and methods for the rapid detection of large numbers of peptides and metabolites in an extremely broad manner that is highly sensitive (e.g., able to detect low-level species), quantitative, and is facilitated by high-throughput measurements. In addition, we will develop procedures and techniques for global metallomic characterization of *Synechocystis* 6803 and *Cyanothece* 51142 organisms. In this project, we focus on developing methods based upon the use of high-resolution capillary liquid chromatography combined with either Fourier transform ion cyclotron resonance (FTICR) or high mass accuracy time-of flight (TOF) mass spectrometry (MS) to increase the comprehensiveness, quantitation and throughput of the proteomics and metabolomics analyses, and leverage the significant separations and mass spectrometry capabilities originally developed at PNNL in the Biological Systems Analysis and Mass Spectrometry group.

Proteomics

The ability to quantitatively measure differential protein abundances is essential to elucidate cellular processes. When using comparative proteomic studies to determine differences in protein abundances between two different conditions, the abundance level of the protein in each state is not a static value, but represents a dynamic protein level dependent upon a variety of multiple mechanisms. This can be described as a steady-state level of protein abundance which is not apparent when there is only one time point or when only comparing two different conditions or states. In the simplest terms, this protein steady state can be described as a function between the rates of protein synthesis competing with the rate of protein degradation, but other regulatory factors can dramatically affect this steady state including the spatial localization and transportation of the protein. The use of a pulse-stable isotopic labeling approach can allow for the calculation of multiple protein turnover rates based upon the quantitative observation of mass spectrometry (MS) peak pairs. These MS pairs correlate with the combination of peptides from newly synthesized proteins incorporating unlabeled amino acid residues during an unlabeled chase period with the previously heavy labeled peptides from original proteins incorporating the heavy labeled residues. The intensity values which can be extrapolated from the identified pairs are then used to calculate the turnover rate of a number of various proteins. The basis of this approach can be applied to any number of culture conditions for the quantitative elucidation of multiple cascade and localization effects.

In applying this approach globally for the study of the membrane biology of *Cyanothece*, the overall objective of the proteomic component of this project is to develop quantitative approaches to study the dynamic abundances of the membrane proteome of *Cyanothece*. Specifically, this will be achieved by a time course based stable isotopic labeling approach as described above, providing a unique level of protein turnover information that will result in a new quantitative capability to support other

projects within this imitative as well as be applicable to most any cell culture based proteomic analysis in the future.

Proteomics-assisted Genome Annotation

Virtually no proteomic analyses had been preformed on this organism prior to this project, so critical initial studies were needed to globally identify and characterize the main proteome response of the diurnal cycle. Additionally, the genome of *Cyanothece* sp ATCC 51142 was simultaneously being sequenced, which provided a unique opportunity to use proteomic data to help guide annotation efforts. Initial studies involved the Pakrasi lab specially growing *Cyanothece* samples in a time course based manner throughout a 48-hr diurnal cycle. Membrane and cytosolic proteins of each time point were isolated and tryptically digested, after which the samples were subjected to capillary liquid chromatography (LC) combined with tandem MS (MS/MS) analysis to initially identify peptides and create a mass and time tag database. To date, we have now identified over 27,000 unique peptides from over 2700 proteins, representing ~50% of the currently annotated genome.

Due to the timeliness of the genome annotation events and using the proteomic data as a guide, an additional 53 proteins with multiple peptide identification were included into the annotated database that would have otherwise been removed during curation and annotation. Additionally, 457 out of 1930 annotated hypothetical proteins (24%) were reannotated as "unknown" due to the detection of multiple peptide sequences. Furthermore, just by using limited quantitative information provided by the LC-MS/MS analysis, protein abundance cycling patterns were clearly observed in the 48-hr time course data; (see Figure 18), which highly correlated with previous transcriptomic measurements. We have now repeated this experiment using more precise ^{18}O labeling based approaches and these data are now being incorporated as a key component into multiple systems biology based correlations and are being integrated into a key initiative-wide manuscript for publication.

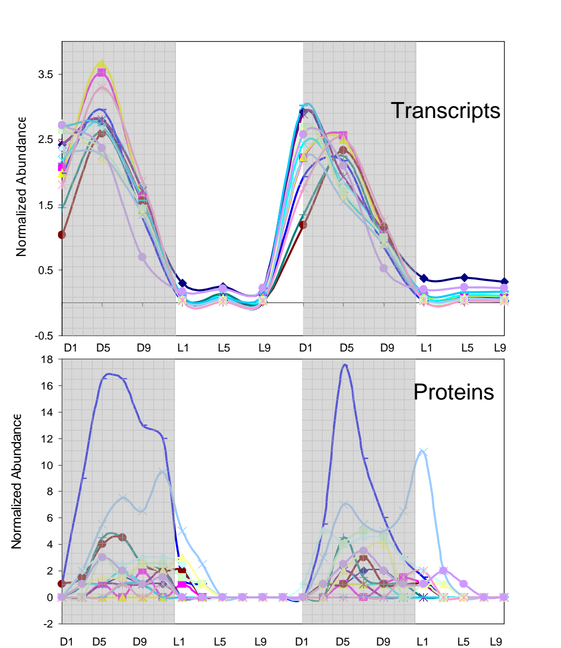


Figure 18. Using a spectral count approach, the protein abundance cycling characteristics of nitrogenases are clearly observed, which highly correlate with transcriptomic analyses.

Regulation of Cell Processes during Diurnal Cycling

To provide more dynamic quantitative detail to the protein abundance patterns observed in the initial data, we began to develop a protein turnover experimental system in which we can distinguish newly synthesized proteins within a certain timeframe as well as potentially assign a rate of synthesis for each protein. Initial studies are now complete, and the results have been quite promising, as we have been able to define the parameters needed for the experimental design and track the incorporation of the heavy labeled peptides in comparison to their light labeled components, see Figure 19. Looking at the ratios of heavy/light incorporation in Fig. 19B, one can clearly see

different rates of incorporation across all identified proteins. We are ramping up to implement this approach for a full diurnal cycle experiment and anticipate that a unique level of information will become available to help elucidate the membrane biology of *Cyanothece*. Other future studies in support of the proteomic efforts will include continuing to improve and augment our quantitative approaches, particularly through informatics developments as these have appeared to be the most challenging to address and overcome, i.e. refining the algorithms for appropriate detection and extraction of the quantitative data from the identified heavy/light labeled pairs.

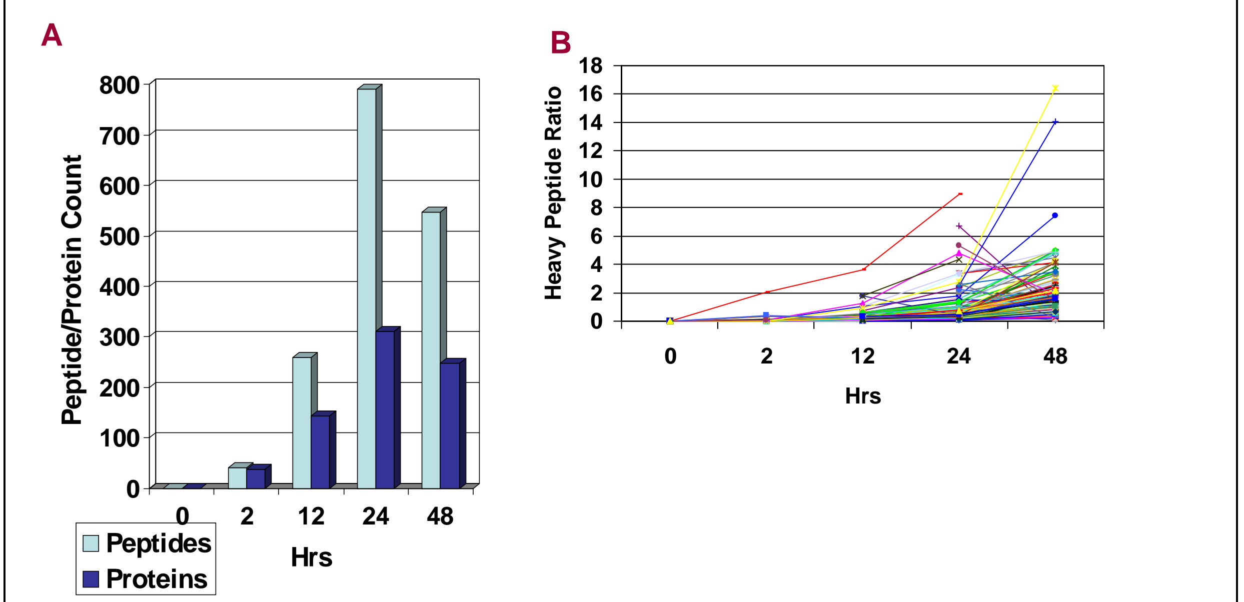


Figure 19. Initial demonstration of a metabolic-based, stable isotopic label incorporation using a heavy labeled Leu residue introduced into the culture media at time point 0 followed by steady state constant light (non-cycling) growth conditions. A) Total number of heavy/light labeled peptides/protein pairs identified across the timeframe. B) Incorporation ratio (heavy/light) for the 200 top proteins across the same timeframe.

Metabolomics

The metabolomics component of this project endeavors to advance the understanding of those metabolites whose concentrations are under circadian and/or light/dark control during diurnal growth of *Cyanothece*, including thylakoid and plasma membrane lipids and those metabolites involved in the glycogen synthesis pathway. This project is directed towards developing approaches and methods for the rapid detection of large numbers of metabolites in an extremely broad manner that is highly sensitive, quantitative, and is facilitated by high-throughput measurements. In addition, targeted metabolite analyses of intermediates in glycogen synthesis will be employed in the final year of this project. All of these approaches will utilize high-efficiency capillary liquid chromatography (LC) coupled with high-mass resolution mass spectrometry (MS).

Regulation of Cell Processes during Diurnal Cycle

We have initially demonstrated the performance of coupled high pressure capillary LC with FTICR analysis of the *Shewanella oneidensis* metabolome. We have leveraged this initial work and evaluated several LC stationary and mobile phases, in order to identify the best combination of both for analyzing water-soluble metabolites by LC-MS. These efforts have resulted in a robust and sensitive

capillary LC-MS method for the analysis of water-soluble metabolites extracted from *Cyanothece* cell lysates (Figure 20).

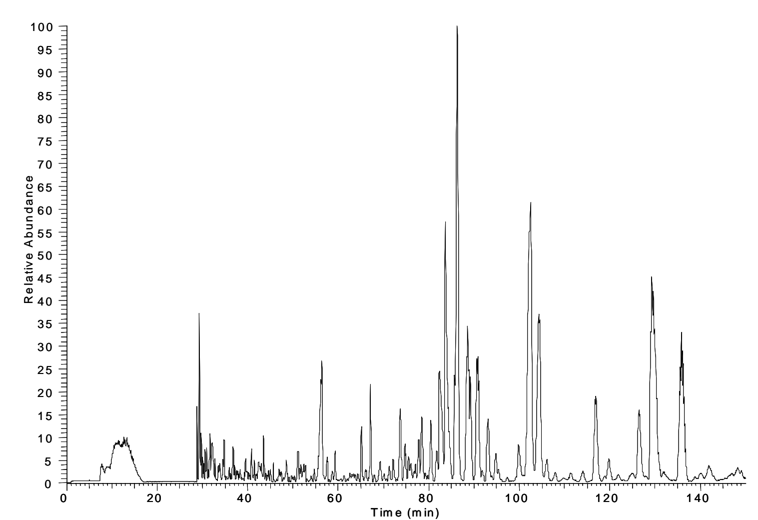


Figure 20. Reversed-phase LC-MS analysis of *Cyanothece* metabolome extracted with isopropanol. Approximately 1,700 metabolite features were identified in this 150 minute separation. LTQ-Orbitrap MS was used for detection (*m/z* 100-1,000).

In order to develop a high throughput approach for metabolite quantitation, our goal is to generally circumvent the extensive use of unlabeled external standards or isotopically labeled internal standards. Instead, we have developed and continue to refine novel software (MultiAlign) for aligning metabolite features observed across multiple datasets, followed by intensity normalization based on the expectation maximization (EM) algorithm. Briefly, this algorithm analyzes the histogram of log ratios of intensities of features common to two or more datasets and finds the peak apex of this distribution by assuming that the histogram is a mixture of a normal density corresponding to unchanged features and uniform density background corresponding to changed features. The EM algorithm is used for calculating the normal part and uniform part of the histogram and the shift in intensity is applied to all features in the aligned dataset. Graphs are produced showing the alignment of the alignee(s) to the baseline and the log ratio intensity histogram of the count of features present in the alignee versus the log of alignee(s)/baseline; the normalized output (cluster #, mass, normalized elution time, and intensity) for the aligned data can then be exported in tab-delimited format for subsequent data processing in programs such as Excel or MatLab. Figure 21 and 22 show the alignment and normalization, respectively, of technical replicates of a *Cyanothece* metabolite extract analyzed by capillary LC-MS.

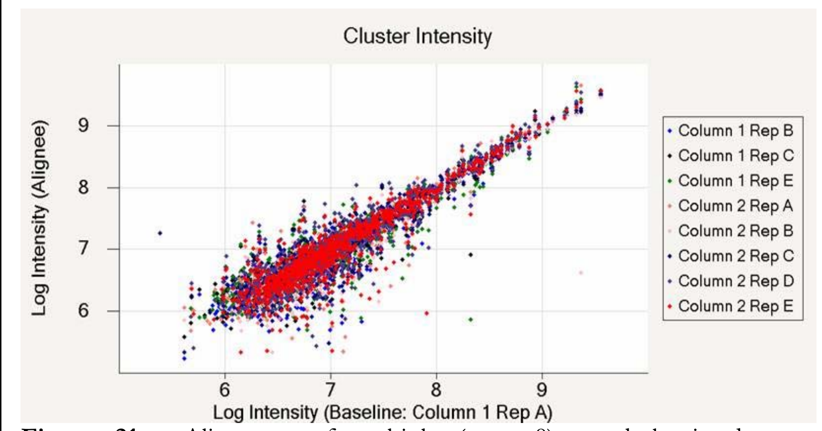


Figure 21. Alignment of multiple (n = 9) metabolomic datasets corresponding to technical replicates of the same *Cyanothece* metabolite extract analyzed on two columns of a dual-column capillary LC system coupled with MS. A single dataset was arbitrarily selected as a baseline for alignment purposes.

LC-MS metabolomics analyses of a 48-hr *Cyanothece* time course experiment were made in order to identify any metabolite features (defined by a measured mass and elution time) demonstrating interesting behavior (i.e. cyclic abundance pattern, unique to light only, unique to dark only, etc.). Features reproducibly observed in multiple analyses of time course samples were grouped by single linkage clustering in two dimensions, mass and normalized elution time (NET), based on user-defined options. For these data, mass and NET tolerances of ± 3 ppm and ± 0.02 NET (representing ± 3 min of a 150-min LC separation) were selected based on empirical observation of mass spectrometer and LC system performance.

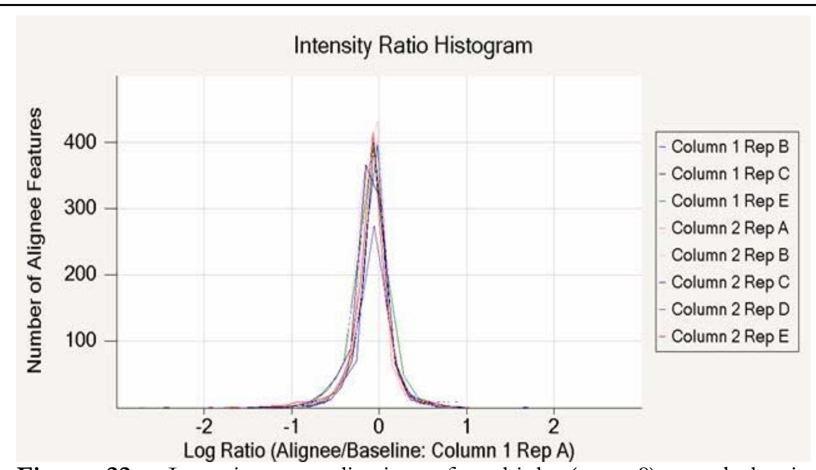


Figure 22. Intensity normalization of multiple (n = 9) metabolomic datasets corresponding to technical replicates of the same *Cyanothece* metabolite extract analyzed on two columns of a dual-column capillary LC system coupled with MS. A single dataset was arbitrarily selected as a baseline for normalization purposes.

After chromatographic alignment, intensity normalization was applied using the EM algorithm, and metabolomic features were then examined for cyclic abundance behavior throughout the time course experiment, or for uniqueness to the dark or light time points. We initially selected 20 metabolomic feature candidates for targeted MS/MS experiments and subsequent structural elucidation based on their interesting cyclic abundance patterns (Figure 23). However, these features were in relatively low abundance in the time course samples, and targeted MS/MS experiments on high mass accuracy instruments did not produce sufficiently abundant fragment ions for structural elucidation. Because of this observation, we will be conducting more targeted metabolite profiling analyses during the final year of this project.

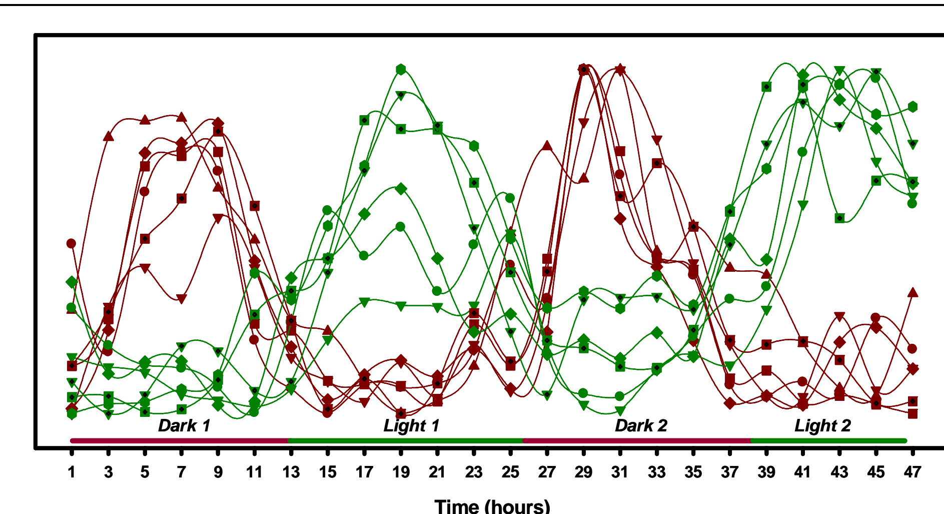


Figure 23. Subset of *Cyanothece* metabolites displaying cyclic abundance patterns during 48-hr incubation in light/dark. Metabolites with abundance maxima during dark hours are shown in brown, whereas metabolites with abundance maxima during light hours are shown in green.

In the final year of this project (FY 2008), we will perform the following tasks:

- Structural elucidation of metabolite features displaying circadian control utilizing *Cyanothece* grown in the presence of $^{13}CO_2$ and $^{15}N_2$. Cells will be lysed and fractionated using LC, followed by heteronuclear NMR analyses of collected LC fractions
- Characterization of the lipid composition of *Cyanothece* thylakoid and plasma membranes during diurnal growth conditions.
- Enrichment and analysis of phosphorylated *Cyanothece* metabolites, particularly those involved in glycogen synthesis.

Metallomics

The metallomics component of this project is directed at developing specialized instrumentation and techniques for comprehensive bio-metal characterization of the *Cyanothece* organism, including determination of metal cycling behavior and the characterization of metallo-proteins.

A wide variety of metal ions are being recognized as intimate players in biological activity, along with growing realization that metals in biological systems do not act alone but rather in concert with other available metals. This recognition has led to adoption of the term metallomics, akin to genomics, proteomics, and metabolomics. Metallomics is the study of the complete complement of metals in a biological system, including free metals, complexed metals, metabolite-bound metals, and protein-bound metals. Determination of minor to trace levels of metals in biological systems will require new approaches to metal analysis, as the concentration range of different metal and metal species in a biological system can span more than 12 orders of magnitude. Inductively coupled plasma mass spectrometry (ICPMS) has demonstrated significant capability in metals analysis, including near-complete metals coverage and a wide (10^8-10^9) dynamic range of detection and quantitation.

