Promise and Challenges in Systems Microbiology: Workshop Summary

Patricia McAdams, Evonne Tang, National Research Council

ISBN: 0-309-53077-6, 48 pages, 6 x 9, (2004)

This free PDF was downloaded from:
http://www.nap.edu/catalog/10934.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

Thank you for downloading this free PDF. If you have comments, questions or want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This book plus thousands more are available at www.nap.edu.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

PROMISE AND CHALLENGES IN SYSTEMS MICROBIOLOGY

WORKSHOP SUMMARY

Patricia McAdams
Evonne Tang

Board on Life Sciences

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS • 500 Fifth Street, N.W. • Washington, DC 20001

Printed in the United States of America.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PLANNING GROUP FOR THE WORKSHOP ON PROGRESS AND PROMISE IN SYSTEMS MICROBIOLOGY

JOHN C. DOYLE (*Cochair)*, California Institute of Technology, Pasadena
ABIGAIL SAYLERS (*Cochair*), University of Illinois, Urbana-Champaign
ALAN HASTINGS, University of California, Davis
DEREK R. LOVELY, University of Massachusetts, Amherst
MICHAEL SAVAGEAU, University of California, Davis
LYLE WHYTE, McGill University, Quebec, Canada

Staff

EVONNE TANG, Study Director
ROBIN SCHOEN, Senior Program Officer
FRANCES SHARPLES, Director, Board on Life Sciences
BRIDGET AVILA, Senior Project Assistant
DANIELLE GREENE, Research Intern
NORMAN GROSSBLATT, Senior Editor

v

BOARD ON LIFE SCIENCES

COREY S. GOODMAN, (*Chair*) University of California, Berkeley
RUTH BERKELMAN, Emory University, Atlanta, Georgia
R. ALTA CHARO, University of Wisconsin, Madison
DENNIS CHOI, Merck Research Laboratories, West Point, Pennsylvania
JOANNE CHORY, The Salk Institute for Biological Studies, La Jolla,
       California
JEFFREY L. DANGL, University of North Carolina, Chapel Hill
PAUL R. EHRLICH, Stanford University, Palo Alto, California
JAMES M. GENTILE, Hope College, Holland, Michigan
LINDA GREER, Natural Resources Defense Council, Washington,
       District of Columbia
ED HARLOW, Harvard Medical School, Cambridge, Massachusetts
DAVID HILLIS, University of Texas, Austin
KENNETH F. KELLER, University of Minnesota, Minneapolis
RANDALL MURCH, Institute for Defense Analyses, Alexandria, Virginia
GREGORY A. PETSKO, Brandeis University, Waltham, Massachusetts
STUART L. PIMM, Columbia University, New York, New York
BARBARA A. SCHAAL, Washington University, St. Louis, Missouri
JAMES TIEDJE, Michigan State University, East Lansing
KEITH YAMAMOTO, University of California, San Francisco

Senior Staff

FRANCES SHARPLES

Preface

In 1993, the National Research Council's Board on Biology established a series of workshops on biotechnology. The purpose of the discussions is to foster open communication among scientists, administrators, policymakers, and others engaged in biotechnology research, development, and commercialization. The neutral setting offered by the National Research Council is intended to promote mutual understanding among government, industry, and academe and to help develop imaginative approaches to problem solving. However, the objective is to illuminate issues, not to resolve them. Unlike study committees of the National Research Council, workshops cannot provide advice or recommendations to any government agency or other organization. Similarly, summaries of workshops do not reach conclusions or present recommendations but instead reflect the variety of opinions expressed by the participants. The comments in this report reflect the views of the workshop participants as indicated in the text.

The first workshop, held in 1996, focused on intellectual property rights issues related to plant biotechnology. Other workshops have focused on broad issues related to developing an agricultural genome project, privacy in biomedical and clinical research, and the field of bioinformatics.

On August 19, 2003, the Board on Life Sciences held the workshop on "Progress and Promise in Systems Microbiology." The systems approach attempts to use comparative, high-throughput assays and mathematical or computational models to generate a picture of system-wide activity that can yield insight into processes operating within the system. The workshop brought together scientists in academe, government, and other institutions

vii

viii

with representatives of the US Department of Energy, the National Science Foundation, and the National Academies. The focus of the workshop was on communication among scientists in different disciplines pertinent to the application of systems biology to microbial ecology, with a goal of facilitating transition among science and application and information dissemination. The training of next-generation scientists with an emphasis on the interface of biology and computational biology was also stressed during the workshop.

Several themes emerged during the discussions, which began with concrete case studies of particular microbial systems and proceeded to research and educational needs to support further development of the field. Microbial ecosystems offer unique opportunities for practical bioengineering applications and attractive model systems for basic biological research. Substantial basic research is needed to develop the theory, software tools, and experimental methods required to support systems microbiology, but the case studies showed that progress has already been made by coupling mathematical modeling with experimentation. There was wide agreement that tight interactions of tool development with practical applications is needed, not only to keep tool development relevant and motivated but also because the current challenges for microbiologists in dealing with the enormous complexity of their systems and data can benefit from incremental progress in modeling, simulation, and analysis tools.

John C. Doyle
Cochair, Workshop on Progress and
Promise in Systems Microbiology

Evonne Tang
Study Director

Acknowledgments

This document presents the rapporteur's summary of the workshop discussions and does not necessarily reflect the views of the round-table members or other participants.

This summary has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published summary as sound as possible and to ensure that the summary meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this summary:

Timothy Donohue, University of Wisconsin, Madison
Phil Harriman, EPSCoR Centers Development Initiative
Reinhard Laubenbacher, Virginia Polytechnic Institute and State University
Gary S. Sayler, University of Tennessee, Knoxville

The review of this summary was overseen by George Kenyon, University of Michigan. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this summary rests entirely with the institution.

Contents

xi

1

Introduction

n 1674, in the Netherlands, Anton van Leeuwenhoek placed a drop of pond water under a glass that he had ground and saw a community of microorganisms never before seen by humans. He called the creatures animalcules. Two centuries later, a French scientist, Louis Pasteur, identified a set of animalcules that required no oxygen for life. Pasteur soon uncovered the extraordinary role of these microorganisms in the universe. By the ebb and flow of their own respiration, they connect the living and mineral worlds and recycle the nutrients that sustain life on Earth. Pasteur called these organisms anaerobes. Today, 150 years after Pasteur's discovery, a lot more is known about the important role of microorganisms on Earth—for example, they produce diverse sources of energy, including methane and hydrogen, and they influence biogeochemical cycles. However, given the number of microbial species on Earth and the complexity of microbial ecosystems, many aspects of microorganisms and of the communities they form remain a mystery.

Microbiologists have become interested in applying "systems biology" to understand and harness complex biological processes in microbial communities. Systems biology has been defined by Ideker et al.[1] as an approach to study "biological systems by systematically perturbing them (biologically, genetically, or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately, formu-

---

[1]Ideker, T., T. Galitski, and L. Hood. 2001. A new approach to decoding life: Systems Biology. Annu. Rev. Genom. Hum. Genet. 2:343-372.

1

lating mathematical models that describe the structure of the system and its responses to individual perturbations." A systems approach, which attempts to use comparative, high-throughput assays, and mathematical or computational models, has been used to generate a picture of system-wide activity that can yield insight into processes operating within a single cell. But the concept of integrating advances in genomics, proteomics, and metabolomics and incorporating them into mathematical models can also be applied to microbial ecosystems, which typically occur in consortia of related and unrelated organisms. Research on microbial communities using a system-based approach could provide a broader perspective on controls on biological processes and how they operate in and among microorganisms.