Metallomic Instrumentation Development

Given the inherent complexity of the proteome and the likely need to determine metals in mixed whole-protein fractions, a need exists for higher performance elemental analysis instruments. A portion of this work was therefore directed at development of a unique, high-resolution inductively coupled plasma mass spectrometry system. Initial experiments are described in which an inductively coupled plasma (ICP) atomic ion source is interfaced to a high resolution, ion trap mass spectrometer which normally samples electrospray ion streams.

While reaction/collision cell ICPMS technology has been shown to remove many conventional (polyatomic) ion interferences, many isotopic isobars remain that will require significant mass resolution to separate, especially in those arising in a more complex protein matrix. Recently a new and operationally simple type of mass spectrometer, the Orbitrap analyzer, was developed that can deliver high mass resolution (R>200,000). This instrument has produced significant sensitivity, resolution and throughput capability enhancements for proteome and metabolome experiments. By combining the robust ICP ion source with this high mass resolution spectrometer, it is hoped that similar benefits can be obtained for atomic MS applications in metallomics. A combined ICP/ESI-linear ion trap (LIT)-Orbitrap MS instrument configuration is desired. This instrument configuration will allow switching of ESI and ICP sources for molecular or atomic ion analysis, enabling metal fingerprinting of proteins by standardizing and comparing elution time profiles.

Work during the year has progressed in designing and building this instrument. A photograph showing the interfacing of an ICP ion source to the LIT-Orbitrap instrument is given in Figure 24. This effort turned out to be a non-trivial exercise as the substitution of a very hot (~5000 °K) ICP ion source for the conventional ESI ion source on the Orbitrap MS instrument created a number of interfacing issues. Work continues on improving the exact interface design to efficiently transfer atomic ions to the mass spectrometer. An excellent indication of success has been achieved, however, in this as indicated in Figure 25, which illustrates a very high-resolution mass spectrum of atomic Rb. The resolution shown here, of better than 240,000 (m/dm), has been unachievable in any other ICPMS instrument to date. Achieving this resolution routinely for metallomic applications will enable more comprehensive and complete characterization of the metallome in *Cyanothece* samples.

Figure 24. Photograph of new ICP-Orbitrap MS.

Metallomic Analyses

Additional metallomic studies have also been carried out using conventional plasma source emission and mass spectrometry methods to examine metal concentration distributions and cycling behavior, and to characterize metal contents of proteins isolated for structural determinations. An example of the latter effort is the determination of metal content in support of the structural determination of a bicarbonate binding protein (CmpA) present in cyanobacteria. The structure determination was led by the Danforth Plant Sciences Center group and is described in their report. As the structure determination was being done, questions arose relative to the presence of a metal in the structure. Analyses of a CmpA protein sample, using ICP-OES and ICPMS, showed the presence of major and stoichiometric amounts of Ca. The presence of the Ca has a noted effect on the structure, and the binding of the Ca ion appears to be tied to bicarbonate ion binding. Minor amounts of other

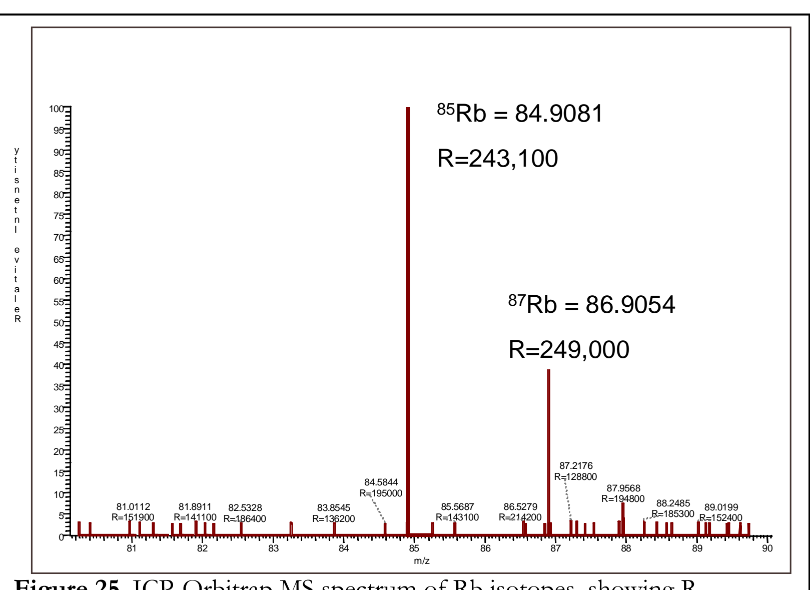


Figure 25. ICP-Orbitrap MS spectrum of Rb isotopes, showing R >240,000 (unparalleled in atomic MS).

metals, including Fe and Zn were also found but these were determined to only be concomitant elements. A publication on the structure determination was produced.

Samples of *Cyanothece* soluble and membrane proteins were obtained from Washington University in St. Louis to evaluate inductively coupled plasma/mass spectrometry (ICP/MS) for the analysis of metals distribution. The results from these analyses indicated Ca, Fe, Co, Cu, Zn, Mo, and Pb concentrations at easily quantifiable levels using ICPMS. Accordingly, experiments and whole cell sampling, at 4-hour intervals over a 48-hour cultivation period, are being conducted for early FY 2008 investigations. A range of metals, to minimally include Fe, Mn, Ca, Zn, Mo, and Cu, will be determined for metal cycling behavior and correlation to proteomic/metabolomic and other results.

Publications, Presentations, and Awards

1. Hager, G.J., C.J. Barinaga, and D.W. Koppenaal. 2006. "Advanced Ion Trap ICPMS Techniques." Presented by Dave Koppenaal (Invited Speaker) at 2006 Asia Pacific Winter Conference, Bangkok, Thailand on November 30, 2006.
2. Koppenaal, D.W., C.J. Barinaga, J.A. Campbell, H.B. Pakrasi, and M.L. Liberton. 2007. "Metallomics - New Techniques for a New -Ohmics." Presented by David Koppenaal (Invited Speaker) at 39th Central Regional American Chemical Society Meeting, Covington, Kentucky on May 21, 2007.
3. Oehmen, C.S., J.E. McDermott, J. Stockel, E.A. Welsh, J.M. Jacobs, T. Metz, A. Dohnalkova, G.W. Buchko, and H.B. Pakrasi. 2007. "Systems approach to understanding molecular mechanisms in the light-dark cycle of *Cyanothece* sp. 51142." Presented by Christopher Oehmen at 9th Cyanobacteria Molecular Biology Workshop, Delavan, Wisconsin on June 6, 2007.
4. Welsh, E.A., M.L. Liberton, J. Stockel, H. Sato, T. Loh, C. Wang, A. Wollam, R.S. Fulton, S.W. Clifton, J.M. Jacobs, R. Aurora, L.A. Sherman, R.D. Smith, R.K. Wilson, and H.B. Pakrasi. 2007. "Proteomics Assisted Genome Sequence of Cyanothece 51142, a Unicellular N_2-fixing Cyanobacterium." Presented by Eric Welsh at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, Wisconsin on June 6, 2007.
5. Koropatkin, N., A.M. Randich, M. Bhattacharyya-Pakrasi, D.W. Koppenaal, H.B. Pakrasi, and T.J. Smith. 2007. "Specificity and Regulation of ABC Transporters." Presented by Nicole Koropatkin at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, Wisconsin on June 6, 2007.
6. Ding, J., C.M. Sorensen, Q. Zhang, H. Jiang, N. Jaitly, E.A. Livesay, Y. Shen, R.D. Smith, and T.O. Metz. 2007. "Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling." *Anal. Chem.* 79:6081-6093.
7. Metz, T.O., Q. Zhang, J.S. Page, Y. Shen, S.J. Callister, J.M. Jacobs, and R.D. Smith. 2007. "Future of liquid chromatography-mass spectrometry in metabolic profiling and metabolomic studies for biomarker discovery." *Biomarkers Med.* 1:159-185.
8. Koropatkin, N., D.W. Koppenaal, H.B. Pakrasi, and T.J. Smith. 2007. "The Structure of a Cyanobacterial Bicarbonate Transport Protein, CmpA." *Journal of Biological Chemistry* 282(4):2606-2614.
9. Welsh, E.A., M.L. Liberton, J. Stockel, H. Sato, T. Loh, C. Wang, A. Wollam, R.S. Fulton, S.W. Clifton, J.M. Jacobs, R. Aurora, L.A. Sherman, R.D. Smith, R.K. Wilson, and H.B. Pakrasi. 2007. "Genome sequence of *Cyanothece* 51142 shines new light on the circadian life style of a cyanobacterium." Submitted *Nature.*

10. Jacobs, J.M., E.A. Welsh, M.L. Liberton, J. Stockel, C.D. Nicora, D.G. Camp, II, H.B. Pakrasi, and R.D. Smith. 2007. "Proteomic Characterization of the Cyanobacteria *Cyanothece* 51142." Poster presentation at the 55th American Society for Mass Spectrometry Conference, Indianapolis, Indiana.

Structural Analysis of Membrane-Associated Proteins

Tom Smith and Nicole Koropatkin

Cyanobacterial growth is severely limited by the availability of several nutrients; bicarbonate, nitrate, iron, and zinc. In the aquatic environment that the cyanobacteria are found, the concentrations of these essential nutrients are exceedingly low. To compensate for this, the cyanobacteria have a number of high affinity, ATP driven pumps (ABC transporters). The basic components are the solute-binding domain that resides in the periplasm, the ion pore in the membrane, and the ATPase motor in the cytoplasm. There are a number of questions with regard to these transport systems. How do these systems specifically recognize particular solutes when there are others with similar characteristics? For example, how does the zinc transporter recognize zinc but not manganese or iron? How does the bicarbonate system bind bicarbonate but not the similar nutrient, nitrate? How are these systems regulated? How does the solute binding domain respond to the ATP motor on the other side of the membrane to release the nutrient into the membrane pore?

To address these questions, our work has elucidated the atomic details of the specificity and regulation of the uptake of several essential nutrients; iron, zinc, bicarbonate, and nitrate. These are the key, limiting nutrients that are essential for *Cyanothece* growth. We determined the atomic structures of the solute binding domains of the nitrate (NrtA), bicarbonate (CmpA), iron (FutA1), and zinc (ZnuA). Each protein was expressed, purified, crystallized in the presence and absence of the various ligands. In this way, we were not only able to see the atomic details of ligand recognition but also to ascertain how the solute was bound and released during membrane transport.

CmpA

Cyanobacteria, blue-green algae, are the most abundant autotrophs in aquatic environments and form the base of the food chain by fixing carbon and nitrogen into cellular biomass. To compensate for the low selectivity of Rubisco for CO_2 over O_2, cyanobacteria have developed highly efficient CO_2-concentrating machinery of which the ABC transport system CmpABCD from *Synechocystis* PCC 6803 is one component. We determined the structure of the bicarbonate binding protein CmpA in the absence and presence of bicarbonate and carbonic acid. CmpA is highly homologous to the nitrate transport protein NrtA. CmpA binds carbonic acid at the entrance to the ligand-binding pocket, whereas bicarbonate binds in nearly an identical location compared with nitrate binding to NrtA.

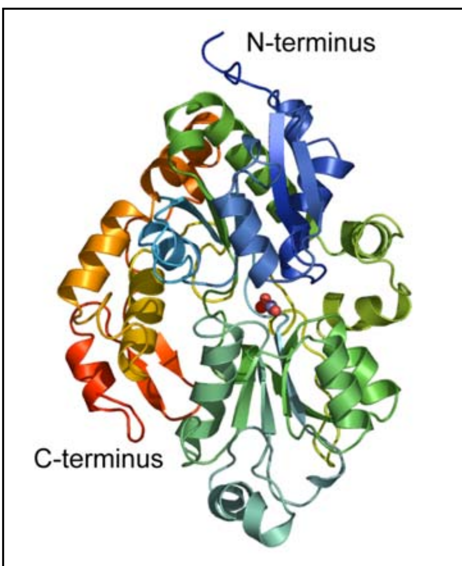


Figure 26. The structure of CmpA with bicarbonate bound (spheres in the middle).

Unexpectedly, bicarbonate binding is accompanied by a metal ion, identified as Ca^{2+} via inductively coupled plasma optical emission spectrometry. The binding of bicarbonate and metal appears to be highly cooperative and suggests that CmpA may co-transport bicarbonate and calcium or that calcium acts a cofactor in bicarbonate transport (see Figure 26).

NrtA

The single most important nutrient for photosynthesis and growth is nitrate, which is severely limiting in many aquatic environments particularly the open ocean. It is therefore not surprising that NrtA, the solute-binding component of the high-affinity nitrate ABC transporter, is the single-most abundant protein in the plasma membrane of these bacteria. Here, we describe the structure of a nitrate-specific receptor, NrtA from *Synechocystis* sp. PCC 6803, complexed with nitrate and determined to a resolution of 1.5 Å. NrtA is significantly larger than other oxyanion-binding proteins, representing a previously uncharacterized class of transport proteins. From sequence alignments, the only other solute-binding protein in this class is CmpA, a bicarbonate-binding protein. Therefore, these organisms created a solute-binding protein for two of the most important nutrients: inorganic nitrogen and carbon. The electrostatic charge distribution of NrtA appears to force the protein off the membrane while the flexible tether facilitates the delivery of nitrate to the membrane pore. The structure not only details the determinants for nitrate selectivity in NrtA but also the bicarbonate specificity in CmpA. Nitrate and bicarbonate transport are regulated by the cytoplasmic proteins NrtC and CmpC, respectively. Interestingly, the residues lining the ligand binding pockets suggest that they both bind nitrate (see Figure 27). This implies that the nitrogen and carbon uptake pathways are synchronized by intracellular nitrate and nitrite.

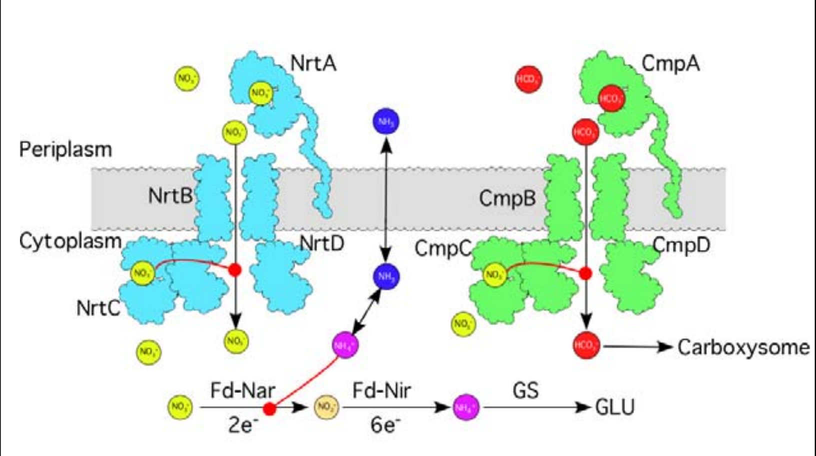


Figure 27. Model for the possible concomitant regulation of nitrate and bicarbonate import by intracellular nitrate.

FutA1

To better understand the mechanism of iron selectivity and transport, the structure of the solute-binding domain of an

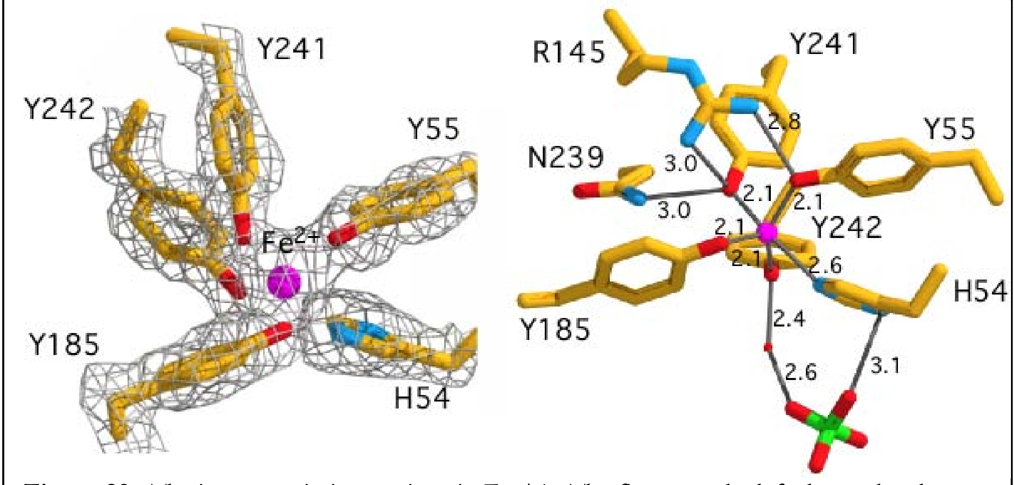


Figure 28. The iron-protein interactions in FutA1. The figure on the left shows the electron density of the bound iron (purple sphere) and the details of the interactions are on the right.

ABC iron transporter, FutA1, was determined in the presence and absence of iron. The iron ion is bound within the 'C-clamp' structure via four tyrosine and one histidine residues (see Figure 28). There are extensive interactions between these ligating residues and the rest of the protein such that the conformations of the side chains remain relatively unchanged as the iron is released by the opening of the metal binding cleft. This is in stark contrast to the zinc binding protein, ZnuA, where the domains of the metal binding protein remain relatively fixed while the ligating residues rotate out of the binding pocket upon metal release. The rotation of the domains in FutA1 is facilitated by two flexible b-strands running along the back of the protein that act like a hinge during domain motion. This motion may require relatively little energy since total contact area between the domains is the same whether the protein is in the open or closed conformation. Consistent with the pH dependency of iron binding, the main trigger for iron release is likely the histidine in the iron-binding site. Finally, neither FutA1 nor FutA2 binds iron as a siderophore complex or in the presence of anions and both preferentially bind ferrous over ferric ions.

ZnuA

A number of bacterial metal transporters belong to the ABC transporter family. To better understand the structural determinants of metal selectivity of one such transporter, we previously determined the structure of the periplasmic domain of a zinc transporter, ZnuA, from *Synechocystis* 6803 and determined that ZnuA binds zinc via three histidines. Unique to these ABC zinc transporters, ZnuA has a highly charged and mobile loop that protrudes from the protein in the vicinity of the metal binding site that we had suggested might facilitate zinc acquisition. To further examine the function of this loop, the structure and zinc binding properties of two ZnuA variants were determined. When the loop is entirely deleted, zinc still binds to the three histidines (see Figure 29). However, unlike what was suggested from the structure of a similar solute binding protein, TroA, release of zinc occurs concomitantly with large conformational changes in two of the three chelating histidines. These structural results combined with isothermal titration calorimetry data demonstrates that there are at least two classes of zinc binding sites; the high affinity site in the cleft between the two domains and at least one additional site on the flexible loop. This loop has approximately 100-fold weaker affinity for zinc than the high affinity zinc binding site and its deletion does not affect the high affinity site. From these results, we suggest that this region might be a sensor for high periplasmic levels of zinc.

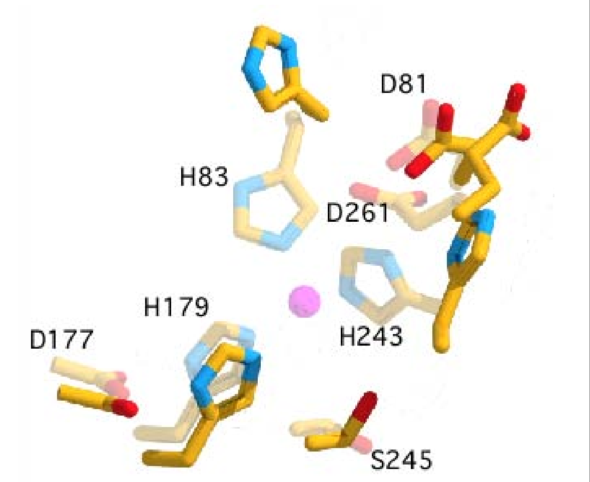


Figure 29. The difference in the ZnuA binding site when zinc is bound (translucent structure) and when it is released (solid structure).

Future projects

1) FutA2: FutA2 is homologous to FutA1. However, it does not appear to be regulated by iron and may be involved in actually protecting cyanobacteria from high iron concentrations. Therefore it apparently does not interact with the transport system and therefore there may be much to learned about how the solute binding proteins interact with the pore. We have

crystallized this protein and are on the verge of determining its structure. This structure and the associated paper will be coming out this year.

2) COP23: COP23 is a circadian oscillating protein identified by this Grand Challenge grant. It is clearly up regulated in the dark. We have cloned and expressed this protein and are trying to crystallize it this coming year.
3) Diflavin flavoprotein: The Grand Challenge collaborators have identified this protein as being highly expressed during the light portion of the circadian rhythm. We have cloned, expressed, and purified this protein. We hope to determine its structure and determine its function this coming year.

Publications, Presentations, and Awards

1. Koropatkin, N.M., D.W. Koppenaal, H.B. Pakrasi, T.J. Smith. 2007. "The structure of a cyanobacterial bicarbonate transport protein, CmpA." *J. Biol. Chem.* 282:2606-2614.
2. Wei, B., A.M. Randich, M. Bhattacharyya-Pakrasi, H.B. Pakrasi, T.J. Smith. 2007. "Possible regulatory role for the histidine-rich loop in the zinc transport protein, ZnuA." *Biochemistry* 46:8734-8743.
3. Bhattacharyya-Pakrasi, M., T. Ogawa, T.J. Smith, X-M. Si, and H.B. Pakrasi. 2007. Submitted *FEMS Letters.*
4. Koropatkin, N., A.M. Randich, M. Bhattacharyya-Pakrasi, H.B. Pakrasi, and T.J. Smith. 2007. "The Structure of the Iron-binding Protein, FutA1, from Synechocystis 6803." *J. Biol. Chem.* 282(37):27468-77.
5. Koropatkin, N.M., H.B. Pakrasi, and T.J. Smith. 2006. "Atomic structure of a nitrate-binding protein crucial for photosynthetic productivity." *PNAS* 103:9820-9825.
6. Koropatkin, N., A.M. Randich, M. Bhattacharyya-Pakrasi, D.W. Koppenaal, H.B. Pakrasi, and T.J. Smith. 2007. "Specificity and Regulation of ABC Transporters." Presented by Nicole Koropatkin at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, Wisconsin on June 6, 2007.

First Protein Structures from *Cyanothece* 51142

Garry W. Buchko, Howard Robinson (BNL), Michael A. Kennedy, and Shuisong Ni

Cyanothece is an intriguing organism that uses circadian rhythms to perform photosynthesis during the day and nitrogen fixation at night. Proteins that play significant roles in the processes are being identified by proteomic, transcriptomic, and metabolomic data collected as a part of this MBGC. While a number of the circadian-associated proteins, such as some of the photosynthesis proteins, are well characterized, many "hypothetical", uncharacterized proteins have also been observed to oscillate in unison with periods of light and dark. To gain an understanding of the biochemical roles these "hypothetical" and/or poorly understood proteins play in the circadian processes, our goal is to determine the structure of these proteins to gain insights on their function. Proteins that have been immediately targeted are the pentapeptide repeat proteins (PRPs), hypothetical proteins associated with nitrogen fixation (DUF683 and DUF269), two proteins associated with nitrogen fixation (NifW and NifZ), the circadian oscillating protein (COP23), and a Kai-protein (KaiB3) with interesting circadian rhythms. The goal is to characterize the properties of these proteins in solution and to determine their structures. Except for the C-terminal of KaiB3, no structural information exists for any of these five proteins; DUF683, DUF269, NifW, NifZ, and KaiB3. Hence, structural

information for these proteins promises to be unique and may provide important insights into the biochemical roles these proteins play in the complex circadian lifestyle of cyanobacteria.

Pentapeptide Repeat Proteins

Through the analysis of complete genome sequences that were becoming widely available in the late 1990s, a novel family of proteins containing tandem pentapeptide repeats of the general consensus sequence A[D/N]LXX was identified. Since the discovery, the number of pentapeptide repeat proteins (PRPs) listed in the Pfam data bank (Pfam00805) has escalated to approximately 3500. While present in eukaryotes, PRPs are found primarily in prokaryotes and especially in cyanobacteria. Thirty-five PRPs have been identified in *Cyanothece* 51142 and they are predicted to be located in the lumen/periplasm (7), plasma membrane (9), and the cytosol (19). Recent proteomic and transcriptomic microarray data with *Cyanothece* suggests that these proteins are being expressed in the cells and at least some of them may be associated with circadian rhythms. To gain a better understanding of the molecular function of PRPs in cyanobacteria and to characterize the effect of sequence variation on structural features of PRPs we have determined the crystal structure of two PRPs from *Cyanothece*, Rfr32 and Rfr23, to 2.1 Å resolution (see Figure 30). The structures of Rfr32 and Rfr23 are the first to be determined from DNA encoded in the genome of *Cyanothece*. Indeed, of the approximately 3500 PRPs listed in the Pfam data bank, Rfr32 and Rfr23 represent the second and third structures, respectively, determined in the PRP family.

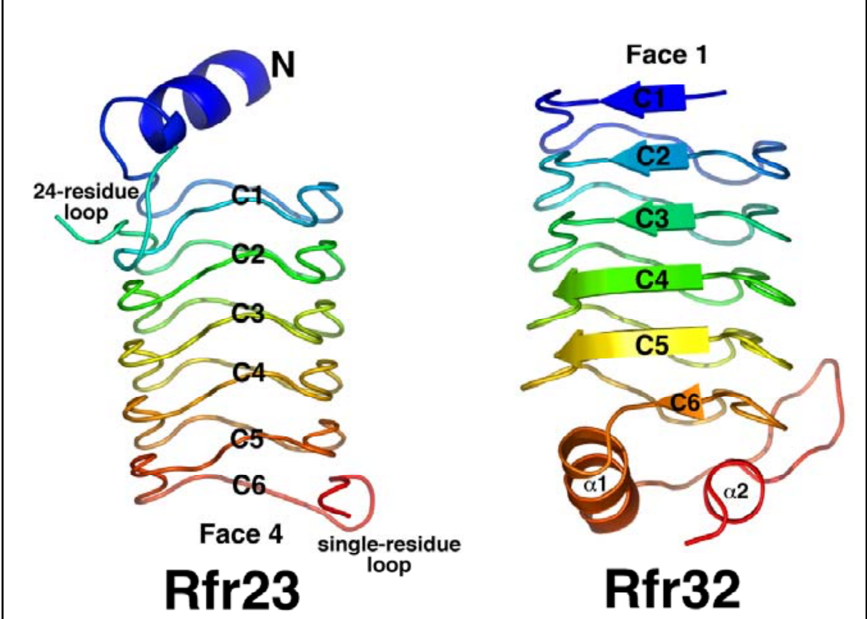


Figure 30. Structures of the first two proteins ever determined from the cyanobacterium *Cyanothece*. Both structures were determined using XRD crystallography methods and have been deposited into the RCSB protein data bank. Rfr32 (2F3I, 2G0Y), Rfr23 (2O6W). The protein is rainbow color-coded starting from the N-terminal (blue) to the C-terminal (red).

As illustrated in Figure 30, the repeated five-residues adopt a novel type of right-handed quadrilateral α-helix, an Rfr-fold, reminiscent of a square tower with four distinct faces. Each pentapeptide repeat occupies one face of the Rfr-fold with four consecutive repeats completing a 20-residue coil, the

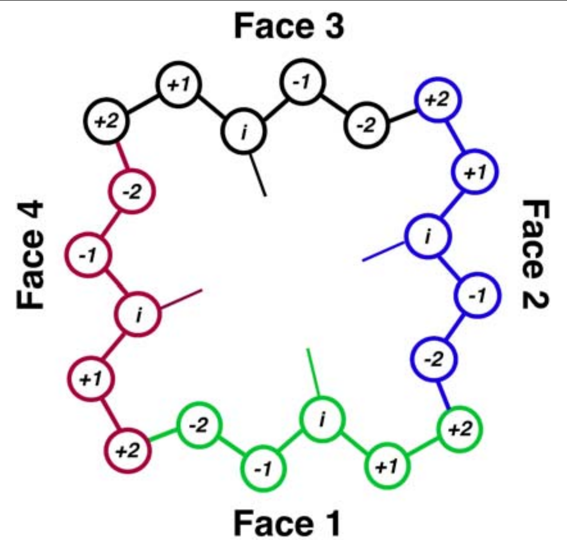


Figure 31. A stylized representation of a complete coil, the higher order structural unit adopted by four tandem pentapeptide repeats in an Rfr-fold.

higher order structural unit of the Rfr-fold. A stylized representation of the coil is shown in Figure 31. Residues at the i-2 and i position are directed towards the interior of the tower while those at the i-1, i+1, and i+2 position are directed towards the exterior. The coils stack on top of each other with a rise of approximately 4.8 Å with every revolution. The tower-like feature is maintained through two distinct, four-residue type II and type IV β-turns that may be universal motifs in maintaining the shape of the Rfr-fold in all PRPs. Given the repetitive nature of the primary amino acid sequence of pentapeptide repeats, it is not surprising that they adopt a similar structure. However, detailed analysis of the two PRP structures from *Cyanothece* suggests that there may be structural variations between PRPs. Such structural variation may reflect different biochemical functions. As highlighted in Figure 32, Rfr23 contains three structural features previously unobserved in PRPs, a 24-residue loop, a single-residue loop, and a disulfide bracket. The 24-residue loop is natively disordered, electron density was observed for only eight of the 24 residues. This loop, which has a net positive charge, may serve as a substrate-binding site between itself and the negatively charged surface of Face 4. On the other hand, the C-terminal single-residue loop may play a role in preventing edge-to-edge aggregation between Rfr-proteins. The disulfide bracket appears to have important structural "control" over the Rfr-fold as NMR and CD spectroscopy show that reduction of this single bond results in the complete, irreversible unfolding of the Rfr-fold. We speculate that in vivo reduction of the disulfide bond may play a role in the self-regulation of the activity of the protein.

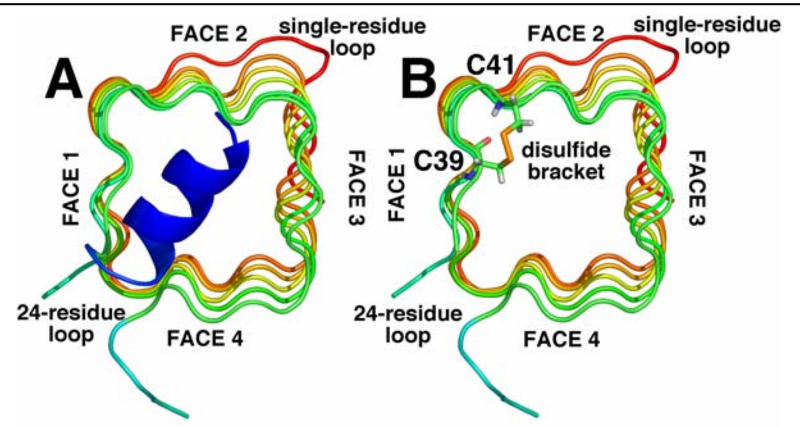


Figure 32. Cartoon representation of the structure of Rfr23 viewed from the top looking down the N-terminal highlighting three interesting features previously unobserved in PRPs. A) Entire protein shown. B) The N-terminal α-helix (M26* - E34) has been removed and the side chains of C39 and C42 included to show the position of the C39-C42 disulfide bond (disulfide bracket).

In total, about 20 of the 35 PRPs have been successfully cloned into *E. coli* expression systems. The majority of these clones are expressed primarily in the insoluble fraction or are highly aggregated in the soluble fraction. However, two of the approximately 20 PRP, Rfr10 and Rfr11, show promise for structural characterization. Rfr10 (183 residues, 18.7 kDa) is a dimer in solution. However, using perdeuterated samples and by collecting data at 45°C, the ^{1}H-^{15}N HSQC spectrum was relatively well dispersed and it was possible to collect much of the data necessary to assign the backbone of Rfr10. Because crystals of Rfr10 all diffract poorly (greater than 4 Å) it looks as if further characterization of Rfr10 will have to make use of nuclear magnetic resonance (NMR) data and this will require analyzing the NMR data sets to assign the backbone. Rfr11 (176 residues, 18.7 kDa) also expresses in the soluble fraction. Unfortunately, the current construct contains an approximately 5 kDa N-terminal tag that is difficult to remove. With the tag left on the protein, the NMR spectrum is too complicated to assign and the protein failed to crystallize under approximately 500 tested conditions. Because proteomics and transcriptomic data show that Rfr11 oscillates signicantly with periods of light and dark, the pursuit of this structure is continuing. A new construct of Rfr11 containing a small, 18-residue (approximately 1kDa), N-terminal tag has been cloned and will shortly be tested in early FY 2008.

DUF683 and DUF269

Analysis of the transcriptomics data has identified two proteins with "domains of unknown function", DUF683 and DUF269, that oscillate with diurnal cycles. The genomic location of these *Cyanothece* proteins are within a large cluster of approximately 35 genes that fix nitrogen. No structural information exists for any proteins in either of these two DUF families. While DUF683 is a small protein of 78 amino acid residues, size exclusion chromatography quickly showed that the protein was not behaving as a 10 kDa monomer. More detailed characterization by NMR spectroscopy shows DUF683 to be a homo-tetramer in solution. As a tetramer it has a molecular weight of ~40 kDa, and consequently, backbone assignment required the preparation of partially deuterated and specifically 15N-labeled samples. With such samples it was possible to completely assign the backbone for DUF683 as illustrated in Figure 33. Further analysis of the circular dichroism spectroscopy data and the carbon chemical shift and NOESY NMR data indicate that DUF683 is essentially all α-helical in solution. As shown in Figure 34, diffraction-quality crystals of DUF683 have also been grown and X-ray diffraction data have been collected out to a resolution of 2.6 Å using synchrotron radiation at Brookhaven National Laboratory. Initial electron-density maps confirm the conclusions reached from the NMR and CD data; the protein is all-helical. Efforts are in progress to complete the analysis of the X-ray diffraction data to obtain a high-resolution structure. Upon determining the structure, it will be possible to use the backbone NMR assignments to probe the biochemistry of DUF683 in solution with chemical shift perturbation and other experiments. The DUF269 gene has only been recently cloned into an *E. coli* expression system and pursuit of its structure will begin immediately in FY 2008.

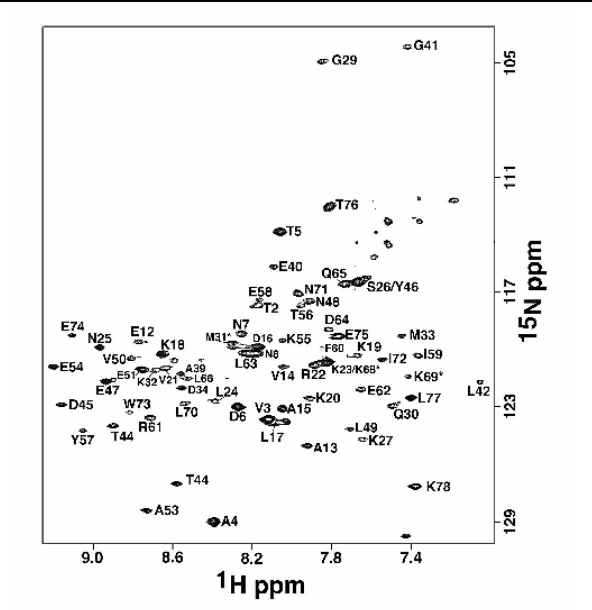


Figure 33. The ^{1}H-^{15}N HSQC spectrum for DUF683 collected at a ^{1}H resonance frequency of 600 MHz (20°C).

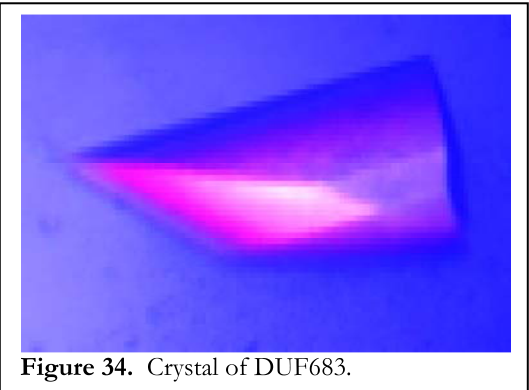


Figure 34. Crystal of DUF683.

NifW and Nif Z

NifW (116 residues, 13.3 kDa) and NifZ (114 residues, 12.9 kDa) are two small proteins located in the 35 gene nitrogen fixation cluster in *Cyanothece*. Proteomic data show that NifW has an interesting pattern of behavior. It has a double maximum, peaking once, as expected, early in the first dark cycle, but it also peaks again near the transition period between dark and light. NifZ does not show any significant oscillations with periods of light or dark, however, earlier studies in other cyanobacteria have shown that NifZ associates with NifW. While both proteins are believed to play some type of role in the nitrogen fixation process, their present biological function remains

unknown. Furthermore, there is no structural information for any protein in the NifW or NifZ family, and hence, any structural information on these proteins will be the first of its kind. Expression constructs containing the genes for both NifW and NifZ have been ordered and one of the goals for FY 2008 is to determine the structure for both proteins.

COP23

Analysis of the sequenced genome of *Cyanothece* revealed that the cyanobacteria contained a gene coding for a protein similar to one labeled COP23 from *Synechococcus*. COP refers to circadian oscillating protein, a protein that was observed to oscillate with periods of light and dark in *Synechococcus*. Transcriptomic and proteomics data on *Cyanothece* show that this similar protein also exhibits circadian oscillations in synch with periods of light and dark. The *Cyanothece* COP23 gene was cloned into an *E. coli* expression system and observed to express approximately 20% of the protein into the soluble fraction. Unfortunately, CD and NMR spectroscopic characterization of the soluble protein indicate that it is unstructured in solution. Efforts are in progress to either refold the protein from inclusion bodies or to optimize the solution conditions for the protein in the soluble fraction. However, it simply may be that COP23 is intrinsically unstructured in solution until it meets the correct substrate(s).

KaiB3

It has long been known that the Kai proteins play important biochemical roles in the circadian rhythm processes. In *Cyanothece*, KaiA, KaiB1, and KaiC1 are in one operon, KaiB2 and KaiC2 in a second operon, and KaiB3 and KaiB4 are by themselves. KaiB3 appears to be present only in non-thermophilic strains of bacteria that fix nitrogen. Transcriptomics data show that the levels of KaiB3 dwarf all the other Kai proteins, with a strong maximum observed in D1 and a minimum in L1. There is one structure for KaiB3 in the protein data bank, however, KaiB3 from *Cyanothece* contains an additional 150 N-terminal amino acids. An expression construct for *Cyanothece* KaiB3 (253 residues, 29.6 kDa) has been ordered and one of the goals for FY 2008 is to determine the structure of KaiB3.

Summary of FY 2008 goals

In FY 2008, the focus of our research will shift from the pentapeptide repeat proteins towards proteins associated with nitrogen fixation. Two manuscripts describing the backbone NMR assignments and crystallization protocols for DUF683 are ready to submit for publication. A structure for DUF683 should be solved and deposited into the RCSB protein data bank shortly. Constructs to express four other proteins associated with nitrogen fixation, DUF269, NifW, NifZ, and KaiB3 have been ordered.

Publications, Presentations, and Awards

1. Buchko, G.W., H. Robinson, H.B. Pakrasi, and M.A. Kennedy. 2007. "Insights into the structural variation between pentapeptide repeat proteins – Crystal structure of Rfr23 from *Cyanothece* 51142." In Press, *J. Struct. Biol.*
2. Buchko, G.W., H. Robinson, S. Ni, H.B. Pakrasi, and M.A. Kennedy. 2006. "Cloning, expression, crystallization, and preliminary crystallographic analysis of a pentapeptide repeat protein (Rfr23) from the bacterium *Cyanothece* sp. 51142." *Acta Cryst.* F62:2151-1254.

3. Buchko, G.W., S. Ni, H. Robinson, E.A. Welsh, H.B. Pakrasi, and M.A. Kennedy. 2006. "Characterization of two potentially universal turn motifs that shape the repeated five-residues fold – Crystal structure of a lumenal pentapeptide repeat protein from *Cyanothece* 51142." *Protein Sci.* 15:2579-2595.
4. Buchko, G.W. 2007. "Biophysical characterization of the solution properties of the pentapeptide repeat protein from *Cyanothece* 51142." *Can. J. Micro.,* in preparation.
5. Buchko, G.W., and H. Robinson, 2007. "Expression, crystallization, and preliminary crystallographic analysis of DUF683 from *Cyanothece* sp. 51142 – a hypothetical protein associated with nitrogen fixation." *Acta Cryst. F.,* in preparation
6. Buchko, G.W. and H.J. Sofia, 2007. "Backbone 13C, 15N and 1H NMR assignments of a hypothetical protein associated with nitrogen fixation - DUF683 from *Cyanothece* 51142." Submitted to *Biomol. NMR Assign.*
7. Oehmen, C., J.F. McDermott, J. Stockel, E.A. Welsh, J. Jacobs, T. Metz, A. Dohnolkova, G.W. Buchko, and H.B. Pakrasi. 2007. "Systems approach to understanding the molecular mechanism of the light and dark cycles of *Cyanothece* sp. 51142." Poster presentation, 9th Cyanobacterial Workshop, Delavan, Wisconsin June 6-10, 2007.
8. Buchko, G.W. 2007. "Towards understanding circadian rhythms in the cyanobacterium *Cyanothece* - Insights into the structure of pentapeptide repeat proteins and DUF683." Invited lecture, Washington State University, Pullman, Washington, June 7, 2007.
9. Buchko, G.W. 2006. "Tales and adventures on the road to the Protein Data Bank (PDB)- The chronicles of DR0079 and Rfr32." Invited lecture, Washington State University- Tri-Cities, Richland, Washington, November 6, 2006.