The US Department of Energy (DOE) sponsored a National Research Council workshop on "Progress and Promise of Systems Microbiology" that complements its Genome to Life program (GTL). GTL's key goal is to attain a basic understanding of thousands of microorganisms and their communities in their habitat and to apply this knowledge toward clean energy production and bioremediation. This workshop, held on August 19, 2003, in Washington, DC, was intended to be a forum for discussion of the tools, technology, and programs that are needed to advance the study of microorganisms through a systems approach.

Although some infrastructure for systems microbiology—such as high-throughput DNA-sequencing tools and methods—is similar to that needed for other genome projects, it also requires additional tools to address issues specific to microbial communities; for example, metagenomics (that is, the analysis of the sum of all genomes in an environment) will facilitate studies of microbial communities. This workshop examined the tools and infrastructure needed to advance systems microbiology. In addition, workshop participants discussed ways to encourage collaboration among scientists of different disciplines because "the combined capabilities and imagination of biological, physical, and computational scientists will be needed to organize creative new venues for discovery," as stated by Frazier et al.[2]

---

[2]Frazier, M.E., G.M. Johnson, D.G. Thomassen, C.E. Oliver, and A. Patrinos. 2003. Realizing the potential of the genome revolution: The Genomes to Life program. Science 300:290-293.

2

Workshop Proceedings

MICROBIAL SYSTEMS: CASE STUDIES

Because the workshop was formatted as open discussions, the planning committee of the workshop selected two contrasting case studies to maintain focus: on the *Geobacter* community, which has well-recognized biotechnologic applications, and on termite hindgut communities. On the basis of the two case studies, workshop participants outlined broad issues, including modeling, infrastructure needed, and education and training, even though some of the points made could be applied to a broader context.

Single-Organism System: *Geobacter*

To begin the workshop, Derek Lovley, (professor and head of the Department of Microbiology at the University of Massachusetts, Amherst) presented the first case study, *Geobacter*. The family Geobacteraceae includes over 40 known species of microorganisms, many of which have a considerable impact on the environment. According to Lovley, geobacters are known for their ability to remediate contaminated water and sediments and their ability to harvest electric energy from their surroundings. Other microorganisms play a role in the breakdown of pollutants but Lovley explained that *Geobacter* is special in that it dominates many microbial communities in subsurface environments in which oxygen is absent but iron oxides are abundant. Molecular studies, which avoid culture bias, have demonstrated that geobacters account for about half the microbial community in environments in which metal reduction is important. Not only are

3

geobacters more prevalent in these environments but they have been found to be the primary microbial species carrying out such processes of interest as iron reduction, uranium reduction, and the degradation of petroleum contaminants. The ability to bring geobacters into the laboratory, where they can be mass-cultured for biochemical and physiologic studies and for genetic sequencing, makes them highly valuable for investigating the interaction of a microbial community with its environment.

Bioremediation

Geobacters have different mechanisms for Fe(III) reduction from other phylogenetically distinct Fe(III) reducers. Lovley described the biochemistry behind their method of iron reduction from Fe(III) to Fe(II) and described how their unique form of respiration enables the application of geobacters for bioremediation of groundwater. According to Lovley, microorganisms that use insoluble Fe(III) oxide as an electron acceptor can have an important function in the carbon and nutrient cycles of aquatic sediments and in the bioremediation of groundwater. These microorganisms are effective in breaking down organic compounds, chlorinated solvents, and a variety of metals. In addition to geobacters' ability to reduce iron because of their unique respiration, these microorganisms can reduce other metals, including gold and uranium. In fact, the species may make up as much as 85% of the microbial communities in the subsurface during the most active phase of an in situ uranium bioremediation project.

Harvesting Electric Energy

In addition to their ability to reduce metals at contaminated sites, geobacters are capable of obtaining electricity from organic waste material. More specifically, geobacters harvest energy, transferring electrons not only within the cells but beyond, and the electrons can be harvested to produce electric current.

Geobacters can inhabit mud, generating electricity while they degrade pollutants. Furthermore, the microorganisms can be housed in an experimental fuel cell and power devices that are buried in the ocean floor and in remote locations. Geobacters can serve as biopowered antenna grids or play a role in novel biosensing devices. They can convert such waste organic matter as sewage sludge, domestic wastes, and industrial organic wastes to electricity; and they can convert renewable biomass to electricity instead of ethanol. They can harvest energy from hot environments with thermophiles, organisms that thrive under warm conditions ($\geq 50°C$). The study of geobacters, accordingly, will benefit DOE in remediation, in the development of cleaner forms of energy, and in biomass conversion to energy.

Genomics

An application of genomics has unraveled some of the key mechanisms that allow geobacters to compete and thrive in subsurface environments. Lovley and his colleagues are pursuing parallel research on previously cultured geobacters, novel culturing strategies to isolate environmentally relevant geobacters, and environmental genomic DNA of "as-yet-uncultured" geobacters to build conceptual models for optimizing uranium bioremediation and electrical energy harvesting.

Sequencing the genome of previously cultured geobacters led to a number of surprising discoveries. For example, he found a novel mechanism that allows geobacters to access Fe(III) oxides. Earlier studies had demonstrated that *Geobacter* species do not produce electron-shuttling compounds that might alleviate the need for the microorganism to contact Fe(III) oxide directly in order to reduce them. That raised the question of how geobacters could find the Fe(III) oxides that serve as their primary electron acceptor in subsurface environments. Although geobacters were believed to be non-motile, inspection of the preliminary genome sequence revealed genes for flagella—slender outgrowth of cells used for swimming and other forms of locomotion—and that suggested that the microorganisms might be motile after all, at least under some conditions. Laboratory studies allowed Lovley to confirm that geobacters are indeed motile, but only when they are growing with insoluble Fe(III) or Mn(IV) oxide as an electron acceptor. The expression of the genes for flagella are highly regulated and coupled with novel chemotaxis to access Fe(III).

The analysis of previously cultured geobacters revealed that the *Geobacter* genome comprises a high degree of gene duplication in cytochrome genes (i.e., certain cytochrome genes have a high percentage of identical sequences). Studies of differential gene expression led to the identification of genes specifically required for Fe(III) oxide reduction.

Other surprises revealed in the genome sequencing include the discovery of genes for new bioremediation capabilities, such as TNT degradation and mercury reduction, and novel central metabolism genes that suggest that geobacters are well adapted for acetate metabolism. Lovley said that data from the *Geobacter* species already sequenced, combined with systematic manipulative experiments that allow analysis of differential gene expression, associated physiologic and biochemical studies, and modeling had proved extremely valuable to his research team. "It has jumped us years ahead of where we would have been in our research."

The genome sequences of four geobacters are available, and up to 10 *Geobacter* pure-culture genomes will be available in another year. In-depth studies of gene expression with whole-genome DNA microarrays and detailed proteomic studies are under way. An in silico model for the central

metabolism of *Geobacter* has been developed. Preliminary environmental genomic studies suggest that uncultured *Geobacter* species in subsurface environments have, at least in some instances, similar gene arrangements to *Geobacter* species available in pure culture. Furthermore, billions of base pairs of genomic DNA extracted from environments in which geobacters predominate will be sequenced over the next 3 years.

In the conclusion of his presentation, Lovley said that application of pure culture and environmental genomic techniques to studies on geobacters has enhanced our understanding of its role in environmental processes. He raised the following issues regarding the application of systems biology in environmentally relevant microorganisms.

• Although techniques for high-throughput analysis of mRNA and proteins are available, high-throughput methods for quantitatively evaluating the metabolome of microorganisms are just as important but are not readily available.
• Ways to model the influence of minor components of a microbial community without a comprehensive analysis of each organism should be developed.
• Microbial ecologist should be trained so that they have enough understanding of basic microbial physiology and environmental science to design and interpret meaningful high-throughput studies on environmental microbial processes.

Multiorganism Systems: Termite Hindgut Communities

Jared Leadbetter (assistant professor of environmental microbiology at the California Institute of Technology) presented the second case study, on termite hindgut communities. As many as 250 species of microorganisms—many of which are strictly anaerobic—inhabit the hindgut of a termite, including bacteria, archaea, and protozoa. The termite shelters the microorganisms and feeds them with food that the insect itself cannot digest. The microorganisms digest the food and metabolize the nutrients into a biologically available meal for the insect. This exquisite architecture of symbiotic processes lies at the core of Leadbetter's work. Specifically, his research focuses on the intricate dynamics of the hindgut community of the California, or Pacific, dampwood termite, *Zootermopsis angusticollis*, and on the mutualistic symbiosis formed between the termite and the microorganisms living in its hindgut.

Z. angusticollis is nothing but a "grinding machine," according to Leadbetter. The termite feeds on wood, but it cannot digest cellulose, so it grinds up the chips and passes them to the hindgut, where the flagellated

protozoa and other microorganisms digest the carbohydrates, converting the sugars into the acetate that the insect uses for energy.

Termites are of interest to the scientific community for a number of reasons, including their role as sources of greenhouse gases, such as carbon dioxide, CO_2, and methane, CH_4. Those gases are released into the atmosphere by the hindgut microbiota. Because of their high biomass densities, termites, including their hindgut microbiota, contribute an estimated 2% of global CO_2 emission and about 4% of global CH_4 emission. Many termites are also the sites of intense nitrogen (N_2) fixation into biologically useful forms, an activity that stimulates the insect's proliferation on and mineralization of a nitrogen-poor diet, thereby introducing net protein into local food webs. Those activities are mediated by the hindgut microorganisms; any intense perturbation of the hindgut community results in the termite's loss of the ability to digest wood, to emit CH_4, and to fix N_2. The more than 2,000 species of termites each serve as host to a different microbiota. Because the microbial communities are different, each species of termite has an array of biochemical processes in its hindgut distinct from other species.

Microorganisms in the Termite Hindgut

"Molecular inventories suggest that there are easily 100-250 microbial species," Leadbetter said. A typical hindgut has a volume of 1-5 μL, but it can contain 40,000 to 50,000 large protozoa and other microorganisms. In one particular gut track isolated from a worker insect, more than seven species of anaerobic protozoa were identified, as were methanogenic archaea, and many aerobic and anaerobic bacterial phyla.

Leadbetter described several recent discoveries about termites, including steep gradients of pH, oxygen, and hydrogen throughout the gut tract. Hydrogen is produced in the anoxic core of the gut tract. As it diffuses out of the gut, its concentrations plummet. In other words, the hydrogen gradient counters the oxygen gradient, and this leaves the peripheral region of the gut with very different physical and chemical environments from the central portion. Because of the physical difference, the distribution of microorganisms in different regions of the gut tract is neither random nor homogeneous. In fact, the termite hindgut can be considered the smallest bioreactor on Earth. After discussing the peculiarities of the hindgut environment, Leadbetter introduced a number of microbial communities that live in the hindgut and presented an array of biochemical processes that are carried out by them. He then described aspects of one of the processes, called acetogenesis—the process of generating organic carbon in the form of acetate from inorganic carbon catalyzed by bacteria—which is the focus of his research at the California Institute of Technology.

Acetogenesis Research

Leadbetter pointed out that the termite gut fluid is typically dominated by spiral-shaped organisms, which phylogenetically are spirochetes. There are eight to 10 morphologically distinct types in one termite, but gene inventories have suggested that there may be as many as 25 or even 100 types of spirochetes in the gut of one termite. Until recently, biologists did not know the role of these organisms. Leadbetter was able to isolate some of them, grow them, and test them in the laboratory. He found that some of the spirochetes are bonafide homoacetogens, anaerobic bacteria that reduce CO_2 to acetate during respiration, also known as H_2-CO_2 acetogens. About 100 homoacetogens have been characterized, and most are clostridia and their relatives. The addition of spirochetes to that list is of interest, not only because of their role in the termite hindgut but because they represent genetic diversity. Leadbetter wants to know how the spirochetal pathway of acetogenesis compares with the "clostridial" Wood-Ljungdahl pathway. Diversity among acetogens suggests the presence of multiple niches even in this 1- to 5-μL environment. If that is the case, the diverse groups of acetogens warrant further investigation.

Molting and Reinoculation

At first glance, the orchestration of all the gut activities would seem to be the most fascinating aspect of termite biology. But the termite also sheds its hindgut contents when it molts. "The termite is born sterile and acquires the microorganisms from its parents and siblings," Leadbetter explained. "Every time it molts, it has to be reinoculated with microorganisms. Therefore, the whole ecosystem is sensitive to extinction. If the host fails, it's all over. Or if the microbial community fails the host, it's really all over." New generations of microorganisms have been passed from termite to termite in this manner for more than 12 million years. Leadbetter suspects that the insect would have shed species that are not needed over such a span of years, so he speculates that most of the 100-250 species of microorganisms play some role in the hindgut. So far, Leadbetter has been focusing on robust systems, but he would like to investigate some questions further. For example, What are the first microorganisms that are reinoculated in the hindgut after the molt? What role do they play?

Future Research Endeavors

In summarizing his research, Leadbetter pointed out that despite the widely recognized role of termites as global sources of greenhouse gases, such as CO_2 and CH_4, little is known about the function of most of the

hindgut microorganisms, and much less about the insect- and atmosphere-derived physicochemical controls on lignocellulose mineralization. The termite hindgut system is ideal for complexity analysis, which focuses on one metabolic process carried out by various contributors (such as cellulose digestion by protozoa or methane production by archaea) without considering the detailed ancillary physiology of the contributors. Such analysis could be conducted at the genetic level if specific genes that encode important enzymes or transport proteins could be identified. The ecosystem can also be perturbed and then the effects compared. Termites are amenable to a range of physiological and ecological studies. The quandary is that many techniques disrupt the architecture of the system, which is known to be important. Better experimental designs are needed to disentangle the physical, chemical, biological, genetic, and temporal architecture of the termite hindgut.

Leadbetter concluded his presentation with the following question and remarks: "Even if the complete metagenome of the hindgut community were available and sorted into all the files it should be in, what information would that reconstruction provide? I'm excited, and I would love to have those data. I'm just not sure how far they will go."