Transcriptional Profiling and Computational Modeling

Rajeev Aurora and Rangesh Kunnavakkam

Computational modeling is an essential component of the systems biology approach because the model is the result of data integration and the source of hypothesis generation. We have developed algorithms and model-building techniques that allow us to fit the data to generate models that can be validated by additional measurements. These models provide biological insights that are uniquely the strength of the systems biology approach.

Similar to other cyanobacteria, *Cyanothece* coordinates different biological processes by an internal clock that controls gene expression as well as resulting biological activities. To identify and study the coordinate regulation of the genes, at the transcriptional level, that underlie these two processes we need tools such as microarrays. For the MBGC project, our goal was to design the microarrays, which will be used for transcriptional profiling and to use these time-course profiles to develop a dynamic model of the *Cyanothece* metabolic network which will be used to understand how these biological processes are coordinately expressed. In our view, the organism uses the highest available energy-rich nutrients (i.e., sunlight in the day and nitrogen at night). The diurnal cycle is then a feed-forward anticipatory circuit that switches the system, including the reorganization of the membranes, to use the appropriate nutrient source. The mechanisms by which this is accomplished are encoded in the genome in the regulatory and enzymatic activities. Our goal is to understand the mechanism of this regulation.

Transcriptional Profiling

We designed a microarray from the first draft assembly of the *Cyanothece* genome. This array has been used extensively to generate mRNA expression profiles and analysis of these profiles at St. Louis University, Washington University at St. Louis (Welsh, Stockel) and at PNNL (McDermott, Oehmen). Recently, as the genome sequencing has been completed we have redesigned the microarray for the updated definitions of the open reading frames. More significantly, the design of version 2.0 includes putative promoter sequences so that these arrays can also be used for chromatin immunoprecipitation (ChIP) on the microarray (ChIP on chip) analyses. These studies can allow one to identify the binding sites of the transcription regulators and furthermore, identify the promoters that are occupied by the transcription factor under any given condition. The version 2.0 design provides a powerful tool to identify regulatory elements at the genome level that control gene expression.

Metabolic Modeling

Using the genomic sequence information we reconstructed the metabolic graph of *Cyanothece*. Graphs are collections of vertices and edges. In our case, the vertices are metabolites and an edge exists between two vertices if an enzyme encoded in the genome can convert one metabolite to another (Figure 35).

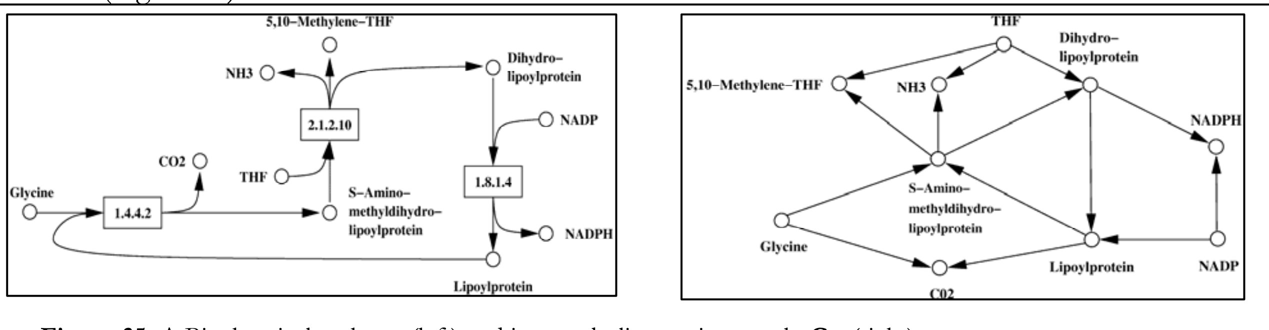


Figure 35. A Biochemical pathway (left) and its metabolic reaction graph, $\mathbf{G_M}$ (right).

Therefore, a metabolic reaction graph (GM) is bipartite directed graph that captures the substrate-to-product relationship. This reconstruction provides the total metabolic capability of the organism and has led to a number of insights. First, we found that relative to *Synechocystis* 6803 there are, most probably, a number of key enzymes that have not yet been annotated. The evidence for this comes from comparison of the metabolic graphs as shown in Table 1 below. The table indicated that there are roughly equivalent number of enzymes in *Cyanothece* (773) as in *Synechocystis* (831), but the number on metabolites are significantly less (roughly half). These results and a more detailed analysis

Table 1. Comparison of metabolic graphs of various organisms

Organism	Vertices (Metabolites)	Edges (Enzymes)
Reference	11,241	3,169
E. coli	3,142	1,178
Synechocystis 6803	2,014	831
S. cerevisiae	2,646	1,026
A. thaliana	3,066	1,136
Cyanothece	1,024	773

indicates that *Cyanothece* has additional metabolic functions that are not present in *Synechocystis* and further a number of enzymes are not yet annotated. Since this approach identifies the "holes" in pathways we can initiate a targeted search of theses enzymes in the genome.

One can also use such a graph to understand the fluxes of input nutrients (carbon, nitrogen, oxygen, etc.) through the system as the reaction carried out by the enzymes are well understood. This allows us to reasonably combine expression profile network with the underlying metabolic network for a diurnal cycle. The GM also shows how changes in a given metabolite will quantitatively affect other metabolites during the switch between photosynthesis and nitrogen fixation phases. A preliminary flux balance indicates that the system is not balanced, in accord with the notion that we are still missing annotation of number of enzymes. As these missing enzymes are identified, we will generate a better model in 2008.

Using the gene expression profiles to identify the enzymes that are cycling with the diurnal phase, we were able to identify a number of pathways and, by inference, the metabolites. Most of the metabolites are as expected and consistent with the expectation of nitrogen and carbon fixation during night and day, respectively. But, there were a few unexpected results. One result is that significant amount of O_2 is produced during the dark, and because O_2 inhibits nitrogenase this result is intriguing. One possibility is that the oxygen is consumed as it is produced, or is bound or compartmentalized in a manner that it cannot access the nitrogenase enzymes. Such inferences are powerful, because the generate new testable hypotheses that could not have been generated from the measurements, and may, for instance, indicate that low-level nitrogen fixation could occur under low light when oxygen evolution is at low levels.

Publications, Presentations, and Awards

1. Welsh, E., et. al. "Genome sequence of *Cyanothece* 51142 shines new light on the circadian life style of a cyanobacterium" Submitted *Nature*.
2. Aurora, R., E. Welsh, J. Stockel, and H. Pakrasi. "Metabolic network reconstruction of *Cyanothece*: Dynamics during the Diurnal Cycle." (Manuscript in preparation).

Computational Biology and Bioinformatics Tools for Understanding the Role of Membrane Proteins in Diurnal and Circadian Processes of Prokaryotes

Christopher Oehmen, Jason McDermott, Lee Ann McCue, Ronald Taylor, Bobbie-Jo Webb-Robertson, and Heidi Sofia

In the spirit of systems approach to biology, the MBGC project teams are collecting large, high-dimensional datasets from multiple analytical approaches in order to investigate the role of membrane systems in the diurnal cycle of the cyanobacteria *Cyanothece* from experiments such as time-course studies. The relevance of this project to the overall goals of the MBGC is that we (1) provide analytical tools, training, software support and access to specialized hardware for managing, analyzing, and integrating the data produced by the Grand Challenge team and (2) provide data analysis and integration support to extract mechanistic, systems-level descriptions of processes fundamental to cell cycling, energy production, carbon handling, and other processes in the context of membrane systems in *Cyanothece*.

The key data sources for studying *Cyanothece* are the sequence and annotation, transcriptomics, proteomics, metabolomics, crystal structures of proteins, and imaging studies aimed at revealing interactions between proteins, membrane systems, and macroscopic structures. The team is culturing *Cyanothece* concurrently at multiple sites, allowing each site to independently devise experiments to investigate different aspects of cycling—but most of these experiments yield data of each modality leading to a large data integration and management challenge. For instance, a single 48-hour time course experiment easily results in 16 full genome microarrays, 24 full proteomics datasets, and 24 metabolite profiles. We focus on tools and analysis supporting a global view of these data sources, such as whole-genome or whole-proteome network inference, whole-proteome differential proteomics, and genome context mining. We also provide analysis support using software tools specially modified to allow for integrating multiple data modalities into a single coherent picture—transcriptomics and metabolomics simultaneously, for instance. These tools and results of data analysis and integration are described in greater detail below.

Software Development, Deployment, and Training

In the first year of the project, we focused on making data analysis software available to the MBGC team. Through this effort, four key tools were made available to the team, three of them being accompanied by training and tutorial presentations. Figure 36 illustrates an example of how we used these tools to identify possible functions of previously uncharacterized proteins suspected to play a role in nitrogenase. The four tools are briefly described below.

1. Similarity Box, a dendogram viewer and browsing tool, was used to perform genome context analysis revealing proteins of unknown function (DUF proteins) which are highly conserved and found in the neighborhood of nitrogenase genes in several species closely related to *Cyanothece*. These targeted proteins are being structurally analyzed by Dr. Garry Buchko.
2. ScalaBLAST, a high-throughput sequence analysis engine, was used to perform large-scale BLAST calculations using the EMSL supercomputer, dramatically speeding up the BLAST calculations required to complete the *Cyanothece* annotation. Over 17000 CPU hours were consumed by this project (over 2 CPU years).
3. SEBINI, a network inference testbed providing access to a variety of network inference methods was modified for use on time course data and used to provide supporting evidence for presence of DUF proteins concurrent with nitrogenase using proteomics data.
4. PQuad, a proteomics results browser has been used to provide additional evidence for the role of key proteins in nitrogenase by providing a global view of differential proteins present in dark vs. light conditions.

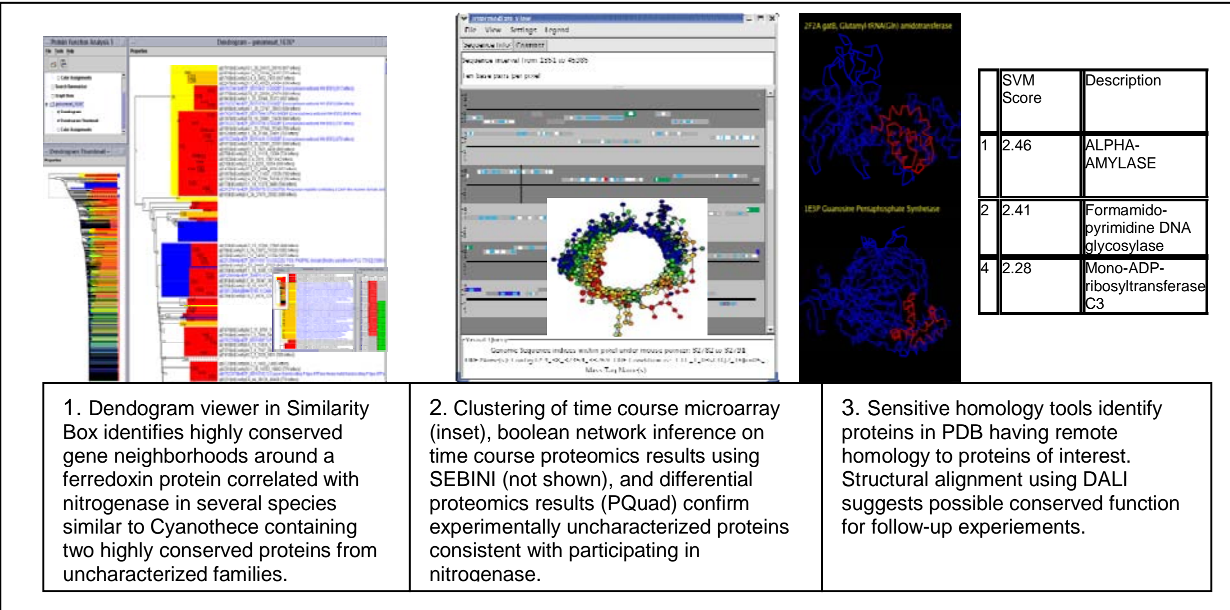


Figure 36. Example analysis pipeline using software tools provided by this project.

Additional experimental software has been applied to data arising from the MBGC including a network viewer (BellaVista) which has been used to show the cyclic nature of genes, proteins, and metabolites in the light/dark time course experiments (Figure 37), as well as integrating across these datasets.

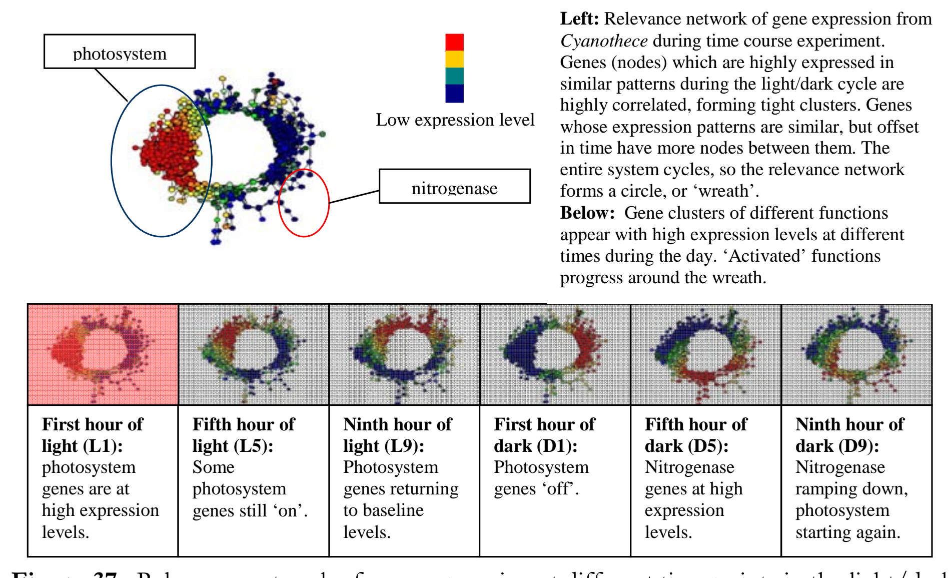


Figure 37. Relevance network of gene expression at different time points in the light/dark cycle of *Cyanothece*.

A support vector machine (SVM)-based sensitive homology detection tool, which has been shown to be substantially more sensitive than BLAST has been used to produce a list of remote homolog candidates for the DUF proteins identified in the first year. We are following up on these candidate homologs to verify the functional roles of the DUF proteins through a collection of collaborative, directed experiments. We plan to test the sugar-binding capability of one of the DUF proteins (suggested by sensitive homology detection) in conjunction with Garry Buchko's protein structure studies. Sugar-binding may play a key role in nutrient sensing, or signaling, which may play a critical role in cell cycling.

Integrated Data Analysis of Diurnal Cycling Data

Data analysis efforts thus far have focused on analyzing microarray, proteomics, metabolomics, and other data sources in the context of the *Cyanothece* time course. Time course gene expression data from the MBGC team was clustered using Pearson correlation with a cutoff to produce relevance networks. Genes with closely matching time course profiles cluster together in this method, which produces a wreath-like graph, shown in Figure 37. Similar analysis on proteomics results reveals similar cyclic characteristic. Adding correlations with metabolomics results to either graph reveals clusters of metabolites having mass consistent with products or reactants consistent with processes associated with genes or proteins, illustrated in Figure 38. This reference frame will allow us to continue exploring correlations between molecular components, and systems-level processes at work in *Cyanothece*, and identify new targets for protein structure studies, metabolites of interest, and processes likely to play important roles in controlling circadian and/or diurnal cycle of *Cyanothece*.

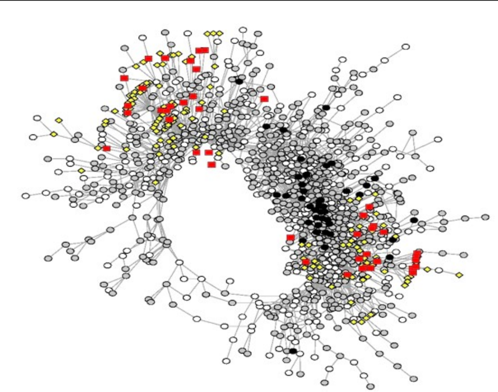


Figure 38. Relevance network integrating 'expression' pattern of proteins (clear nodes) and metabolites (red and black nodes) at a single time point in the light/dark cycle. This approach can help 1) give evidence for the identity of clusters of metabolites which correlate to specific cell functions or pathways, and 2) give insight into system-level processes that cannot be captured in a single data modality.

A second set of collaborative experiments is aimed at finding how cell division is related to the light/dark cycle in *Cyanothece*. This endeavor will extend our findings from transcriptomics, proteomics, and metabolomics studies with a new set of experimental conditions in tandem with imaging studies aimed at characterizing the trajectory of the cell population through the cell division cycle. This is likely to involve key regulatory processes that affect and may react to membrane organization.

Publications, Presentations, and Awards

1. Taylor, R.C., et al. 2006. "SEBINI: Software Environment for Biological Network Inference." *Bioinformatics* 21:2706-2708.
2. Webb-Robertson, B.J., et al. 2007. "PQuad - Visual Platform for Proteomics Peptide and protein Data Exploration for Prokaryotic Systems." *Bioinformatics* 23(13):1705-7.
3. Oehmen, C.S., et al. 2007. "Systems Approach to Understanding Molecular Mechanisms of the Light-Dark Cycle in Cyanothece sp. 51142." 9th Cyanobacterial Molecular Biology Workshop, Delavan, Wisconsin on June 6-10, 2007 (poster).

Data Assimilation, Visualization, and Mining

Elena S. Peterson, Kyle Klicker, Leigh Williams, and Gordon Anderson

The initiative's biology research involves the generation of very large volumes of data using a number of different technologies, such as genomics, proteomics, metabolomics, and high-resolution imaging. Integration of datasets presents a challenge requiring the development of new capabilities. In managing the raw data, analysis methods must be documented and all metadata from these analyses must be recorded. Additionally, the procedures used for organism growth and sample preparation must be documented and made available to other researchers and data analysts. After data are generated, a number of computational tools can be used by biologists and bioinformaticists to generate hypotheses about a particular biological question. These tools are often hard to gather and use because of data access, formatting, and annotation issues.

Under this project, we have gathered and customized a set of data management tools to facilitate the collection and organization of experimental data and metadata. This collection of tools aided the MBGC researchers by helping to manage experimental design data, results data, and to help facilitate the collaborations through sharing of raw data results and experimental protocols. We created customized software that brought together several existing prototype analytical tools and data management systems to enable the researchers' data workflow. The outcomes of this project include a storage mechanism for experimental results, access to external data sources useful in analysis, a web portal for public dissemination of MBGC literature and data, and customized access to bioinformatics tools.

Collaborative Data Sharing

During FY 2005, efforts focused on gathering data, storage and sharing requirements, and collecting existing tools to leverage as the research was just getting started. We provided a document management portal to allow for the collection of preliminary research data such as sample preparation operating procedures, drafts of manuscripts, and early results data. We also made available some existing software tools through secured access portals and live web tutorials. We purchased the hardware needed to support the long-term storage and processing of the various data types that will be generated. We also worked with Dr. Richard Smith's proteomics pipeline team to begin to connect MBGC researchers to their proteomic data results through an externally accessible system. The leveraging of the existing data storage model saved us the effort of recreating a system to store MBGC proteomic and metabolomic data, while allowing us to customize access of the data by bioinformatic tools in later years.

Connectivity to Software Applications

During FY 2006, we focused on the application and customization of bioinformatics analysis tools. We took the PQuad software (Platform for Proteomics, Peptide, and Protein data exploration) developed for the visualization of peptide-based proteomic data and redesigned it for the exploration of differentially expressed peptides and proteins of interest in *Cyanothece*. Significant time was spent with bioinformaticists in customizing the software to visualize high-throughput proteomics data in a rapid and intuitive manner. We found that the size of the data sets made the software too slow to be useful, so we developed new algorithms for loading and preparing the data

for visualization. It will now handle entire prokaryotic proteome datasets in under a minute. The original PQuad software was also missing some key functionality centered on comparative analysis. Because of the customized modifications made to the software, PQuad was used by Dr Chris Oehmen's team to identify proteins that were differentially expressed in light and dark conditions with high confidence based on peptide coverage during the diurnal cycling time course experiments.