EXPLORING MICROBIAL SYSTEMS

The biological processes of microorganisms have been explored at multiple levels of observation, from gene regulation to the interactions of pathogens with their hosts. Recently, scientists have been interested in adopting so-called systems approaches to understand and harness complex biological processes. Although systems analyses have most often focused on the operation of a single cell or on the individual biological processes in an organism, microorganisms do not typically exist as individuals or even as pure cultures. Rather, they occur in consortia of related and unrelated organisms in natural communities or in engineered fermenters. Thus, additional challenges in systems microbiology research include the need to define a system (be it a single-cell system or a consortium) and its subsystems and the need to integrate multiple scales. Including research on microbial communities in system-based studies could provide a broader perspective on controls of biological processes and how they operate in and among microorganisms.

Definition of a System

Before discussing how best to explore microbial systems, participants in the workshop considered some characteristics of a model biological system. Participants decided that such a system must have a fixed number of

inputs and outputs. It should be a single isolated entity with distinct bound-aries. A half-cell, for example, is not a good system. A whole cell, however, is often a reasonably good system. A single-organism community (such as *Geobacter*) and a multiorganism system (such as in the termite hindgut) are other examples of good systems. A biological system, in fact, can be on any scale. A tidal salt marsh is an example of a whole intact ecosystem with multiple variables with respect to the flow of water in and out of the system. A system can be on the molecular scale. For example, a network of genes that function by inducing and suppressing each other is a system.

Connecting the coarse-grain description of a single-organism or multiorganism system with the fine-grain molecular system is a challenge. "There is a tool, however, to help to connect the dots," said John Doyle (professor of electrical engineering and control and dynamic systems at the California Institute of Technology and workshop chair). "That tool is modeling and modeling enables the additional tools of simulation and analysis."

Daniel Drell (DOE) cautioned that modeling a microorganism or a microbial system cannot be done all at once; such systems have to be modeled in pieces. Those thoughts were echoed by several participants later during the workshop. Michael Savageau (University of California, Davis) suggested that one strategy for systems microbiology is to understand the subsystems and try to put them together. As an example, he said that understanding all processes at the RNA level would not provide information on how a cell works, but knowing the interactions between mRNA, protein, metabolites, receptors, and sensors in a subsystem and connecting the subsystems would allow an understanding of how the whole cell system functions. Lovley agreed that measuring every variable possible—for example, every protein or every protein-protein interaction—may not be the best approach to study a system, because of the large amount of data generated. That led to a discussion on the modularity of microbial systems, that is the extent to which a system is divided into modules connected by interfaces.

Even though living things are diverse, there is a striking unity among them, particularly at the microbial level. All forms of life share the genetic code, many biochemical processes and similarities in cell structure. In fact, J. Doyle said that microorganisms appear to be well-designed systems with a high degree of modularity that facilitates domain swapping and horizontal transfer.

J. Doyle predicted that scientists one day will be able to model biological systems by taking advantage of the modularity of microorganisms. Gene swapping appears to be a source of confusion, he said, but the ability of systems to swap genes suggests a degree of modularity. Systems with a high degree of modularity are easier to model than those without it. Microbial ecosystems are very attractive for that reason. As soon as biologists understand more about the principles, the processes and how they function

together can be sorted out quickly. Although biologists are already familiar with some of the principles, they know them implicitly, incompletely, and more at the small-component level than at the network level. Engineering often moves ahead by integrating modules as much as by creating new ones.

Savageau pointed out that if biologists make the wrong assumptions about how the subsystems are partitioned, it may be difficult to link all the subsystems to obtain a picture of the whole system. "There are many similarities between biology and engineering, but there are some differences. In engineering, you design the systems so that the interfaces are clean. In biology, you have to discover them." In a microbial consortium, the system can be partitioned at the community level, the cellular level, and the subcellular level. The proper partitioning and the integration of different levels of subsystems, particularly given the role of evolution in creating diversity within those subsystems, pose another challenge to applying systems-based approach to microbial ecosystems.

Geobacter and Termite Hindgut as Model Systems

The workshop planners chose *Geobacter* and termite hindgut organisms as extreme cases of single-organism and multiorganism systems for comparison and to maintain the focus of discussion. The workshop participants considered whether those two systems illustrate all the challenges faced by microbiologists who apply the systems approach to study other systems. Participants concluded that the two model ecosystems represent only heterotrophic systems—photosynthesis does not play a role in either. Both are discrete systems that can be studied in isolation, so they do not illustrate the difficulties of studying such complex systems as soil microbial systems that cannot be studied in isolation and do not have controlled inputs. Nevertheless, a systems-level approach to understanding, simulating, and designing cells based on the two communities may contribute substantially to biologists' genetic engineering of microorganisms and plants for a variety of industrial and agricultural applications.

In comparing the two systems in a systems-biology context, Gary Sayler (University of Tennessee, Knoxville) noted a large knowledge gap regarding their cellular communication and cell-to-cell signaling. Sayler explained that all kinds of communication go on in the environment. "It occurs between and within species. It occurs between plants and microorganisms. And now there is evidence that human hosts communicate with microorganisms. Undoubtedly, those processes are coming under fundamental exploration at all levels, including mathematical simulations. They are probably operating in microbial systems, too, but we don't exploit those in ferreting out where they are." Sayler said that it would not surprise him if Leadbetter came back next year and said that the termite was releasing a

pheromone that was communicating with the methanogens in holding down population. "There is a lot of exciting work going on in systems biology that requires the integration of biology, mathematics, and engineering," Sayler said.

Leadbetter shared a story about communication between a bacterium living in close association with a phototroph, an organism that harvests energy from light. One of the fascinating aspects of this consortium, Leadbetter said, is that the phototroph contains a chromophore that is attracted to light of a specific wavelength. The bacterial cell has a flagellum and can actually swim toward light, but it does not have the chromophore, so it cannot detect that light. "There has to be some kind of communication, not just at the transcriptional level, but actual, behavioral, real-time communication between the phototroph that receives the gradients of light and the bacterium that swims toward that light. These are two organisms with very rapid communication related, in this case, to gradients of light."

J. Doyle thought that Leadbetter suggested an interesting example of a deductive approach that draws inferences about what might be happening by ruling out what cannot be happening. "You might know, for example, what bandwidth the control system has to have, and from that you can infer how fast the signal has to cross the system. Thus, you might know that the signal probably cannot diffuse through the membrane, so it probably has to have some coupled receptors between the two organisms or some other mechanism, such as data channels, to go between two membranes. Just by knowing the speed of response, therefore, you could know several things about what the mechanism could not be. This cellular communication could not involve transcription, for example, and you could rule out other things as well."

MODELING—THE TOOL FOR UNDERSTANDING SYSTEMS

J. Doyle began the discussion about modeling by stating that "modeling can play a huge role in helping us to understand biological systems. What is needed is software infrastructure and new theory for systems biology."

The first point Doyle made is that simulation technology is essential to connect multiple scales. He added, however, that existing simulation technology is inadequate because biological dynamics are stiff and stochastic with lots of random variables and multiple scales. Processes in biological systems occur in different time scales, and fast dynamics coexist with slow dynamics. The statement holds true for modeling cells or collections of cells. Further complicating the dynamics is the fact that biological systems are extremely stochastic with different counts of random molecules in different time scales. Modelers are still struggling to develop a more unified view of stiffness that encompasses the different timescales and stochasticity.

Connecting scales in biology is different from connecting scales in physics and chemistry; only some components of a biological system matter enormously and others don't. Some components in a biological system could be varied without disrupting the system because of its robustness. To "bootstrap" (that is, to draw statistical inferences based on a data-based simulation) an understanding of a system, modeling some of the detailed components that do not greatly disrupt the system can be deferred until later.

Doyle's second point is that as critical as simulation technology is, simulation alone is not an adequate tool. Suppose that all the genomic data of an ecosystem are available. In principle, the characteristics of the available metabolic and cellular processes are known and a model could be constructed to describe the ecosystem. However, many characteristics would still be unknown. If new data become available, an exponentially large number of simulations need to be performed within the parameters to determine whether the proposed model is consistent with all the available data. Simulation alone is not a scalable way to analyze a model, and the alternative to simulations requires mathematics—a novel approach even within the engineering community.

Finally, Doyle said fundamental laws that govern the organization of biological networks still need to be discovered. For instance, the conservation of energy is a fundamental principle that applies in all biological systems, but other principles are not as well known. If such general principles were known, it would be recognized that there are only a few ways to do some things. Once the laws governing how networks can be organized are known, the key components that should be included in a model can be more easily identified. Therefore, biologists and modelers need to identify some of the hard constraints in biological networks. Without knowledge of the governing principles, scientists may never truly understand how a system functions even if all its components are known. Most of the governing principles are related to efficient energy flow, particularly resistance to fluctuations, Doyle explained. "If you don't exploit these laws, the complexity will be bewildering and overwhelming."

Doyle continued: "Microbial ecosystems are a fantastic domain in which to study the basic questions in multicellularity that are important at the core of biology, let alone the obvious immediate applied benefits that everybody is well aware of in terms of just basic research. Modeling microbial systems is going to be a very good challenge. If the modeling is not well done, however, or if we don't develop the right new tools, we will never be able to figure out how these systems work."

Eberhard Voit (Medical University of South Carolina) agreed with Doyle that taking static snapshots of a system and comparing snapshots are insufficient for studying regulation of systems. Modeling should be based on dynamic measurements and detailed study of responses to defined inputs.

The frequency and duration of sampling and measurement depend on how fast the dynamics are. Thus tools for systems microbiology should include those that could measure systems dynamics to ensure data collection at appropriate timescales. Computation infrastructure for modeling must be able to integrate dynamic measurements of multiple scales.

Workshop participants agreed that the flood of data from high-throughput techniques makes it clear that biological research needs to move forward in a more quantitative manner than it has in the past. Biologists need engineers to help them model complex systems and build technology infrastructure.

Identifying the Goals for Modeling Biological Systems

Biological and mathematic modeling serves two main purposes: elucidate the structures of biological processes or make predictions. Mathematical ecologist Alan Hastings (University of California, Davis) said that successful modeling requires defined multiple goals and that different models need to be used to reach different goals. That is more true of biological systems than of physical systems because biological systems are more complex. But the art of modeling is knowing what to leave out, and what should be left out depends on the questions to be answered. Hastings believes that biologists should define the goals and identify the questions before they begin modeling. The art of knowing what components to leave out of the model results from the interplay between biologists and modelers. Often, biologists consider many components important, so they develop almost a one-to-one scale model that includes most components.

Modeling Biological Structure

Modelers can help biologists to identify key components that elucidate the structure of a biological system. Hastings said that one of the most valuable outcomes of modeling is knowing where to put our effort—knowing the data to collect that will make the most difference in our understanding. Even fairly crude models can be useful in pinpointing the most important items to study.

Applying Governing Principles to Predictive Modeling

J. Doyle suggested that an approach to the predictive modeling of microbial systems is to consider some of the governing principles that are universal in complex organisms because biologists do not have the time or the resources to collect data for every parameter. Instead of deciding which variables to leave out of the model, one can construct the model on the basis of universal governing principles. For example, retardation of growth

by nutrient deprivation is a governing principle that could be used to model growth. The desired outcome would be a simple yet robust model.

He referred to the demand theory of gene regulation, pioneered by Savageau, which says that the negative mode will be selected for the control of a gene whose function is in low demand in an organism's natural environment, whereas the positive mode will be selected for the control of a gene whose function is in high demand. A good example of gene transcription that switches "on" and "off" in accordance with the demand theory of gene regulation is the lactose operon, more commonly known as the lac operon.

According to Doyle, demand theory proved to be able to predict regulatory strategies that connect environments to organisms. The underlying idea is that the characteristics of the organism provide insight into environment and vice versa. Thus, knowledge of an organism's environment can be used to make inferences and predictions about the organism. Doyle explained, "I'm hoping this is a paradigm for what we are going to want to do with ecosystem problems. Bootstrapping is going to be essential. The point is that we are going to have to model systems that we know little about initially, so it's often going to be in very coarse resolution to start with." Doyle reiterated that modeling microbial systems requires new technology infrastructure that would allow multiscale modeling with variable resolution.

Referring to the discussion on the different laws and governing principles in biological systems, Leadbetter pointed out that some variables probably will be irrelevant to geobacters. To illustrate his point, Leadbetter used the day and night cycle and seasonality as examples. Those two variables probably have little impact on gene expression in geobacters. However, if the goal is to predict the number of people who will wear sunglasses and go ice skating in Central Park in New York, the day and night cycle and seasonality are important to consider. The day and night cycle and seasonality may be the governing principles for some processes and have minimal effect on others. Doyle thought that that was a good example because it is the changes that result from the two cycles that make a big difference in New Yorkers' response strategies. To reiterate the importance of governing principles, Doyle pointed out that the demand theory considers how variations in an environment determine a gene-expression pattern.

Matt Kane (National Science Foundation) talked about how geobacters' relationship to oxygen is similar to that of some other organisms. He asked whether relationship to oxygen is the controlling factor that could be investigated in studying an organism's relationship to its environment. J. Doyle agreed that that is the kind of questions in which modeling could be valuable because the hard constraints can be identified. He recounted how Lovley had conducted his research to learn about *Geobacter*. Lovley had

hypothesized that there must be an organism performing a particular reaction in that environment, so he created a culture environment that emphasized this reaction in which the organism survived and he isolated those organisms. Doyle thought that was a good test for identifying hard constraints.

A simple model might also provide valuable insight into spatial aspects of an ecosystem. For example, in the *Geobacter* case study, models with few parameters might provide insight into the effect of acetate depletion in one location. Again, Doyle pointed out an instance in which Lovley wanted to explain why sulfur reducers were able to overrun the geobacters and so he constructed a simple model. Lovley reasoned that once the geobacters' preferred electron acceptor is depleted, the next preferred one is sulfate; the sulfur reducers flourish because of the abundant amount of acetate that is available for their consumption. Doyle pointed out that although Lovley constructed a model with few interactions, it provided valuable insights. Doyle elaborated that it is easy to include just a few more interactions, but the combination of multiple interactions becomes bewildering. "Fairly simple models that don't have any genes, and don't have many details, can help you sort out what might be the next thing to do. You won't always get the right answer, but it suggests how you might do the next experiment. That's how you bootstrap yourself into deeper understanding."

To end this part of the discussion, Lovley referred to his research with geobacters. He repeated the story about how the use of models predicted phenotypes that he could never have sorted out only by looking at the genome. In part, that was because the *Geobacter* mutant grew better on iron than the wild type. "That never would have entered my mind."

Are Detailed Data Needed for Modeling?