In FY 2007, we integrated with the Bioinformatics Resource Manager team to provide MBGC researchers with several existing analysis tools for mining of proteomics, metabolomics, and microarray data. We prepared a version of the BRM software that would allow access to raw proteomic data for MBGC collaborators only. We also provided infrastructure support for deploying the SEBINI network inference software to external collaborators. SEBINI was used by Drs. Ron Taylor and Chris Oehmen to create inferred networks from the diurnal cycling time course data and is now being used by other collaborators to do the same.

Data Management, Archival, and Dissemination

During FY 2007, we focused on storing and integrating the experimental metadata that are associated with the proteomics, metabolomics, microarray, and image data that are starting to emerge from our MBGC researchers. We provided access to a newly developed web portal for data sharing, called the Experimental Data Management System, which allows for project and experimental hierarchies to be created and sample information to be entered. It also allows for supplementary protocol files to be attached to experiments and raw data to be attached to the samples for a complete workflow of data generation and management. We also designed and deployed a public web site to disseminate literature about the initiative and published data from the collection of projects (http://mbgc.emsl.pnl.gov/).

Publications, Presentations, and Awards

1. Webb-Robertson, B.M., E.S. Peterson, M. Singhal, K.R. Klicker, C.S. Oehmen, J.N. Adkins, and S.L. Havre. 2007. "PQuad - A visual analysis platform for proteomic data exploration of microbial organism." *Bioinformatics* 23(13):1705-1707.
2. Oehmen, C.S., J.E. McDermott, J. Stockel, E.A. Welsh, G.W. Buchko, J.M. Jacobs, T. Metz, A. Dohnalkova, G. Orr, E.S. Peterson, and H.B. Pakrasi. 2007. "Systems approach to understanding molecular mechanisms in the light-dark cycle of *Cyanothece*." Presented by Christopher Oehmen at 9th Cyanobacteria Molecular Biology Workshop, Delevan, Wisconsin on July 7, 2007.
3. Oehmen, C.S., L. McCue, J.N. Adkins, K.M. Waters, T. Carlson, W.R. Cannon, B.M. Webb-Robertson, D.J. Baxter, E.S. Peterson, M. Singhal, A. Shah, and K.R. Klicker. 2006. "High-Throughput Visual Analytics for Biological Sciences: Turning Data into Knowledge." Presented by Christopher Oehmen (Invited Speaker) at Supercomputing 2006, Tampa, Florida on November 14, 2006.
4. Peterson, E.S. 2007. "Strategies for collecting and managing experimental data for future analysis." Presented by Elena S. Peterson (Invited Speaker) at Oregon Graduate Institute, Portland, Oregon on February 16, 2007.

Summary and Future Plans

In less than 30 months, the MBGC project has provided a tangible proof for the concept of a Grand Challenge project that has capitalized on the unique large-scale facilities at EMSL. Today's genomics based systems biology projects in general need collaborations between scientists and technologists in diverse disciplines. The MBGC research group has evolved into a cohesive team that has wonderfully adopted the central philosophy of interaction and interdependence large-scale biology projects. During this project period, *Cyanothece* 51142, a relatively unknown cyanobacterium, has emerged as one of the most attractive bacterial systems to study circadian rhythms at various levels of physiology, metabolism, and structure. A testament to this assertion is the recent funding of two independent projects, one at PNNL and one at Boston University, by the DOE-GTL program to study aspects of *Cyanothece* 51142 biology that are complementary to those in MBGC. In addition, as described earlier, the same program last year approved a proposal from Principal Investigators Himadri Pakrasi and Louis Sherman to sequence six additional *Cyanothece* strains that have been isolated from different ecological niches all over the world. This effort is already underway at DOE's Joint Genome Institute (JGI). Together, these activities have transitioned the initial research objectives of MBGC to a much broader level, and it is conceivable that in the near future, a *Cyanothece* collaboratory will be formed to provide a forum for interactions between all researchers working on *Cyanothece*.

The future goal of the MBGC project will be to use different *Cyanothece* strains for CO_2 assimilation and bioenergy production. All current biofuel enterprises result in the production of copious quantities of CO_2. Although conversion of plant feedstock to CO_2 as a waste product of biofuel fermentative processes may be considered a carbon neutral process, the amount of CO_2 produced will be staggering, as biofuel production becomes a viable source of energy. Thus, at the least, it would be incumbent to consider processes that might use CO_2 as a viable starting material for the synthesis of value-added products. At the most, if the bulk of the CO_2 produced could be "bioassimilated" on a large scale, an important contribution towards decreasing the amount of CO_2 emitted into the atmosphere would result.

Photosynthesis is the ultimate source of all fossil fuels. It has been the dominant energy conversion and storage process on Earth for at least three billion years. It has been fine-tuned by evolution so that photosynthetic organisms are ubiquitous in nearly every known habitat that supports life and forms the basis of essentially all food chains. Given the success of photosynthetic organisms as living solar energy conversion systems, it is natural that they be investigated as vehicles for providing energy for human use, either directly or as the inspiration for biomimetic systems. The products of energy storage in natural photosynthetic organisms are used primarily for cell growth and reproduction. The production of large amounts of easily harvestable, high-energy products is not usual in natural microbial photosynthetic systems. For this reason, the use of genetically manipulated organisms in which a large fraction of the stored energy has been diverted into nonnative metabolic pathways is attractive. This will require that any large-scale production facility use closed bioreactors to avoid the introduction of genetically engineered organisms into the environment. Closed bioreactors are also highly desirable in terms of water use characteristics, as water loss is minimal in such systems. This permits them to be sited in arid environments that are unsuitable for either traditional crops or open pond systems. Long-term energy farming applications such as energyplexes will certainly use such environments in order to be economically feasible.

An appealing aspect of the use of photosynthetic microorganisms is that they do not produce macroscopic structures with robust mechanical properties, such as lignocellulose, that need to be broken down to recover the stored energy. The basic mechanisms of light-energy capture and primary photosynthetic energy storage are essentially the same in all photosynthetic organisms, so any system in which the cellular complexity is low is attractive. In principle, a large fraction of the biomass in such microbes is easily recovered as usable fuels, or organisms can be engineered to produce recoverable fuels such as hydrogen or liquid alcohols. Another advantage of photosynthetic microbes is that they are easily genetically manipulated. Genome sequences for many *Cyanothece* strains will soon be available, and robust systems are being developed for stable introduction and expression of foreign genes from a variety of sources.

The primary challenges to using photosynthetic microorganisms in bio-energy production are in selecting suitable organisms, engineering them to produce the appropriate easily recovered high-energy products, as well as developing efficient and cost-effective production facilities, such as photobioreactors. In this project, we plan to use oxygenic cyanobacteria.

Future Research Directions and Needs

- Use *Cyanothece* strains as model prokaryotic photosynthetic organisms to develop a detailed understanding of phototrophic prokaryotic energy conversion and CO_2 utilization systems.
- Adopt a systems biology approach to understand the regulation of metabolism as a function of environmental conditions. It is important to develop a *Cyanothece* strain as a laboratory model so that we can determine the important characteristics needed when seeking new organisms in nature.
- Apply molecular biology and genetics to construct organisms with appropriate metabolic properties. For example, development of strains with higher ethanol production, more lipid storage for biodiesel production, higher levels of H_2 evolution and more metabolite storage.
- Culture of *Cyanothece* strains to harvest the solar spectrum from 400 nm to 900 nm for the production of metabolites such as biodiesel and H_2.
- Develop the energyplex concept. A long-term goal is to develop the cyanobacterial system for use as part of an energyplex that will integrate recycling CO_2 from a power generation facility. The *Cyanothece* cells will use photosynthesis to first-capture carbon and then use the reduced carbon for the production of biofuels and other valuable compounds.

The MBGC effort initiated by DOE and EMSL has provided a significant start to the attainment of these longer-term goals. In addition to the provision of basic research findings and improved understanding of the versatile *Cyanothece* organism, the MBGC has demonstrated the use and value of large, integrated-team, focused research at a unique DOE user facility. It has also raised the collective scientific consciousness regarding the potential energy and CO_2 mitigation possibilities of photosynthetic organisms.

Appendices and Reference Information

A. Call for Concept Papers

B. Awarded Proposal

C. Advisory Committee

D. New Capabilities Resulting in Increased Scientific Impact of EMSL

E. Publications/Presentations/Award

F. Advisory Committee Report

G. Participants Contact Information

H. Executive Summary - Biogeochemistry Grand Challenge

A. Call for Concept Papers

A Scientific Grand Challenge in Membrane Biology

Call for Concept Papers

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a multiprogram national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington, is seeking concept papers for a Scientific Grand Challenge in the area of membrane biology. Membrane processes are critical to solving highly complex problems related to energy production, carbon sequestration, bioremediation—all mission areas of the U.S. Department of Energy (DOE)—and other problems of national interest in energy and environmental science. The solutions to these problems require an approach that not only involves multidisciplinary teams of experts in the scientific area of choice, but that also require access to a range of cutting-edge instrumentation and computational resources not available at a single institution. This Scientific Grand Challenge will build interdisciplinary "team science" in biology while using multiple capabilities at EMSL to achieve a systems-level understanding of membrane processes.

Background

EMSL houses researchers who perform molecular studies focused on solving the major environmental challenges facing DOE and the nation. The facility contains a broad spectrum of world-class scientific resources that integrate theory, modeling, and simulation with experimentation and enable novel approaches for solving such difficult challenges.

In Fiscal Year 2004, EMSL is implementing a set of Scientific Grand Challenges to address important scientific questions using a collaborative approach. These Scientific Grand Challenges—each expected to last three to five years in duration— will have a set of well-defined goals that must lead to enduring scientific impact and may involve additional capability development. They will focus on critical milestones in the advancement or use of science, will be user driven, will take advantage of the full spectrum of resources and technical expertise available at EMSL, and will support DOE mission areas.

Successful concept papers submitted for EMSL's Scientific Grand Challenges will pose scientific questions that cannot readily be addressed without access for substantial periods of time to the full range of scientific instrumentation, computational resources, and research teams located at EMSL. In addition to the capabilities of the EMSL, these Scientific Grand Challenges will require teaming of scientists from multiple disciplines and use of specialized instrumentation at the participants' home institutions, and possibly use of capabilities located at other DOE user facilities (e.g., Joint Genome Institute, synchrotron light sources). The concept papers must clearly identify a scientific leader who will be responsible for the overall management of the proposed work and who is able to attract some of the best research scientists in the applicable research area.

Research Area of the Call for Concept Papers

One research area of interest to EMSL's Scientific Grand Challenges—addressed by this Call for Concept Papers—is biological membrane processes in cells (e.g., energy transduction, metal/ion regulation and transport, photosynthesis, signal transduction, cellular membrane architecture, dynamics of membrane proteins, and regulation of conformation states of proteins). Understanding membrane processes requires a systems-level analysis of fundamental cellular processes including the genetics, biochemistry, and biophysics of such processes, the characterization of which are particularly well suited to the capabilities of EMSL. Accomplishment of this Scientific Grand Challenge will elucidate how these membrane processes fit into the overall cellular physiology and ecology.

EMSL Signature Characteristics and Capabilities

EMSL's signature characteristics include

- integrated theory, modeling, and simulation with experimentation
- multidisciplinary teams and collaborative modes of operation to solve major scientific problems
- research teams responsible for developing extraordinary tools and methodologies
- scientists who design experimental strategies and operate state-of-the-art instruments
- education and training in the use of sophisticated instrumentation and computational systems and approaches
- a cyber infrastructure that facilitates productive remote interactions
- the charter to deliver capability in a transparent manner and to facilitate user outreach.

EMSL also boasts unparalleled resources and infrastructure in high-performance computing and informatics, nuclear magnetic resonance spectroscopy, multimodal optical spectroscopies, and imaging technologies, as well as advanced mass spectrometry capability for global proteomics—all of which are essential to progress in the study of membrane biology. Supporting these technologies are world-class researchers from PNNL, including chemists, biologists, physicists, and software engineers, as well as instrument designers and builders.

Process for Submitting Concept Papers

EMSL requests submission of a concept paper outlining the Scientific Grand Challenge goal, approach, technical requirements, and expected scientific and technical outcomes. Interested parties should become familiar with EMSL and its capabilities (http://www.emsl.pnl.gov/). Concept papers of no more than five pages must be postmarked by **May 14, 2004**, to the address below. An external review committee will select concepts for further development based upon scientific merit and potential for sustained scientific impact, appropriateness of EMSL resources for investigation of the challenge, and alignment with DOE mission needs in biology. EMSL will then host a workshop(s) for the successful concept team(s) to develop a scientific plan, strategy, and resource requirements for the Scientific Grand Challenge, and this information will be used for final selection of the challenge to be implemented. EMSL staff will provide assistance to the successful team in implementing their Scientific Grand Challenge. To submit concept papers or obtain further information, contact:

J. W. Rogers, Jr., PhD
Director, William R. Wiley Environmental Molecular Sciences Laboratory
P.O. Box 999, MSIN K1-46
Richland, Washington 99352
(509) 376-5328
jw.rogers@pnl.gov

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

(*Science* 304(5667):16 April 2, 2004; *ASM News* 70(5):256 May 2004)

B. Awarded Proposal

Scientific Grand Challenge in Membrane Biology: A Concept Paper

Systems Analysis of the Dynamics of Membrane Architecture, Composition and Function in Cyanobacteria

Project Leader: Himadri B. Pakrasi, Department of Biology, Washington University

Co-investigator: Louis A. Sherman, Department of Biological Sciences, Purdue University

Co-investigator: Teruo Ogawa, Department of Biology, University of Tokyo

Co-investigator: Bijoy K. Ghosh, Department of Electrical & Systems Engineering, Washington University

Co-investigator: Rajeev Aurora, Department of Electrical & Systems Engineering, Washington University

Key Collaborators at PNNL

Richard D. Smith and David G. Camp	High-throughput membrane proteomics
Yuri A. Gorby and Jim K. Fredrickson	Controlled cultivation of cyanobacteria
H. Peter Lu and Steven D. Colson	Image analysis of cyanobacterial cells
George S. Michaels and H. Steven Wiley	Network inference and modeling

Grand Challenge Questions

The ultimate goal of this project is to use systems biology approaches to determine the underlying network that governs the forms and functions of cyanobacterial membranes and their components. The key questions that confront us are:

1. What is the relationship between plasma membrane and thylakoid membrane, two intracellular membrane systems in cyanobacteria?
2. How does the transcriptome of cyanobacteria respond to environmental changes such as light intensity, CO_2 level, metal concentrations and circadian rhythm?
3. How does the cyanobacterial membrane proteome change under similar perturbations, and how well does such dynamic pattern relate to that of the transcriptome?
4. What are the functional changes in key biological membrane processes (energy transduction, photosynthesis, metal homeostasis and signal transduction) under the same environmental perturbations?

The cellular components are always in flux, and molecular machines assemble, function and disassemble as a function of time and environmental alterations. It is imperative to utilize a systems biology approach and integrate all temporal information into a predictive, dynamic model to understand the functioning of a cell, in general, and the cellular membrane systems, in particular.

Introduction

Many have called the twenty-first century the 'biological century'. The central approach in most of the biological sciences during the last half of the twentieth century has been based on reductionism, and has resulted in a massive increase in our knowledge of individual cellular components. However, we have also realized that while we know how the individual 'nuts and bolts' function in isolation, our understanding of how these components interact with each other to define the cellular, organismal, and population level behaviors of living beings has remained far less sophisticated.

The recent achievements in genome sciences have opened new paradigms in biological research. Complete genome sequences have become available for many microorganisms and a few eukaryotes, including humans and plants. To extract useful information about the biological functions of any organism from such sequences, the challenge is to develop and use new computational and experimental tools to integrate data at various levels of biologically relevant interactions. Such a 'Systems Biology' approach should provide enabling technologies to examine complex biological processes, which should in turn result in an integrated and predictive understanding of how an organism behaves and responds to environmental changes.

Clearly, the successful implementation of such a systems approach will require collaborative interactions between biologists, computer scientists, mathematicians and model builders, engineers, physicists, chemists, and perhaps specialists in other disciplines. The current call for a scientific grand challenge in membrane biology at PNNL offers an excellent and timely opportunity to assemble a team of academic researchers, and scientists at PNNL to initiate a systems biology project for a fundamental understanding of membrane processes in cyanobacteria.

Organisms

Cyanobacteria are oxygenic photosynthetic prokaryotes that make significant contributions in harvesting solar energy in the biosphere, planetary carbon sequestration, metal acquisition and H_2-production. The genomes of more than a dozen different cyanobacterial species have been sequenced. During this project, we plan to focus on two cyanobacterial species: *Synechocystis* sp. PCC 6803, and *Cyanothece* sp. ATCC 51142. *Synechocystis* 6803 is a fresh water organism (Figure 1), and is arguably, the most widely studied cyanobacterium during the past decade. It is also the first photosynthetic organism with a completely sequenced genome (1996), and has nearly 3700 annotated genes (www.kazusa.or.jp/cyano/cyano.html). This species is naturally transformable, and targeted gene replacements can be readily generated. Whole genome microarrays for this organism have been

developed, and the project leader and all of the co-investigators have significant collective experience in designing, utilizing and analyzing such microarray-based experiments. Another genome-level project headed by Wim Vermaas (a close colleague of the project leader) was funded in the 'Microbial Cell Project'. A yeast two-hybrid protein interaction map is also being generated at the KAZUSA institute. Finally, a large collection of gene knockout mutants is available to the research community.

Cyanothece is a marine organism that is capable of oxygenic photosynthesis, nitrogen fixation as well as heterotrophic growth in the dark. Most importantly, this organism displays one of the *most striking circadian rhythms identified in any prokaryote* (1). This is needed because the enzyme that fixes atmospheric N_2, nitrogenase, is highly sensitive to oxygen, an obligatory product of oxygenic photosynthesis. Because this is a unicellular organism, it cannot form specialized cells to protect its nitrogenase. Thus, the strain uses time as a way of separating the peaks of photosynthesis from peaks of nitrogenase activity (Fig. 2). The genome of *Cyanothece* 51142 has not been sequenced yet. Recently, Pakrasi and Sherman have submitted a proposal to the Washington University Genome Sequencing Center for draft sequencing of this genome.

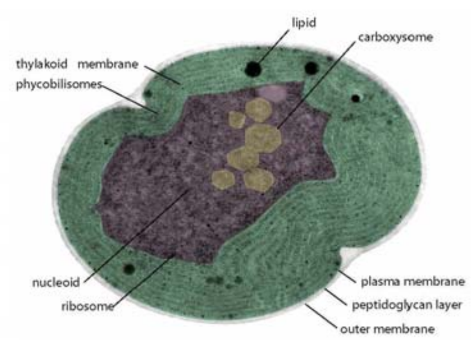


Figure 1: A transmission electron micrograph (false colored) of a dividing cell of *Synechocystis* 6803 (from the CYANOBASE website). Indicated are various subcellular locations and macromolecular assemblies.

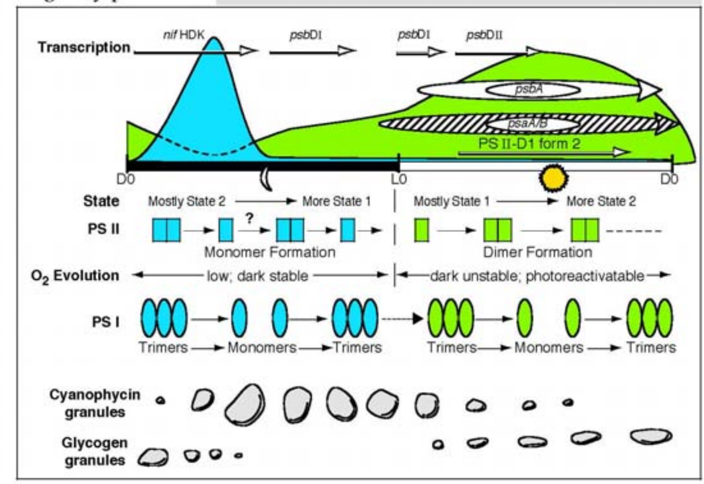


Figure 2. Diagram of metabolic rhythms in *Cyanothece* 51142. This summary indicates the diurnal patterns of N_2-fixation in the dark and photosynthesis in the light. In addition, the proposed changes in photosystem organization and observed changes in accumulation of cyanophycin and glycogen granules are presented.