Timothy Donohue (University of Wisconsin, Madison) was troubled by J. Doyle's vision that biologists did not have to worry about having all the data and details. His concern was echoed by several participants. Using Lovley's work on *Geobacter* as an example, Donohue said that he would like his datasets to be robust enough for a model to be able to predict whether a particular gene is present. Although sequence analysis could have predicted the presence of a particular gene in *Geobacter*, Lovley purified the samples and demonstrated the presence and location of the gene.

Doyle, however, clarified that data are always incomplete. Some biologists believe that all the information are in the gene-coding sequence, but there are large gaps in knowledge of cell dynamics. In dynamic situations, regulation is much more complex and cannot be predicted from genomics alone. "The good news is that you can defer knowing all those details. You can get somewhere and make predictions. You can move ahead without

having to know everything. You can have incomplete data, incomplete models, and incomplete knowledge; but you use what you have to predict the next discovery."

Lovley does not worry about every component of central metabolism in *Geobacter*, which he can model fairly well, but respiration is not understood and cannot be sorted out by looking at genomic sequences. Until he has that information, he cannot predict other responses in *Geobacter*. At some point, the function of a particular subset of genes needs to be elucidated. He asserted that "every genomic project should study the physiology of an organism."

Leadbetter agreed with Lovley's view that the structure and architecture of some ecosystems, such as the tidal salt marsh, can be elucidated on the basis of a few physical characteristics. For example, if the effects of oxygen, pH, sulfite, and light are taken into account, the biological structure of sediment, in which there might be 1,000 species, can be understood. The stratification of the salt marsh can be clearly modeled on the basis of knowledge of those four components. A day and night cycle and maybe seasonality can be included to improve understanding of the underlying architecture of the community. But every system is different. "I would love to have that type of information on the termite, but I don't know what the four components are. If I did, it would change everything. With the termite, we need to find out what the critical components are before predictions can be made."

J. Doyle urged the group to keep their eyes on the big picture and to apply high-value-added details only to models so as to reach their goals. "People who model human diabetes processes are helping in drug design, but they have no idea what most of the details are. The idea that you have to know how every gene works with every protein is unrealistic. The point is to figure out what we can do now that enables us to move ahead without having to know everything."

Jizhong Zhou (Oak Ridge National Laboratory) supported Doyle's premise of the need to stay focused on the big picture. Microbial systems are very much a "black box," and not all the details are necessary for constructing models, he said. Hastings spoke of the advances in population models made possible over the last 15 years by quantitative statistical approaches for matching them to data. Scientists are gathering and structuring information on what used to be black boxes.

Doyle suggested the importance of focusing on key areas. "For example, if we were going to look at metabolite concentrations in some detail, some would give us an enormous amount of information, and some would give us little. You could spend the next decade spending huge amounts of money in measuring concentrations of metabolites in glycolysis in *Geobacter*, for example, and you would learn almost nothing. But some things are show-

stoppers. For example, understanding of some aspects of respiration is necessary for knowing how organisms function. So you have to measure in a targeted way. If you are looking at genes and proteins, you don't want to do everything. You want high-value-added data."

Hastings suggested that ecologists had also struggled with the issue of whether detailed data are needed for modeling. "About 60 years ago, ecologists split into two camps and they haven't talked to each other much since. One camp looked only at energy flow and exchange of materials. The other camp looked only at the dynamics of species numbers. We keep hitting the same issues: can you just look at input and output without considering which species are and are not present? Or do you have to look only at the species that are present? Those are fundamental questions.

The group considered the issue from a variety of perspectives and, with J. Doyle, concluded that it is clear that taking dynamic, in situ measurements is difficult. Modeling can play a big role in combining sequencing, which provides a static picture and lots of data—with a few key dynamic measurements. That type of work can be expensive, but it is worth doing because dynamic measurements and responses to inputs are essential for learning about regulation, according to Voit. There is no doubt that snapshots are valuable and necessary, but eventually dynamic experiments and analyses have to be performed if we are to understand the structure and function of a whole system.

To conclude the discussion on modeling, Doyle said, "Again, the point is, you want to have enough information to be able to say, 'I don't know what all the dynamics are, but I know that the basic dynamics in the cell of greatest interest in this system occur here.' So you know where to look without knowing the entire answer. That's the sort of bootstrapping that we hope will continue to work."

TOOLS FOR SYSTEMS-MICROBIOLOGY STUDIES

After discussing modeling of the two case studies, the group turned its discussion to the tools, databases, and personnel that are needed to facilitate the modeling of microbial systems. A systems approach to research in microorganisms and microbial communities requires novel experimental tools for detecting and measuring biochemical processes—such as gene expression, metabolic activity, and protein production—across the entire system, whether the system is defined as a genome, a cell, or a cell culture. To make sense of information generated from measurements of system-wide activity, assays must be repeated continually, each time monitoring the system under a different set of conditions.

The statistical analysis of the resulting data is likely to be computationally intensive, and mathematical and computer models are needed to

capture relationships in a way that enables the visualization and quantification of the underlying events that give rise to system-wide activity. Analysis of the data will be facilitated by knowledge about pathways, molecular relationships, and operational mechanisms—knowledge that can be acquired through traditional hypothesis-driven studies. Combining data from multiple research approaches will enhance our understanding of biological systems.

Experimental Tools

Workshop participants were asked to think about the technologies and tools that could facilitate systems-microbiology studies. What technologies are needed for exploring microbial systems? What are the corresponding in vivo and in situ tools that would provide insights into the dynamics? What technologies are going to be available for dynamic measurements?

Donohue believes that many creative scientists could develop tools for database management, dynamic measurements, and so on, if they had appropriate funding. Sayler suggested that an environmental genome for prokaryotes would be a useful tool. "My students don't even want to make microarrays", Sayler said, adding that he would like to purchase such tools as a phylogenetic microarray so that his students could be actively engaged in the experimental work, such as perturbation experiments to study responses, instead of spending time in developing a microarray. A genotyping chip with diagnostic sequences could help to determine which microorganisms are present in a community, their activity and dynamics. An appropriate chip-based analysis can be very informative despite its low sensitivity. Sayler suggested that more commercial biotechnology companies are needed to produce tools that facilitate and accelerate biological research. One such company is Biolog, in Hayward, California, which makes the Phenotype MicroArrays™ that can quantitatively measure thousands of cellular phenotypes at once. The company also makes systems that can rapidly identify over 1,900 species of aerobic and anaerobic bacteria, yeasts, and fungi [http://www.biolog.com/main.html].

As an example of one funding mechanism that enables investigators to buy microarray chips, Leadbetter credited the Cystic Fibrosis Foundation [http://www.cfri.org/home.htm] for developing an especially creative strategy for generating research about cystic fibrosis. "The Foundation makes metric chips available for about $150 if you write to them with a short proposal and say, 'I'm going to study this particular aspect of *Pseudomonas aeruginosa*. It may have some connection, some way, or somehow, to cystic fibrosis.' You can then buy 10 chips for $1,500, and they even give you the software free. This policy has stimulated research in the field: it changes the dynamics of researchers who might be afraid to fail with a $2,000 chip."

Lyle Whyte (McGill University) asked what high-throughput techniques, aside from phylogenetic microarrays, funding agencies like DOE can support. Zhou said that typical microarrays are easy to make but insufficient for some kinds of investigations. For the termite hindgut, for example, which may have more than 100 species of microorganisms in a single insect, new high-density microarrays are necessary. However, even if there were microarrays for 100 microorganisms from the hindgut of one termite, there are more than 2,000 species of termites to consider.