Controlled Cultivation

For this systems biology project, it is imperative that the cyanobacterial cultures are grown in continuous or semi-continuous mode in illuminated fermentors (photobioreactors). Currently, we grow our batch cultures in shake flask or air-bubbled bottles. In such cultures, the environmental conditions (in particular the light intensity and CO_2 concentration) for cyanobacterial cells are poorly described and ever changing. Clearly, we need to use cultivation technologies that provide continuous monitoring and control of culture conditions so that well-characterized and reproducible samples are produced. During recent months, Sam Kaplan and Himadri Pakrasi have discussed with Yuri Gorby and Jim Fredrickson at PNNL the designing of fiber optics illuminated fermentors for the cultivation of photosynthetic

microbes. The microbial growth facility at PNNL will soon have such reactors in place. Both *Synechocystis* 6803 and *Cyanothece* 51142 strains will be grown in such photobioreactors under defined environmental conditions (light, CO_2, metal concentrations etc.). Most importantly, cultures will be grown in sufficiently large volumes so that aliquots of the same sample are used for parallel transcriptomics, proteomics, ultrastructural and functional studies. Careful analysis of such well-defined data sets will greatly facilitate the network inference and model building activities.

Imaging of Cyanobacterial Cells

Cyanobacteria are unique prokaryotes, since they contain a differentiated membrane system. In addition to the envelope layer consisting of outer membrane (OM), periplasm and plasma membrane (PM), these Gram-negative bacteria have the intracellular chlorophyll-containing thylakoid membrane (TM) involved in oxygenic photosynthesis. Recently, the Pakrasi group has described a method to purify TM and PM from *Synechocystis* 6803 (2). Such membrane preparations will be employed to determine the localization of various key membrane proteins in each membrane, as well as in many of the protein complexes to be studied in this project.

In cyanobacteria, the TM is the site for both photosynthetic and respiratory electron transfer reactions. Thus, the two photosystem complexes, PSI and PSII, function in the TM. However, one of the recent unexpected findings was that many of the central components of both PSI and PSII are also present in the purified PM. Furthermore, these proteins are assembled in the plasma membrane as chlorophyll-containing multiprotein complexes, and are capable of undergoing light-induced charge separation (2). Based on these data, the Pakrasi group has proposed that the PM, and not the TM, is the site for the early steps of biogenesis of these two complexes in cyanobacteria. Such a model predicts movements of partially assembled protein complexes from one membrane system to the other. Such translocations can occur if the thylakoid membranes are contiguous with the plasma membrane, similar to what has long been known for anoxygenic photosynthetic bacteria such as various *Rhodobacter* species. However, our recent EM studies have failed to observe such continuities. Another possibility is that there may be directional flows of membrane vesicles from one membrane to the other. Although the exact mechanism is currently unknown, the dynamic behavior of various protein complexes from one subproteome to another (cycling) has become abundantly clear. A central objective of the proposed research is aimed at probing such a dynamic picture using large-scale proteomic approaches.

Like other bacterial cells, cyanobacteria are relatively small in size. The Sherman group will use advanced electron microscopy techniques to obtain high resolution, *albeit* static pictures of the membrane architecture in both *Synechocystis* 6803 and *Cyanothece* 51142 cells. In addition, we'll access the correlated atomic force microscopy and confocal fluorescence lifetime imaging microscopy expertise in the Cellular Observatory program at EMSL to examine the dynamic behaviors of the membrane systems in live cyanobacterial cells.

Microarray analysis of the cyanobacterial transcriptome

For global gene expression studies, two different versions of DNA microarrays for *Synechocystis* 6803 are available to us. An earlier version is a spotted array (developed jointly by the Burnap and Sherman labs) that has PCR fragments corresponding to the majority of the chromosomal genes on it. A more recent version is a printed 60-mer oligonucleotide microarray (Pakrasi, Aurora and Ghosh groups) that has been produced by the Agilent Co., and includes an additional 460 plasmid-encoded ORFs. Once the genome of *Cyanothece* 51142 is sequenced, we will design printed oligonucleotide arrays for this organism also.

The microarray measurements will be differential expression values of the perturbed state relative to the unperturbed state. We will use different perturbations, such as light intensity, CO_2 concentration, available metal (Fe, Mn, Zn and Cu) concentrations, etc. For *Cyanothece*, we will collect samples at various stages of the circadian cycle.

Analysis of the Dynamic Nature of Cyanobacterial Membrane Proteomes

A *specific objective* here is to enable the studies of comprehensive and quantitative proteomics (both global and subproteomic), and the kinetics associated with specific protein complexes from *Synechocystis* 6803, that has been cultured under experimental conditions designed to follow the incorporation of ^{15}N as a stable isotope pulse label. Operationally, the cell can be subdivided into as many as seven fractions (Fig. 3): outer membrane, periplasm, plasma membrane, cytoplasm, carboxysome, thylakoid membrane and thylakoid lumen. Determining the relative abundance of proteins in *Synechocystis* cells exposed to different environmental conditions is essential to the study of cellular metabolism and responses to environmental changes. In addition to the characterization of the global proteome of *Synechocystis* through the development of the AMT tag database, the characterization of six membrane-associated as well as the cytoplasmic subproteomes derived from subcellular fractionation will be accomplished. Of significant interest is information on both relative protein abundances (*e.g.*, the relative changes resulting from new environmental conditions) and on the absolute abundances of proteins and their modifications. It is noteworthy that ~1100 predicted proteins in *Synechocystis* 6803 are integral membrane proteins, and only ~150 of them have been identified so far. The advanced capabilities in membrane proteomics at EMSL should help us fill this significant gap in the basic knowledge about cyanobacterial membranes.

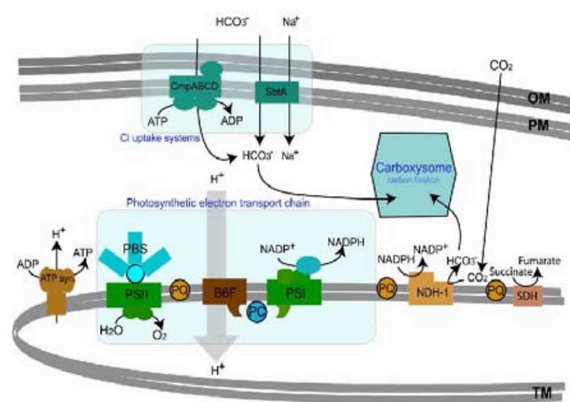


Figure 3: A schematic diagram of a section of a cyanobacterial cell and a few representative protein complexes. The thylakoid membrane (TM) harbors the following proteins: PSII, cyt b6f, PSI, NDH-1 and SDH complexes, whereas the plasma membrane (PM) harbors CmpABCD and SbtA. Also shown is a carboxysome into which HCO_3^- enters, converted to CO_2 and fixed by RuBP carboxylase (Rubisco). ATP synth., ATP synthase; Ci, inorganic carbon; PBS, phycobilisome; PC, plastocyanin; PQ, plastoquinone.

Two important aims in this project are to (a) determine the protein compositions of key membrane protein complexes that are key contributors to energy transduction, photosynthesis, metal acquisition and signal transduction processes in cyanobacterial cells, and (b) examine the possible dynamic time-dependent changes in the relative amounts of different proteins in any given protein complex under altering environmental conditions. Examples of such membrane proteins are: PSI and PSII, phycobilisomes, CO_2/HCO_3^- uptake proteins, ABC metal permeases and the membrane bound sensor components of two-component signal transduction systems. In collaboration with Dick Smith, David Camp and colleagues at EMSL, the Pakrasi group has initiated proteomic studies to (A) generate an AMT map of *Synechocystis* 6803, and (B) analyze the abundances of protein components of His-tagged PSII complex isolated from this organism. In particular, we plan to employ pulse labeling, time course experiments with ^{15}N, to follow the incorporation of ^{15}N into proteins associated with specific protein complexes. These will be important indicators of the kinetics of association, dissociation, synthesis, and turnover of these purified protein complexes. In addition, we plan to develop software tools to support calculation of the rate of ^{15}N incorporation, and quantitative information.

Functional Assays

A variety of assays will be used to determine the functional properties of the cyanobacterial cells under different environmental conditions. For example, room temperature fluorescence kinetics analysis will

be employed for non-invasive measurements of photosynthetic electron transfer processes. ICP-MS will be used to profile the metal contents of various cellular and subcellular preparations. In *Cyanothece*, the dynamic nature of cellular physiology will be monitored by measuring N_2 fixation (a dark function) and photosynthetic O_2 evolution (a light function). Biochemical assays for the accumulation of glycogen (in the light) and accumulation of cyanophycin (an amino acid polymer that accumulates in the dark after N_2 fixation) will determine the major endpoints of central metabolism. Most importantly, we plan to initiate collaborations with specialists in the high-field magnetic resonance facility at EMSL to develop and utilize EPR and NMR-based assays to track critical electron and proton transfer reactions during cyanobacterial energy transduction processes.

Network Inference and Model Building

Effective progress in this project will clearly require interactions between biologists, computational scientists and other scientists and engineers. These interactions are necessary to develop not only the means to acquire the large, complex data sets required to understand the properties of complex biological systems, but also the tools to rapidly analyze those data (3). We have outlined a superb biological system for the analysis of membrane phenomena, but it will require systems biology to develop models, to make predictions that can be tested experimentally and to infer cellular networks using various models. The challenge is to provide methodologies for transforming high-throughput data sets into biological insights about the underlying mechanisms. Although high-throughput assays can provide a quasi-global picture, the details are inherently noisy, and several types of observations must support the conclusions, as we have outlined.

An essential feature of systems biology is the construction of predictive models. We and others have begun to use transcriptome data for predictive modeling of regulatory networks (4). Global analysis of the proteome poses the next great challenge. First, proteins exhibit enormous dynamic ranges—from 1 to a million copies per cell. Second, the patterns of protein expression are dynamic over temporal, spatial and physiological parameters. Third, proteins are modified, they assemble/disassemble into functional machines, and they are degraded—each of these relates to some aspect of their function. This is the reason that proteomics represents such a critical component of this proposal and why modeling these dynamic aspects is so important. The Ghosh and Aurora groups plan to collaborate with Steve Wiley, George Michaels and their colleagues in the Biomolecular Systems Initiative at PNNL for the network generation and model building activities.

Management Plan

Since *at least* five research groups at four different geographical locations will be members of the external science contributor team, the success of this project will depend on the coordination and management of all activities. As the project leader, Himadri Pakrasi will be responsible for the overall coordination, whereas each co-investigator will coordinate the activities in his own research group. Pakrasi will also interact with the coordinator of the PNNL science contributor team to facilitate overall research activities in this project.

To evaluate the scope of such a grand challenge project, a workshop will be hosted at EMSL. We plan to invite other external and PNNL experts in both experimental and computational areas to participate in such a workshop. Based on the discussions during this workshop, one to three other external investigators will be identified as additional co-investigators for such a project. Most importantly, based on such discussions (in particular in brain-storming sessions), and the collective recommendations of the workshop participants, decisions will be made on a focused plan for this collaborative project to answer specific Grand Challenge questions.

References

1. Schneegurt MA, Sherman DM, Nayar S and Sherman LA (1994) Oscillating behavior of carbohydrate granule formation and nitrogen fixation in the cyanobacterium, *Cyanothece* sp. ATCC 51142. J Bacteriol 176: 1586-97.
2. Zak, E., Norling, B., Maitra, R., Huang, F., Andersson, B. and Pakrasi, H.B. (2001) The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. Proc. Natl. Acad. Sci. USA; 98: 13443-48.
3. Hood, L. (2003) Systems biology: integrating technology, biology and computation. *Mech. Ageing Develop*. 124:9-16.
4. Friedman, N. (2004) Inferring cellular networks using probabilistic graphical models. *Science*. 303: 799-805.

C. New Capabilities Resulting in Increased Scientific Impact of EMSL

One of the objectives of the MBGC is to increase the scientific impact of the EMSL National Scientific User Facility. This can be seen through the technology and capability developments that have occurred in the research projects as well as capital equipment investments that were made by the EMSL program to support the research needs of the MBGC.

The MBGC research project consists of a combination of internal (PNNL PIs) and external research activities lead by Himadri Pakrasi. The internal research projects contribute to the MBGC research objectives as well as develop new capabilities and methods. These new capabilities and methods are deployed in the EMSL and made available to the user community. This has resulted in significant advancement in EMSL capability and scientific impact of the EMSL. The internal research projects and each project PI are listed below:

- *First Protein Structures from Cyanothece 51142*, PI GW Buchko
- *Morphological, Functional and Redox Studies of Synechocystis 6803 and Cyanothece 51142 sp. Bacterial Membrane Complexes by Methods of Electron Microscopy*, PI AC Dohnalkova
- *Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function: Proteomic, Metabolomic, and Metallomic Characterization*, PIs JM Jacobs, TO Metz, DG, DW Koppenaal
- *Dynamic Changes in Molecular Expressions and Interactions along the Circadian Rhythm*, PI Galya Orr
- *Computational Biology and Bioinformatics Tools for Understanding the Role of Membrane Proteins in Diurnal and Circadian Processes of Prokaryotes*, PI CS Oehmen
- *Data Assimilation, Visualization, and Mining*, PI ES Peterson
- *Ecophysiological Investigation of Cyanobacteria Using Controlled Cultivation,* PI JC Scholten

Each of these projects has had an impact on the scientific capabilities made available to the EMSL user community and in the following sections are examples of how the capabilities developed are drawing new research to EMSL.

Proteomics-Assisted Genome Annotation

Sequencing and Annotation of the Genome of *Cyanothece* 51142 was completed with the aid of high throughput proteomics and computational capabilities at EMSL. In order to support the throughput needs of this research and to support future work in this area additional proteomics analysis instrumentation was purchased by EMSL. This includes a Linear Ion Trap instrument and a high performance mass spectrometry instrumentation used for quantitative studies. In addition to the proteomics capability ScalaBLAST, a high-throughput sequence analysis engine, was used to perform large-scale BLAST calculations using the EMSL supercomputer, dramatically speeding up the BLAST calculations required to complete the *Cyanothece* annotation. Over 17000 cpu hours were consumed by this project (over 2 cpu years). Applying these EMSL capabilities to annotation of the *Cyanothece* genome enabled the reclassification of nearly 25% of the predicted hypothetical proteins.

Another impact from this research is ScalaBLAST has been made available to the Joint Genome Institute (JGI) via a new Computational Grand Challenge collaboration. Through this collaboration,

ScalaBLAST will be used to process the entire collection of microbial genomes using the EMSL Molecular Science Computing Facility (MSCF) supercomputer, MPP2. JGI is using these results to provide a browsable pre-computed sequence homology tool allowing users to find highly conserved proteins across all sequenced microbes as well as a collection of viral and eukaryotic genomes. This resource is accessible by a wide scientific user community. This partnership illustrates the impact EMSL can make through technology advances enabled, in part, by the EMSL MBGC program and its unique facilities like the MSCF.

These developments were enabled through the research in following two projects:

- *Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function: Proteomic, Metabolomic, and Metallomic Characterization*, PIs JM Jacobs, TO Metz, DG, DW Koppenaal
- *Computational Biology and Bioinformatics Tools for Understanding the Role of Membrane Proteins in Diurnal and Circadian Processes of Prokaryotes*, PI CS Oehmen

Development of Proteomics and Metabolomics Capabilities

Cyanothece is an intriguing organism that uses circadian rhythms to perform photosynthesis during the day and nitrogen fixation at night. Proteins that play significant roles in the processes are being identified by proteomic, and metabolomic data collected at EMSL.

LTQ Orbitrap high performance mass spectrometer.

The proteomics capability developments under the MBGC are having, and will continue to have, broad impact across many studies that will utilize the EMSL User program. Achieving an understanding of the dynamic nature of static protein abundance represents the next step in proteomic measurements and will become integral in correlating actual protein synthesis with mRNA expression information. We have already facilitated other projects to begin integrating this type of quantitative approach in their pipeline of analyses and see an increasing demand for this type of analysis and expertise.

The metabolomics component of this research endeavors to advance the understanding of those metabolites whose concentrations are under circadian and/or light/dark control during diurnal growth of *Cyanothece*. A hybrid ion-trap-Orbitrap mass spectrometer (LTQ-Orbitrap, ThermoFisher) with MS/MS and high-mass resolution capability was purchased and utilized for global metabolomic and lipidomics studies of *Cyanothece* grown in light/dark cycles. Additionally a new metabolomics capability has been added to the overall EMSL suite of capabilities. This includes a new automated high-pressure LC system devoted to separations of either water- or lipid-soluble metabolites, as well as the overall approach for downstream data processing and analysis (discussed above)

Metabolomics automated high-presure LC system

These developments were enabled through the research under the following two projects:

- *Systems Analysis of the Dynamics of Membrane Architecture, Composition, and Function: Proteomic, Metabolomic, and Metallomic Characterization*, PIs JM Jacobs, TO Metz, DG, DW Koppenaal

- *Ecophysiological Investigation of Cyanobacteria Using Controlled Cultivation,* PI JC Scholten

Development of Structural Biology Capabilities

The maintenance of a strong structural biology presence at EMSL for Users was one of the major goals of the MBGC. The publication of manuscripts describing two protein structures, Rfr32 and Rfr23, attests to the successful impact the Grand Challenge had on this goal.

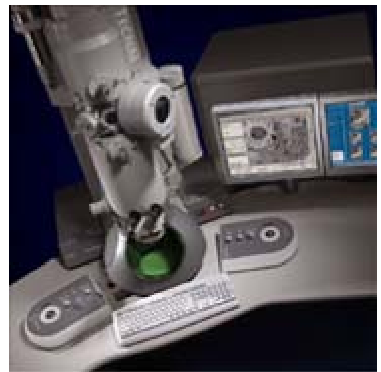
The crystallography process may be divided into three steps: crystal growth, data collection, and data analysis. Of these steps, establishing an expertise and laboratory environment to grow, harvest, and store macromolecular crystals (under liquid nitrogen) is the most important step in protein crystallography (you cannot do anything without a good crystal). The crystals for Rfr32, Rfr23, and DUF683 were grown in a laboratory in EMSL. The second step, data collection, was facilitated with the EMSL single-molecule X-ray diffractometer that made it possible to screen some crystals before sending them to synchrotrons. Such a screening capability makes it possible to i) determine if the crystal diffracts (eg: is it really is a crystal?), and to ii) send only the best crystals to synchrotrons (synchrotron time is at a premium). The third step, data analysis, requires the computational facilities and raw data. The last process essential involves spending time working through problems to become efficient. The structures for Rfr32 and Rfr23 were refined using software present at EMSL.

These developments were enabled through the research under the following project:

- *First Protein Structures from Cyanothece 51142*, PI GW Buchko

New imaging capabilities in EMSL

EMSL acquired a new devoted biological TEM with two unique capabilities of TEM tomography and a cryostage. After the capabilities were developed and demonstrated, EMSL has received a remarkable response to an annual capability call for user proposals that will demonstrated this new capability: 8 groups interested in starting collaborative projects using these new TEM capabilities, including a group from Korea. By bringing these new capabilities on board, and we will strive to support these and other top-science projects.



Biological TEM with cryostage and tomography capability.

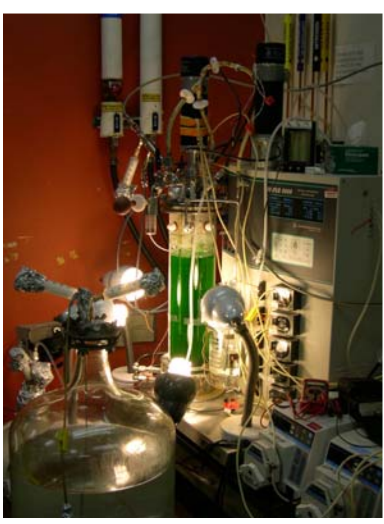
Non-conventional fluorescence laser microscopy and spectroscopy techniques have been developed enhancing EMSL capabilities in molecular and cellular imaging. The techniques include single-molecule fluorescence, fluorescence lifetime (FLIM) and quantitative fluorescence resonance energy transfer (FRET) imaging and spectroscopy in real time and cellular space. In addition to adapting conventional fluorescence methods to the investigation of cyanobacteria, the new approaches have been specifically developed and applied to investigate molecular processes in the intact cyanobacterial cell. The technical developments, which strengthened the capabilities offered in EMSL, will be applied to other user programs that focus on bacterial cells and other organisms that are of interest to the environmental impact of energy production or nanomaterial exposure. The MBGC project has established a solid collaboration between investigators at PNNL, Washington and Purdue Universities, and has familiarized investigators across several institutes with EMSL strong capabilities in molecular and cellular imaging. The project has established the technical and scientific foundation for the continuation and

expansion of this and other projects centered on the CO_2 uptake and concentrating systems in cyanobacteria.
These developments were enabled through the research in following two projects:

- *Morphological, Functional and Redox Studies of Synechocystis 6803 and Cyanothece 51142 sp. Bacterial Membrane Complexes by Methods of Electron Microscopy, PI AC Dohnalkova*
- *Dynamic Changes in Molecular Expressions and Interactions along the Circadian Rhythm, PI Galya Orr*

Controlled Cultivation Capabilities

Development and operation of a fully instrumented photobioreactor that allows for cultivating Cyanobacteria under well defined and controlled conditions is critical to the MBGC research and represents new capablity. This capability not only provides a more complete understanding of the physiological response of cells to changes in environmental conditions, but also provides high quality, reproducible samples that will enhance the effectiveness of biological information obtained from advanced analytical technologies. EMSL invested capital equipment funds to allow the development of three systems, located at PNNL, Washington University in Saint Louis, and the final system in Purdue University. This capability allows controlled growth experiments performed at different institutions with cyanobacteria to be done with a constant set of conditions. This controlled cultivation allows better interpretation and integration of experimental data generated by project collaborators.



Photobioreactor developed at EMSL and deployed at three sites.

These developments were enabled through the research under the following project:

- *Ecophysiological Investigation of Cyanobacteria Using Controlled Cultivation,* PI JC Scholten

Data Management for Systems Biology

Because this MBGC research involved collaboration with multiple institutions we had to develop a way to provide the key EMSL capabilities to those institutions. A key part of that was providing access to the Proteomics data managed in PNNLs PRISM system to external collaborations through the Bioinformatics Resource Manager (BRM) software. We set the stage for the ability for other collaborators on other projects to also be able to access their PRISM data directly through BRM. This was possible because of coordinated efforts with EMSL's CaNS team and our project to work through security and technical issues to make this process possible and seamless to the users. An Experimental Data Management System was enhanced and is being utilized to manage all of the experimental data generated by this collaborative research project. This capability allows investigators to share their data across multiple institutions in a secure environment.

These developments were enabled through the research under the following project:

- *Data Assimilation, Visualization, and Mining*, PI ES Peterson

EMSL Capital Equipment Investments

The EMSL National Scientific User Facility receives an annual operating budget as well as an annual capital equipment budget use to refresh and enhance its capabilities. The capital investments from EMSL are designed to increase the user facilities scientific impact by developing capabilities that will draw high profile users to the facility. EMSL has shown its commitment and support of the MBGC by making significant capital investments aimed at the scientific goals of this research program. Listed below are a summary of the investments described above that have created new or improved capabilities at EMSL and are supporting the MBGC.

Instrument	Cost	Investment year
Cryo TEM,	$860K	2005
LTQ Orbitrap	$247K	2005
3 bioreators	$220K	2005
LTQ Linear Ion Trap	$275K	2005

This is a total investment of $1.6 million in new and improved capabilities that have enabled the development of new user capabilities as a result of the MBGC research and development.

D. Publications/Presentations/Awards

Publications

1. Buchko GW, S Ni, H Robinson, EA Welsh, HB Pakrasi, and MA Kennedy. 2006. "Characterization of two potentially universal turn motifs that shape the repeated five-residues fold - Crystal structure of a lumenal pentapeptide repeat protein from Cyanothece 51142." *Protein Science,* 15(11) pp. 2579-2595
2. Liberton M., RH Berg, J Heuser, R Roth, and HB Pakrasi. 2006. "Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803." *Protoplasma*, 227: 129-138 doi:10.1007/s00709-006-0145-7
3. Koropatkin NM, HB Pakrasi, and TJ Smith. 2006. "Atomic structure of nitrate-binding protein crucial for photosynthetic productivity." *Proceedings of the National Academy of Sciences of the United States of America,* 103(26):9820-9825. doi:10.1073/pnas.060251703
4. Buchko GW, H Robinson, S Ni, HB Pakrasi, and MA Kennedy. 2006. "Cloning, expression, crystallization and preliminary crystallographic analysis of a pentapeptide-repeat protein (Rfr23) from bacterium *Cyanothece* 51142." *Acta Crystallographica. Section F,* 62(12):1251-1254. doi:10.1107/S174430910604663X
5. Oehmen CS, and J Nieplocha. 2006. "ScalaBLAST: A Scalable Implementation of BLAST for High Performance Data-Intensive Bioinformatics Analysis." *IEEE Transactions on Parallel and Distributed Systems,* 17(8) pp. 740-749
6. Dohnalkova A, CL Bilskis, and DW Kennedy. 2006. "TEM Study of Manganese Biosorption by Cyanobacterium Synechocystis 6803." In *Microscopy and Microanalysis 2006, Chicago, Illinois, USA, July 30 – August 3, 2005. Published in Microscopy and Microanalysis*, vol. 12, no. Supplement S02, pp. 444-445. Cambridge University Press, Cambridge, United Kingdom. doi:10.1017/S1431927606064051
7. Ma W, Y Deng, T Ogawa, and H Mi. 2006. "Active NDH-1 complexes from the cyanobacterium Synechocystis sp. strain PCC 6803." *Plant and Cell Physiology,* 47(10) pp. 1432-1436
8. Taylor RC, A Shah, C Treatman and M Blevins. 2006. "SEBINI: Software Environment for Biological Network Inference." *Bioinformatics*, 22(21) pp. 2706-2708 doi:10.1093/bioinformatics/btl444
9. Liberton M and HB Pakrasi. 2007. "Membrane Systems in Cyanobacteria." In 'The Cyanobacteria: Molecular Biology, Genomics and Evolution' (Herrero, A. and Flores, E., eds.), pp. 271-287, Horizon Scientific Press, Norwich, UK
10. Kaplan A, M Hagemann, H Bauwe, S Kahlon, and T Ogawa. 2007. "Carbon acquisition by cyanobacteria; mechanisms, comparative genomics and evolution." In 'The Cyanobacteria: Molecular Biology, Genomics and Evolution' (Herrero A. and Flores E., eds.), Ch. 12, Horizon Scientific Press, Norwich, UK
11. Koropatkin N, DW Koppenaal, HB Pakrasi, and TJ Smith. 2007. "The Structure of a Cyanobacterial Bicarbonate Transport Protein, CmpA." *Journal of Biological Chemistry,* 282(4) pp. 2606-2614 doi:10.1074/jbc.M610222200
12. Aurora, R, Y Hihara, AK Singh and HB Pakrasi. 2007. A Network of Genes Regulated by Light in Cyanobacteria. *OMICS: A Journal of Integrative Biology*, 11(2) pp.166-185 doi:10.1089/omi.2007.4323

13. Wei, B., AM Randich, M Bhattacharyya-Pakrasi, HB Pakrasi, and TJ Smith. 2007. "Possible Regulatory Role for the Histidine-Rich Loop in the Zinc Transport Protein, ZnuA." *Biochemistry*, 46(30) pp. 8734-8743 doi:10.1021/bi70763w
14. Webb-Robertson BM, ES Peterson, M Singhal, KR Klicker, CS Oehmen, JN Adkins, and SL Havre. 2007. "PQuad - A visual analysis platform for proteomic data exploration of microbial organism." *Bioinformatics,* 23 (13):1705-1707 doi:10.1093/bioinformatics/btm132
15. Ding J, CM Sorensen, Q Zhang, H Jiang, N Jaitly, EA Livesay, Y Shen, RD Smith, and T Metz. 2007. "Capillary LC Coupled with High-Mass Measurement Accuracy Mass Spectrometry for Metabolic Profiling." *Analytical Chemistry*, 79(16) pp. 6081-6093 doi:10.1021/ac070080q
16. Ogawa T, and H Mi. 2007 "Cyanobacterial NADPH dehydrogenase complexes." *Photosynthesis Research,* 93 (1-3):69-77 doi:10.1007/s11120-006-9128-y
17. Dohnalkova A, R Mendoza, C Gassman, and C Bilskis. 2007 "Creating 3D reconstruction of cyanobacterium *Cyanothece* sp. by alignment of serial TEM tomograms from consecutive plastic sections." *Microsc. Microanal.* 13(Supp 2), 1338-1339
18. Buchko GW, H Robinson, HB Pakrasi, and MA Kennedy. 2007. "Insights into the structural variation between pentapeptide repeat proteins - Crystal structure of Rfr23 from Cyanothece 51142." *(In Press, J. Struct. Biology)*
19. Metz T, Q Zhang, JS Page, Y Shen, SJ Callister, JM Jacobs, and RD Smith. 2007. "The future of liquid chromatography-mass spectrometry in metabolic profiling and metabolomic studies for biomarker discovery." *Biomarkers in Medicine* 1(1):159-185. doi: 10.2217/17520363.1.1.159
20. Koropatkin N, AM Randich, M Bhattacharya-Pakrasi, HB Pakrasi, and TJ Smith. 2007. "The Structure of the Iron Binding Protein, FutA1, from Synechocystis 6803." *J. Biol Chem.*, 282(37): pp. 27468-27477
21. Xu M, T Ogawa, HB Pakrasi, and H Mi. 2007. "Identification of the CupB Protein Involved in Constitutive CO_2-uptake in the Cyanobacterium, Synechocystis sp. Strain PCC 6803." *(Submitted J. Bacteriology)*
22. Buchko GW, H Robinson, and HJ Sofia. 2007. "Expression, solution properties, crystallization, and preliminary crystallographic analysis of DUF683 from Cyanothece sp. 51142 – a hypothetical protein associated with nitrogen fixation." *(Acta Crst. Sec. F, in preparation)*
23. Welsh EA, ML Liberton, J Stockel, H Sato, T Loh, C Wang, A Wollam, RS Fulton, SW Clifton, JM Jacobs, R Aurora, LA Sherman, RD Smith, RK Wilson, and HB Pakrasi. 2007. "Genome sequence of Cyanothece 51142 shines new light on the circadian life style of a cyanobacterium." *(Submitted Nature)*
24. Bhattacharyya-Pakrasi M, T Ogawa, TJ Smith, X-M Si, and HB Pakrasi. (2007) FEMS Letters (*Submitted FEMS Letters)*

Presentations

1. Sherman, L. A., 2005 "The unicellular, diazotrophic cyanobacterium Cyanothece--genomic sequence and metabolic rhythms." Marine Cyanobacteria: Evolution, Function and Genomes. Stockholm, Sweden, August 24-27, 2005.

2. Koppenaal DW, and CJ Barinaga. 2005. "Metallomics and ICPMS - new opportunities for unique instrumentation and applications." Presented by David W. Koppenaal (Invited Speaker) at PACIFICHEM 2005, Honolulu, HI on December 15, 2005.
3. Campbell AA, JM Kuperberg, HB Pakrasi, and JM Zachara. 2006. "Unique Tools for Unique Science at a DOE National Scientific User Facility." Presented by Allison A. Campbell, J. Michael Kuperberg, Himadri Pakrasi, John M. Zachara at AAAS Annual Meeting, St. Louis, MO on February 17, 2006.
4. Welsh, E. A., Sherman, L. A. and Pakrasi, H. B. 2006. Genome Sequence Analysis of Cyanothece sp. ATCC 51142, a Unicellular Nitrogen-fixing Cyanobacterium. 1st Annual Finishing the Genome in the Future Meeting, Santa Fe, NM, May 4-5, 2006.
5. Dohnalkova A, CL Bilskis, and DW Kennedy. 2006. "TEM Study of Manganese Biosorption by Cyanobacterium Synechocystis 6803." Presented by Alice Dohnalkova at Microscopy & Microanalysis 2006, Chicago, IL on August 3, 2006.
6. Min, H-T. and Sherman, L. A. 2006. Genetic Transformation and Mutagenesis in the Unicellular Diazotrophic Cyanobacterium Cyanothece sp. ATCC 51142. 12th Inter-national Symposium on Phototrophic Prokaryotes, Pau, France, Aug 27-Sept. 1, 2006.
7. Oehmen CS. 2006. "Dealing with the sequence data avalanche." Presented by Chris Oehmen (Invited Speaker) at N/A, Richland, WA on September 14, 2006.
8. Koppenaal DW, CJ Barinaga, SJ Ray, DA Rogers, GM Hieftje, ML Liberton, and HB Pakrasi. 2006. "Metallomics - New Analytical Techniques for the Post-Genomic Era." Presented by Steven Ray at 33rd Annual Conference of the Federation of Analytical Chemistry and Spectroscopy Studies, Lake Buena Vista, FL on September 27, 2006.
9. Oehmen CS, J Nieplocha, and HJ Sofia. 2006. "ScalaBLAST: High-throughput BLAST calculations for genome-scale bioinformatics." Presented by Christopher Oehmen at Presentation to DOE Program Manager, Richland, WA on October 18, 2006.
10. Buchko GW. 2006. "Tales and Adventures on the Road to the Protein Data Bank - The Chronicles of DR0079 and Rfr32." Presented by Garry Buchko (Invited Speaker) at Washington State University, Richland, WA on November 2, 2006.
11. Welsh, E. A., Liberton, M., Stockek, J., Pakrasi, H.B., Min, H., Toepel, J. and Sherman, L.A. 2007. Genomics and Systems Biology analysis of Cyanothece sp. ATCC 51142, a unicellular, diazotrophic cyanobacterium with robust metabolic rhythms. Plenary lecture, 15th International Nitrogen Fixation Congress, Cape Town, South Africa, Jan. 21-28, 2007.
12. Hongtao Min and Louis A. Sherman. 2007. Genomics and Genetic Transformation in the unicellular, diazotrophic cyanobacterium, Cyanothece sp. ATCC 51142. 15th International Nitrogen Fixation Congress, Cape Town, South Africa, Jan. 21-28, 2007.
13. Pakrasi, H.B. (2007) Membrane Biology Grand Challenge: Engagement of a DOE National User Facility for Integrative Studies of an Energy Producing Photosynthetic Bacterium, Plenary talk at the Joint Genomics: GTL Contractor-Grantee Workshop V and Inter-Agency Conference on Metabolic Engineering 2007, Bethesda, MD. Feb 11-14, 2007
14. Felmy AR. 2007. "Andy Felmy's ERSP poster." Presented by Andy Felmy (Invited Speaker) at ERSP Annual PI meeting, Landsdowne, VA on April 16, 2007.
15. Koppenaal DW, and DK Lemon. 2007. "Aligning Strategy and LDRD Investments at Pacific Northwest National Laboratory." Presented by David W. Koppenaal, Douglas K. Lemon (Invited Speaker) at DOE-SC LDRD Program Review, Richland, WA on May 16, 2007.
16. Koppenaal DW, CJ Barinaga, JA Campbell, HB Pakrasi, and ML Liberton. 2007. "Metallomics - New Techniques for a New -Ohmics." Presented by David Koppenaal (Invited Speaker) at 39th Central Regional ACS Meeting, Covington, KY on May 21, 2007.

17. Min H, and LA Sherman. 2007. "Genetic transformation and recombination in the unicellular, diazotrophic cyanobacterium Cyanothece sp. ATCC 51142." Presented by Hongtao Min at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 6, 2007.
18. Toepel J, and LA Sherman. 2007. "Metabolic regulation of the cyanobacterium Cyanothece sp. ATCC 51142 when grown in light/dark or in continuous light." Presented by Jorg Toepel at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 6, 2007.
19. Xu M, H Mi, AK Singh, HB Pakrasi, and T Ogawa. 2007. "Role of carbon cycling in protecting cyanobacterial cells from high light stress by dissipating excess light energy." Presented by Teruo Ogawa at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 6, 2007.
20. Welsh EA, ML Liberton, J Stockel, H Sato, T Loh, C Wang, A Wollam, RS Fulton, SW Clifton, JM Jacobs, R Aurora, LA Sherman, RD Smith, RK Wilson, and HB Pakrasi. 2007. "Proteomics Assisted Genome Sequence of Cyanothece 51142, a Unicellular N_2-fixing Cyanobacterium." Presented by Eric Welsh at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 6, 2007.
21. Stockel J, EA Welsh, R Aurora, ML Liberton, H Sato, LA Sherman, and HB Pakrasi. 2007. "Robust oscillation of metabolic processes: global gene expression in Cyanothece sp. ATCC 51142." Presented by Jana Stockel at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 7, 2007.
22. Koropatkin N, AM Randich, M Bhattacharyya-Pakrasi, DW Koppenaal, HB Pakrasi, and TJ Smith. 2007. "Specificity and Regulation of ABC Transporters." Presented by Nicole Koropatkin at 9th Cyanobacterial Molecular Biology Workshop, Lake Delavan, WI on June 6, 2007.
23. Buchko GW. 2007. "Towards understanding circadian rhythms in the cyanobacterium Cyanothece - Insights into the structure of pentapeptide repeat proteins and DUF683." Presented by Garry Buchko (Invited Speaker) at Chemistry Department at Washington State University, Pullman, WA on June 7, 2007.
24. Welsh, E. A., many collaborators and Louis A. Sherman. 2007. Complete genome Sequence of Cyanothece sp. ATCC 51142. DOE Joint Initiative Meeting "Finishing Genomes in the Future." Santa Fe, NM, June 16-20, 2007.
25. Toepel, J. and L.A. Sherman. 2007. Metabolic regulation of the cyanobacterium Cyanothece sp. ATCC 51142 when grown in light-dark or continuous light. IV European Phycological Conference, Oviedo, Spain, July 22-26, 2007.
26. Scholten JC. 2007. "Light and Dark Metabolism of Cyanothece Strain ATCC 51142 During Continuous Cultivation." Presented by Hans Scholten at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.
27. Jacobs JM, EA Welsh, J Stockel, ML Liberton, MA Gritsenko, CD Nicora, HB Pakrasi, and RD Smith. 2007. "Dynamic Quantitative Characterization of Cyanothece sp. ATCC 51142." Presented by Jon Jacobs at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.
28. Metz T, J Ding, Q Zhang, CM Sorensen, JC Scholten, GW Buchko, JA Campbell, J Stockel, EA Welsh, J Toepel, LA Sherman, and HB Pakrasi. 2007. "LC-MS-based metabolic profiling studies of Cyanothece sp. ATCC 51142." Presented by Tom Metz at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.
29. Buchko GW. 2007. "First Protein Structures from Cyanothece 51142." Presented by Garry Buchko at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.

30. Dohnalkova A, WJ Harvey, EA Hill, and CL Bilskis. 2007. "Creating 3D Reconstruction of Cyanothece by TEM tomography." Presented by Alice Dohnalkova at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.
31. Oehmen CS, JE McDermott, and ES Peterson. 2007. "Systems Approach to Understanding Molecular Mechanisms in the Light-Dark Cycle of Cyanothece sp. 51142." Presented by Chris Oehmen, Jason Mcdermott at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.
32. Orr G, DJ Panther, and JL Phillips. 2007. "Dynamic Changes in Molecular Expressions and Interactions along the Circadian Rhythm." Presented by Galya Orr at EMSL MBGC Advisory Committee Review, Richland, WA on August 21, 2007.

Awards

1. Alice Dohnalkova received The Diatome US First Place award for "Creating 3D reconstruction of cyanobacterium *Cyanothece* sp. by alignment of serial TEM tomograms from consecutive plastic sections". The award was presented to Alice by the president of the Microscopy Society of America at the Annual meeting in Ft Lauderdale (Aug 2007).