Zhou suggested proteomics as another field in which DOE could help to advance development so that the proteins that carry out a function can be identified and how they function together can be understood.

James Frederickson (Pacific Northwest National Laboratory) emphasized the importance of basic biological research. "I think it's safe to say that we don't know the entire function of networks, even in *E. coli*, which has been studied to death. We still don't know the function of 30-40% of genes. Moreover, the known genes may serve functions that we are not aware of. We haven't even talked about proteins; identifying the proteins that carry out the work of the cell is important, and we need to understand how proteins interact and work together to carry out functions at the cellular level. Many fundamental issues need to be understood before we can move on to other things."

Frederickson noted that a lot of the technologies being developed—such as proteomics and transcriptome analysis—are based on using a large biomass of cells, around 10^9 cells. But there is large variation among individual organisms. Furthermore, many assays are based on in vitro techniques that need to be confirmed in vivo. There is much room for developing approaches and technologies that are centered on confirming such assays.

Many new techniques have been developed to allow single-cell observation, according to Patrick Dennis (National Science Foundation). The classic example is the observation of bacterial motility, which was observed recently by taking pictures with a camera. "That's a type of situation where you can extract a tremendous amount of information by looking at single cells, rather than populations." Other novel technologies will contribute to an increased understanding in this field.

Databases and Ontologies

There are no analytic tools for taking data on gene expression and comparing the experimental results to decipher the biological meaning. For example, Sayler conducted a study, in collaboration with researchers in chemistry and medicine, on the influence of natural anticancer compounds on the human gene array. In the first round of experiments, they narrowed

down the potential anticancer compounds to 2,000 cytokines and chemokines. "The data that came out of the first series of experiments were good. The problem was that the biologists had to work virtually on a gene to gene basis to find out whether a gene response had any relevance to the cancer process." Tools are needed to analyze such large databases. J. Doyle commented that although software would aid biologists in comparing massive amounts of data and deciphering their meaning, it is generally expensive to develop, and so far no one wants to pay for it.

Donohue emphasized the importance of access to reliable, updated databases and the ability to compare data. He found that databases are overwhelmed with 2-year old annotation and cannot be updated fast enough. Different data on a given organism may be found in several databases, depending on who submitted the data. Moreover, the same organism may be labeled differently because of outdated information. Often, different software is built for each microorganism, and this makes cross-comparisons difficult. Doyle questioned whether the fundamental issue is that the correct abstractions for organizing biological information need to be developed or that programmers are still needed to piece together well-developed abstractions. Donohue believes that it is probably both. If databases are linked together seamlessly, researchers should be able to access all the records for a given organism in different databases.

Another tool is an ontology of functional modules that allows researchers to organize vast amounts of scientific data electronically. (An ontology is a description of the concepts and relationships pertinent to an entity or community of entities.) The objective of an ontology is to enable communication between informatic systems in a way that is independent of individual system technologies, information, architecture, and application domains. A structured language that can be put into the database or the ontology allows a computer to look for associations between data. An ontology forces scientists to think systematically about intermediate-level architectural issues and about the different levels of systems. Doyle suggested that not only does a current ontology need to be developed to help today's researchers in studying systems microbiology, but the framework should allow ontologies to evolve systematically and robustly. The hardest thing for software engineers to develop is a system that is both usable now and also flexible enough to adapt to technologies of the future, Doyle told the group.

There has been some effort by software engineers to connect databases and to create a distributed system whereby researchers can view their own annotation in a central searchable system. Any new database would be registered to the central system so that researchers can query the central system and obtain information from different databases. Donohue is not convinced that the databases are well connected, not to mention the short-

comings of some databases, such as GenBank, the National Institutes of Health genetic-sequence database that is an annotated collection of all publicly available DNA sequences. Donohue has found many errors in GenBank, which he says must be updated. "As an end user, when I go looking for something, I know it should be set up differently." Not only is GenBank often outdated, but the databases are not linked properly and are not "talking to" one another, he said. And how genes are identified is confusing. "Bacteriologists don't know what's out there. They need a Web site like the yeast Web page where from a gene you can get chip data, proteomic data, hybrid data, and functional data and go right to the citation—that's what everyone wants."

Doyle believes that basic research on linking databases still needs to be done. He suggested the Systems Biology Markup Language (SBML) as a success story in the context of modeling. SBML is a computer-readable format for representing models of biochemical reaction networks. The development of SBML benefited from the fact that few groups were working on a dynamic systems level and on control-regulation levels, so the groups could easily agree on standards. Most of the major simulation systems enable models to be shared and published in forms that other researchers can use even in different software environments. The key to SBML's success is having good software engineers who concentrate on making the system work. Doyle suggested that DOE could conceivably facilitate development in this field.

Doyle posited the hydraulic system of the Boeing 777 as a perfect example of systems that had problems initially and great success later. The hydraulic system of the Boeing 777 has two separate models, he explained: one describes the layout of the hydraulic lines, and the other the dynamics of the hydraulic system as a component of the control system. The first model is important because the people who wire up the lines are different from the people who decide where the structures go. In the past, when engineers built an airplane, the specifications called for running a hydraulic line directly through a big piece of metal, which, of course, cannot be done. They had to decide whether they should move the hydraulic line or drill a new hole. Before they could drill a new hole, however, they had to consider whether it would compromise the structure of the plane. Boeing has since invested about $1 billion in software so that it can test the system virtually and eliminate those sorts of errors. The software system was expensive, but Boeing made the money back in saved manufacturing and other costs. The model of the dynamics of the hydraulic system had to be developed independently of the model of the layout. A goal of research is to have more integrated models, and this is a big issue in biology.

Doyle again pointed to software that is generally extremely expensive and difficult to build. No one wants to pay the bill, despite the likelihood

of enormous long-term savings. In the case of systems microbiology, the fundamental issue is that the correct abstractions for organizing information need to be developed. In some cases, the abstractions exist but await skilled programmers to put them together properly.

LOOKING AHEAD—COLLABORATION AND EDUCATION

With the promises of microbiology well within reach, the group explored what might be done to facilitate exploration in molecular ecosystems. How might DOE and other funding agencies fuel research and accelerate the process? What are the challenges of integrating disciplines? How do we cross disciplinary boundaries in the short term? What can we do in the long term to educate and train participants to join efforts in the quest for scientific discovery?

Collaboration

Among the more obvious ways to encourage collaborative projects is to create more funding opportunities to drive research. Fredrickson suggested that DOE and other agencies build such goals into their programs. Funding for each project needs to be increased to accommodate the size and scope of interdisciplinary research and to convince investigators to collaborate and form teams that spark innovative research. Fredrickson remarked that opportunities need to exist before researchers can come together.

Finding funds is not so easy, according to Dennis. The National Science Foundation (NSF), for example, has a limited budget to support all types of research; historically, single-investigator grants seem to be the most cost-effective and yield the most return on the research dollar. "We've just heard how important basic science is for understanding lac operon, for example. We can take money away from single investigators and dump it into large collaborative interdisciplinary projects, but it's not going to have the same impact."

Dennis went on to say that NSF has a new Frontiers in Integrated Biological Research program that provides $1 million a year for 5 years. Many researchers believe that such funding is insufficient to meet the goals at hand and that pulling together a large team for a short period like 5 years is not realistic. Furthermore, such collaboration is risky, particularly for younger researchers who will be judged on whether they got grants from NSF and what has come out of their laboratories. Whether they participated in a group project and their achievements in collaboration with others are considered less important.