E. Advisory Committee

CHAIR - Dr. Jack Johnson, SCRIPPS
Department of Molecular Biology
10550, N. Torrey Pines Road
La Jolla, CA 92037
Ph. 858-784-2947
Fx. 858-784-8660
jackj@scripps.edu

Professor Mark Ellisman, University of California, San Diego
Center for Research in Biological Systems
9500 Gilman Drive, MC0608
La Jolla, CA 92093-0608
Ph. 585-534-2251
Fx. 858-534-7497
mark@ncmir.ucsd.edu

Dr. James Fredrickson, Pacific Northwest National Laboratory
Biological Sciences Division
PO Box 999, MSIN P7-54
Ph. 509-376-7063
Fx, 509-376-4909
jim.fredrickson@pnl.gov

Dr. Gary Siuzdak, SCRIPPS
Center for Mass Spectrometry
10550 North Torrey Pines Rd, BCC007
La Jolla, CA 92037
Ph. 585-784-9415
Fx. 858-784-9496
siuzdak@scripps.edu

Professor James Liao, University of California
5531 Boelter Hall
420 Westwood Plaza
Los Angeles, CA 90095
Ph. 310-825-1656
Fx. 310-206-4107
liaoj@seas.ucla.edu

F. Advisory Committee Report

Membrane Biology Grand Challenge (MBGC)
LDRD Review – August 21 – 22, 2007

Introduction

Charge Responses

1. Scientific Impact of EMSL

Findings

- The Advisory Committee has concluded that the MBGC has made excellent progress towards its stated research goals given the relatively short period of time this project has been underway. The research is having a very positive impact on biology in general and on EMSL's reputation within the scientific community. The AC also concluded that there is potential for even greater impact and EMSL recognition in the future. Because of the challenging nature of the research and its extensive complexity – beginning with an organism that no one knew much about – the investigators have made very good technical progress and are well-positioned to have a major impact. The experimental system (i.e., *Cyanothece*) and associated scientific questions (changes in membrane architecture and composition during light-dark cycling) are well-suited to EMSL's capabilities in biology and the scientific team that has been assembled.
- The MBGC has clearly expanded the general awareness and use of EMSL's capabilities in a synergistic manner to address a challenging and potentially impactful scientific problem.
- There has been "aggressive cooperation" among all members of the MBGC team including EMSL management.

Comments

- The AC members appreciated the extensive connections between external and internal contributors and it is clear they have established excellent working relationships. The committee felt that there was inadequate preparation for this review and that many of the connections, the flow of information, and even the specific set of questions being addressed were not made clear in the individual presentations. It was only after the question and answer session that many of these issues were clarified.
- The AC commends the open communications that exist among all members including the sharing of ideas and relatively raw data. A great sense of trust obviously exists among the various members that have come together as a team in a relatively short period of time. EMSL and the MBGC leaders have done an excellent job fostering teamwork and encouraging information sharing among team members. This has also had the advantage of establishing good interactions among staff at PNNL who previously have not worked together. This has the

benefit of creating a team and set of capabilities that can be applied to other problems requiring multidisciplinary and integrated capabilities in the future.

Recommendations

- Develop a synthesis article to help raise the visibility of the scientific impact of EMSL and this grand challenge. An article that takes a "systems" view of the transition between the light and dark cycles and integrates gene expression, proteomics, and metabolomics data, at minimum, would make an excellent target.
- The AC also suggests that the team consider using a bottom-up approach to allow proteomics, metabolomics, and informatics to drive the experimental design. Given the limited time for the MBGC there is a need for the team to stay focused on an integrated set of experiments and analyses.
- Consider a publication in *Microbe* or a similar type of trade publication to gain additional exposure of the grand challenge both in terms of the science issue being addressed but also describing the grand challenge operational model and how EMSL capabilities are being used in a synergistic manner.
- The MBGC team is encouraged to formally develop the connections between external and internal major players of this grand challenge, and the relationships within EMSL. Future reviews and documents describing the MBGC should contain this information as an introduction.

2. Scientific and Technology Outcomes and Progress

Findings

- Given the time and funding constraints, science and technology progress has been outstanding. The model system and the questions being asked are directly relevant to DOE's missions in energy and environment. Although some areas are more mature than others, all of the essential capabilities are in place and some initial experiments, albeit preliminary, have yielded some exciting results regarding insights into how *Cyanothece* "rewires" itself for growth and metabolism during light and dark phases.
- Standardization across laboratories is also very good. For example, the emphasis on the standardization of cultivation conditions across laboratories and the use of bioreactors to obtain cultures under highly controlled conditions is to be commended and has had a positive impact on the quality of the gene expression and proteomics data.
- The major science objectives are being realized as expected; MBGC priorities and DOE/EMSL's expectations as described in the original call for proposals are consistent and being met.
- Some project successes and highlights where new insights being gained by the integrated efforts were not presented but would have been helpful. In general the individual presentations were not connected making it difficult for the AC to see how they fit into the bigger picture. Dry runs and additional preparation probably would have resulted in better communication of the overall project to the AC.

Comments

- Overall, the AC was extremely enthused about scientific and technology process.

Recommendations

- For next review, possibly a very important review by BERAC to assess the outcomes of the MBGC, the team needs to spend more time in preparation and conduct at least one dry run of the presentations.
- The AC found the EMSL grand challenge model to be extremely attractive for addressing difficult scientific issues that are not readily resolved via a single investigator project. The use of EMSL in such a synergistic manner is unique among the DOE user facilities and clearly has the potential to make even greater contributions to DOE science program to address major national problems in energy and environment. The AC encourages DOE to continue to support and consider expanding the grand challenge concept. EMSL, DOE, and its scientific mission will greatly benefit from enhancement of support and continuity.

3. Strategy and Objectives

Findings

- The scientific approach matches well with the topic. Good progress has been made on bringing the various technical pieces together to address the problem.
- EMSL and PNNL in general are fostering the cultural environment necessary for multidisciplinary science.
- A bottom-up approach would allow proteomics, metabolomics, metabolics and informatics done on light-dark cycle experiments to drive the design of more targeted experiments in the future. Focusing further structure and composition of the thylakoid membrane, its composition, biogenesis, and relationship to the cell and outer membranes would also be appropriate.

Comments

- It was not certain how some of the technical pieces, such aspects of cell imaging, were contributing to the overall MBGC and how they would add to the integrated knowledge of the light-dark cycle of *Cyanothece*.
- Publication of integrated results in a high-profile peer-review journal will be essential to the overall successes of this grand challenge.

Recommendations

- The project leaders should ensure that all team members and subtasks have the potential to contribute to the overall integrated effort or the efforts should be re-directed to tasks that will.

- It will be helpful to have greater communication in the imaging area of the effort. Identification and elimination of redundancies in this area will allow better focus and greater productivity. While technology development is part of EMSL's charge, it should not be done at the expense of the scientific mission of the MBGC. The leadership should agree on the distribution of effort that is considered to be part of MBGC. The needs of the MBGC may dictate additional technology development, but that should probably be done within EMSL and not be directly part of the MBGC.

G. Participants Contact Information

Principal Investigators

Himadri Pakrasi
Principal Investigator
Washington Unversity in St. Louis
One Brookings Drive
Campus Box 1137
St. Louis, MO 63130-4899
Ph. 314-935-6853 / Fx. 314-935-6803
pakrasi@biology2.wustl.edu

David Koppenaal
Co-Principal Investigator
Pacific Northwest National Laboratory
EMSL
P.O. Box 999, MS K8-98
Richland, WA 99352
Ph. 509-376-0368 / Fx. 509-376-1919
david.koppenaal@pnl.gov

Washington University in St. Louis Collaborators

Bijoy Ghosh
Ph. 314-935-5818
ghosh@netra.wustl.edu

Michelle Liberton
Ph. 314-935-8133
mliberton@biology2.wustl.edu

Jana Stockel
Ph. 314-935-4964
stockel@biology2.wustl.edu

Eric Welsh
Ph. 314-935-8133
ewelsh@biology.wustl.edu

Pacific Northwest National Laboratory Collaborators

Gordon Anderson
Ph. 509-376-9558
gordon.anderson@pnl.gov

Christina Bilskis
Ph. 509-376-4328
christina.bilskis@pnl.gov

Garry Buchko
Ph. 509-373-9826
garry.buchko@pnl.gov

Jie Ding
jie.ding@pnl.gov

Alice Dohnalkova
Ph. 509-376-3908
alice.dohnalkova@pnl.gov

Allison Campbell
Ph. 509-376-6688
allison.campbell@pnl.gov

James A. Campbell
Ph. 509-376-0899
james.campbell@pnl.gov

Yuri Gorby
yuri.gorby@pnl.gov

Eric Hill
Ph. 509-372-3949
eric.hill@pnl.gov

Jon Jacobs
Ph. 509-376-9702
jon.jacobs@pnl.gov

Peter Lu
peter.lu@pnl.gov

Lee Ann McCue
Ph.509-375-2912
leeann.mccue@pnl.gov

Jason McDermott
Ph. 509-372-4360
jason.mcdermott@pnl.gov

Shuisond Ni
shuisong.ni@pnl.gov

Michael Kennedy
michael.kennedy@pnl.gov

Tom Metz
Ph. 509-376-8333
thomas.metz@pnl.gov

Christopher Oehmen
Ph. 509-376-2038
christopher.oehmen@pnl.gov

Galya Orr
Ph. 509-376-9592
galya.orr@pnl.gov

David Panther
Ph. 509-376-5050
david.panther@pnl.gov

Elena Peterson
Ph. 509-372-4573
elena@pnl.gov

Johannes Scholten
Ph. 509-376-1939
johannes.scholten@pnl.gov

Heidi Sofia
Ph. 509-372-4216
heidi.sofia@pnl.gov

Christina Sorensen
Ph. 509-376-6712
christina.sorensen@pnl.gov

Richard Smith
Ph. 509-376-0723
dick.smith@pnl.gov

Ronald Taylor
Ph. 509-372-6568
ronald.taylor@pnl.gov

Katrina Waters
Ph. 509-375-3907
katrina.waters@pnl.gov

Steve Wiley
Ph. 509-373-6218
steven.wiley@pnl.gov

Bobbie-Jo Webb-Robertson
Ph. 509-375-2292
bobbie-jo.webb-robertson@pnl.gov

Qibin Zhang
Ph. 509-373-9682
qibin.zhang@pnl.gov

Purdue University Collaborators

Louis Sherman
Ph. 765-494-8106
lsherman@purdue.edu

Hongtao Min
Ph. 765-494-8106
mhongtao@purdue.edu

Debra Sherman
Ph. 509-494-6666

dsherman@purdue.edu

Tina Summerfield
Ph. 765-494-0560
tsummer@purdue.edu

Jorg Toepel
Ph. 765-494-0560
jtoepel@purdue.edu

Saint Louis University

Rajeev Aurora
Ph. 314-977-8891
aurorar@slu.edu

Rangesh Kunnavvakkam
Ph. 314-977-8891
kunnavrv@slu.edu

The Donald Danforth Plant Science Center

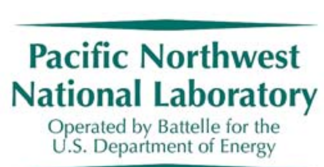
Thomas Smith
Ph. 314-587-1451
tsmith@danforthcenter.org

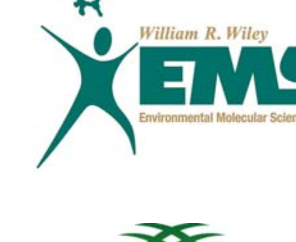
Nicole Koropatkin
Ph. 314-587-1458
nkoropatkin@danforthcenter.org

Shanghai Institute of Plant Physiology & Ecology

Terou Ogawa
Ph. 86 21-54924135
ogawater@xd6.so-net.ne.jp

**Participants past/present; original affiliations shown*





H. Executive Summary – Biogeochemistry Grand Challenge

Biogeochemistry Grand Challenge Concept

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a U.S. Department of Energy (DOE) national scientific user facility, undertook a Grand Challenge research effort in biogeochemistry to align with research programs in the Environmental Remediation Sciences (ERSD) and Life Sciences Divisions (LSD) of the DOE's Office of Biological and Environmental Research (BER). Specific programs within BER that are relevant to the Biogeochemistry Grand Challenge (BGC) include the Environmental Management Sciences (EMSP), and Genomes: Genomes to Life (GtL) Programs. Although the effort has been led by PNNL's Dr. John Zachara and Dr. Jim Fredrickson, the BGC is a coordinated, multi-investigator research effort focused on resolving a major scientific issue not accessible to the single investigator. The science theme was debated and identified by a multidisciplinary group of experts at a workshop entitled "Earth-Life Interaction at the Microbe-Mineral Interface," held at the Pacific Northwest National Laboratory (PNNL) from November 4-6, 2003 in Richland, WA (workshop report available upon request). The topic identified for the grand challenge, "Electron Transfer Across the Microbe-Mineral Interface" is a cutting-edge research issue with potentially broad impacts. Advanced experimental and computational capabilities in EMSL and other DOE user facilities have been leveraged to address this complex science issue. The team, consisting of PNNL staff and external collaborators from academic institutions across the U.S., Canada, and the U.K. is truly interdisciplinary with expertise including microbiology, biochemistry, protein structure and electrochemistry, molecular biology, mineral physics and surface chemistry, and molecular modeling. The BGC effort was commenced in January 2005 and concludes at the end of the 2007 fiscal year. Accomplishments, including a number of high-visibile publications, are summarized below.

Hypothesis

PNNL's Biogeochemistry Grand Challenge (BGC) has investigated mechanisms of electron transfer at bacterium – mineral interface. The well studied metal-reducing bacterium *Shewanella oneidensis* (MR-1) was used as a model organism, and Fe(III) and Mn(III/IV) oxides that buffer the redox potential of subsurface and aquatic sediments are being used as solid phase electron acceptors. The BGC evaluated the hypothesis that the outer membrane decaheme *c*-type cytochrome, MtrC (SO1778) is the terminal electron donating protein on the microorganism surface that couples with the metal oxide surface to accomplish electron transfer.

Scientific Accomplishments

Evaluation of the BGC hypothesis has led to a number of significant scientific accomplishments and findings, all of which are consistent with the original hypothesis. Example findings include:

- The outer membrane decaheme-heme *c*-type cytochromes MtrC and OmcA were isolated and purified and found to rapidly bind and transfer electrons to Fe and Mn oxides.
- MtrC and OmcA exhibited electrochemical activities, as determined by potentiometric titrations and protein film voltammetry, over the ranges of +100 to -400 mV and 0 to-300 mV, respectively, consistent with their ability to function as a carrier of electrons from the

quinol pool of the inner membrane to Fe(III) mineral oxidants exterior to the outer membrane.

- EPR spectra of MtrC and OmcA revealed the presence of broad, complex features indicating the presence of magnetically spin-coupled low-spin *c*-hemes.
- Mutants deficient in MtrC and OmcA were unable to reduce Fe and Mn minerals but retained the ability to reduce complexed, soluble forms of the minerals.
- Immunomicroscopy revealed the presence of MtrC on the outside of MR-1 cells in dense patches scattered over the cell surface and in association with extracellular polymeric substances (EPS) that bound Fe oxide.
- Molecular dynamics simulations in combination with *ab initio* calculations revealed that electron transfer through a perfect hematite lattice is fast and thus electrons are predicted to diffuse away from the initial injection point at an appreciable rate, allowing for rapid turnover of electron-accepting surface sites. Molecular dynamics simulations also revealed that direct contact between heme groups in cytochromes and an iron oxide surface is possible.

EMSL Capabilities Utilized

- Electron Paramagnetic Resonance Spectroscopy (heme center electrochemistry)
- Mass Spectrometery (protein and protein complex characterization)
- Scanning Probe Microscopies (nanometer scale surface characterization)
- Scanning and Transmission Electron Microscopies (imaging of all forms)
- Surface Enhanced and Resonance Raman Spectroscopy (heme identification and localization)
- Molecular Beam Epitaxy (synthesis of thin films)
- Photoelectron Diffraction (surface structure)
- Mössbauer Spectroscopy (Fe coordination environment)
- X-ray diffraction Spectroscopy (mineral structure)
- Molecular Science Computing Facility (numeric experiments of electron transfer)

Publications

1. **Bose, S., B. H. Lower, M. F. Hochella Jr., Y. A. Gorby, D. W. Kennedy, D. E. McCready, and A. Madden.** 2007. Bioreduction of hematite nanoparticles by *Shewanella oneidensis* MR-1. Geochimica et Cosmochimica Acta **Submitted**.
2. **DiChristina, T., D. Bates, J. Burns, J. Dale, and A. Payne.** 2006. Microbial metal reduction by members of the genus *Shewanella*: novel strategies for anaerobic respiration. *In* L. Neretin (ed.), Biogeochemistry of Anoxic Marine Basins. Kluwer Publishing Co., Dordrecht, NL
3. **DiChristina, T. J., J. K. Fredrickson, and J. M. Zachara.** 2005. Enzymology of electron transport: energy generation with geochemical consequences., p. 27-52, Molecular Geomicrobiology, vol. 59.
4. **Eggleston, C. M., J. Vörös, L. Shi, B. H. Lower, T. C. Droubay, and P. J. S. Colberg.** 2007. Binding and direct electrochemistry of OmcA, an outer-membrane cytochrome from

an iron-reducing bacterium, with oxide electrodes: A candidate biofuel cell system. Inorganic Chimica Acta **DOI: 10.1016/j.ica.2007.07.015**.

5. **Harthorne, R. S., B. N. Jepson, T. A. Clarke, S. J. Field, J. Fredrickson, J. Zachara, J. N. Butt, and D. J. Richardson** 2007. Characterization of *Shewanella oneidensis* MtrC: A cell surface decaheme cytochrome involved in respiratory electron transport to extracellular electron acceptors. J Biol Inorg Chem **DOI 10.1007/s00775-007-0278-y**.
6. **Kerisit , S., and K. M. Rosso.** 2005. Charge transfer in FeO: A combined molecular-dynamics and *ab initio* study. Journal of Chemical Physics **123:**224712.
7. **Kerisit, S., and K. M. Rosso.** 2006. Computer simulation of electron transfer at hematite surfaces. Geochimica et Cosmochimica Acta **70:**1888-1903.
8. **Korenevsky, A., and T. J. Beveridge.** 2007. The surface physicochemistry and adhesiveness of *Shewanella* are affected by their surface polysaccharides. Microbiology **153:**1872-1883.
9. **Lower, B. H., L. Shi, R. Yongsunthon, T. C. Droubay, D. E. McCready, and S. K. Lower.** 2007. Specific bonds between an iron oxide surface and outer membrane cytochromes MtrC and OmcA from *Shewanella oneidensis* MR-1. Journal of Bacteriology **189:**4944-4952.
10. **Marshall, M. J., A. S. Beliaev, A. C. Dohnalkova, D. W. Kennedy, D. A. Saffarini, K. M. Kemner, M. I. Boyanov, B. Lai, D. E. Culley, S. B. Reed, L. Shi, Z. Wang, J. S. McLean, V. Bailey, C. J. Simonson, M. F. Romine, J. M. Zachara, and J. K. Fredrickson.** 2006. *c*-Type cytochrome-dependent formation of U(IV) nanoparticles by *Shewanella oneidensis*. Public Library of Science Biology **4:**1324-1333.
11. **Marshall, M. J., A. E. Plymale, D. W. Kennedy, Z. Wang, S. B. Reed, A. C. Dohnalkova, C. J. Simonson, C. Liu, D. A. Saffarini, M. F. Romine, J. M. Zachara, A. S. Beliaev, and J. K. Fredrickson.** 2007. Hydrogenase- and outer membrane *c*-type cytochrome-facilitated reduction of technetium(VII) by *Shewanella oneidensis* MR-1. Environmental Microbiology **DOI: 10.1111/j.1462-2920.01438.x**.
12. **Ross, D. E., S. S. Ruebush, S. L. Brantley, R. S. Hartshorne, T. A. Clarke, D. J. Richardson, and M. Tien.** 2007. Characterization of protein/protein interactions involved in iron reduction by *Shewanella oneidensis* MR-1. Applied Environmental Microbiology **10.1128/aem.00146-07:**AEM.00146-07.
13. **Shi, L., B. W. Chen, Z. M. Wang, D. A. Elias, M. U. Mayer, Y. A. Gorby, S. Ni, B. H. Lower, D. W. Kennedy, D. S. Wunschel, H. M. Mottaz, M. J. Marshall, E. A. Hill, A. S. Beliaev, J. M. Zachara, J. K. Fredrickson, and T. C. Squier.** 2006. Isolation of a high-affinity functional protein complex between OmcA and MtrC: Two outer membrane decaheme *c*-type cytochromes of *Shewanella oneidensis* MR-1. Journal of Bacteriology **188:**4705-4714.
14. **Shi, L., T. C. Squier, J. M. Zachara, and J. K. Fredrickson.** 2007. Respiration of metal (hydr)oxides by *Shewanella* and *Geobacter*: A key role for multiheme *c*-type cytochromes. . Molecular Microbiology **65:**12-20.
15. **Wigginton, N. S., K. M. Rosso, B. H. Lower, L. Shi, and M. F. Hochella, Jr. .** 2007. Electron tunneling properties of outer membrane decaheme cytochromes from *Shewanella oneidensis*. Geochimica et Cosmochimica Acta **71:**543-555.