Another question centered on what kind of research NSF is looking for. The agency claims to support creative, imaginative science, but J. Doyle

argued that in general innovative proposals are almost guaranteed to get a "poor" rating from some reviewers and to be doomed to fail. Dennis confirmed that NSF is looking for creative, imaginative science, which he defined as the kind of small-scale science that might have a big impact down the line—studies on restriction enzymes, for example. Dennis defended his agency's practice of limiting the lifetime of research grants by explaining that science evolves. Resources may need to be reallocated to address different issues at different times. "NSF must pay attention to what the scientific community is saying."

Lovley said that at first collaborating with modelers on *Geobacter* was difficult for him; later he accepted that he did not have to know everything about an organism. "Once I got beyond that, it became a good experience because the people who do the modeling are engineers, and they ask questions we never had thought to ask, because they have a different perspective. I think it's gone incredibly smoothly even though, in our case, we are on opposite coasts."   Hastings was quick to point out that the engineers doing the *Geobacter* modeling had to learn enough biology to be able to communicate with Lovley; this led the workshop group to discuss biologists' and engineers' problems in learning enough about each other's field to be able to communicate.

Undergraduate and Graduate Programs

There was a consensus that an interdisciplinary approach to science begins at the undergraduate and graduate levels. How to bring it about was less clear. One suggested strategy involved transforming graduate-level microbiology courses into one integrative microbiology course to make room for quantitative courses, but this strategy was unpopular with the biologists in the group. Lovley objected to universities' reducing their offering of microbiology courses. "What if you like microbial genetics?" he asked. Leadbetter agreed with Lovley and elaborated on how some of his students are forced to devote as much as 80% of their time and effort to passing quantitative courses at the expense of the other courses they are taking. The students end up not getting much out of any of the courses. "I'm much more excited about getting people who like what they do and know how to do different things. Then we can try to build connections between them. And because they like working together, they will probably communicate."

The workshop participants supported the notion that curriculum constitutes a serious problem. But what courses should be eliminated to make room for new ones? Can a subject be compressed? How? How can people in different disciplines be trained with enough overlap to be able to communicate with each other? How much mathematics does a biologist need?

Quantitative courses enable biologists to assimilate and trust information that they get from computer scientists, said Donohue, "but we need to talk with our friends in the computer field to know the courses that are relevant." Donohue explained that the quantitative courses that his biology students are considering either have many requirements that his students cannot meet or have no relevance to their work in biology. The group agreed that courses in statistics, calculus, and computer science would help biologists in their research and studies. Those courses should be rigorous in presenting problems and examples in biology, chemistry, physics, and economics that science students can relate to.

Dennis suggested that biology needs to be taught in a mathematical framework, which has rarely been done in the past. In contrast, engineering and many other disciplines have been teaching mathematics in their own contexts. Biologists may have been teaching in cubicles—chemistry, biology, physics, mathematics, computer science, and engineering—and a more integrated curriculum needs to be developed. Core courses intended for students outside a discipline should be designed carefully with a specific focus, according to Savageau. As it is now, when nonbiologists first see a biochemistry text, they are overwhelmed by it and quickly lose interest. However, it is possible to teach the principles of biochemistry without all the details that specialists think are interesting.

Hastings proposed that scientists change how they think about doing science. He suggested that scientists in different disciplines consider using a common language in their collaborative work. Otherwise, they will be talking past one another. He described a course that he teaches to graduate students, some of whom are majoring in applied mathematics and in population biology. His students in different disciplines communicate with each other, he said. "If you can get the students to sit down together—outside class and without faculty members around—that's the real key. It will break down the barriers. We need to design projects with questions of interest to you as a microbiologist and questions of interest to you as an engineer. That's what develops collaborative skills and builds expertise in other backgrounds over time. We don't do a good job of teaching people to collaborate."

Future scientists may be encouraged to collaborate through carefully designed programs at the undergraduate and graduate levels, but postdoctoral researchers may be encouraged by the funding of projects that force them to work across disciplines with biologists, mathematicians, and computer scientists. Experienced scientists can look for new short courses specially designed to share knowledge across disciplines. A number of participants told the group about various short courses that lasted from a couple of hours to a couple of weeks. Frank Doyle (University of California,

Santa Barbara) stated that the time he spent on such special short courses had been worthwhile.

Learning enough to be able to speak the language, however, is insufficient. Working toward something that draws everyone together is the higher ideal. A number of workshop participants suggested that collaboration would be encouraged by identifying concrete challenges in subjects of common interest at the intersection of various disciplines. It is important not to focus systems microbiology only on microbiology, modeling, and computational biology; that would exclude knowledge essential to comprehending the complexities of these systems. Biochemistry, organic and inorganic chemistry, physics, economics, and other disciplines all play roles.

CONCLUDING REMARKS

In the final discussion, Donohue cautioned that the value of basic biological research and benchwork—such as genome sequencing, which Lovley had described earlier as the tool that led to the discovery that geobacters have genes for flagella—should not be neglected. The goal should be to train a group of well-rounded students and postdoctoral researchers. "Ten years from now, you don't want to have a group of students and postdoctoral scientists who know how to look at chip data and make models but don't know how to purify an enzyme."

J. Doyle, workshop chair, brought the meeting to a close when he said that he had learned about two important systems over the course of the workshop and, as an engineer, was becoming as interested in microorganisms as the biologists in the room were. As mentioned earlier, the shared interest across disciplines is the key to driving the exploration of microbial ecosystems. But Doyle said that the biology community has yet to recognize the need for specialists to develop theory and software infrastructure in order to advance systems biology. "One of the things the biology community and the funding agencies have to recognize is that if you want to have technology that works, you have to spend the money to build infrastructure. Not everyone needs to know how to do that, but somebody does. That is where systems biology has to go. You must have the infrastructure to address problems. Microorganisms are the foundation of the biosphere, but they have been neglected."

3

Summary of Key Points

More than two dozen experts from across the country met at the National Academies to discuss the promise of systems microbiology and to identify the multiple strategies and novel experimental tools needed to explore these systems. Although formal conclusions and recommendations will not come from this workshop, many insights into the future of the field can be gleaned from the discussion described here. The major points are summarized below.

APPLICATION OF THE SYSTEMS APPROACH

• The study of single-organism and multiorganism microbial communities, as illustrated by the two case studies, can greatly benefit from application of systems biology. To apply that approach successfully, we need to define systems with closed boundaries and identify the goals of studies.

MODELING

• Modeling can play a major role in helping us to understand biological systems. Software infrastructure and other interoperable tools for systems microbiology are needed to facilitate modeling.

• Successful modeling requires defined multiple goals. If the goal of modeling is to understand biological processes, modelers can help biologists to identify key components that reveal biological structure. For predictive modeling, modelers can help biologists to identify parameters that exert the most influence and the governing principles that place con-

27

straints on the variables to be modeled.   Model parameters should be those
that have a large influence on the predicted outcome.

TOOLS

• Having commercial access to assays, gene chips, and other novel
tools of biotechnology would streamline scientific research by allowing
biologists to concentrate on fundamental questions, rather than preparing
the actual tools. Access to inexpensive tools, such as the affordable gene
chips provided by the Cystic Fibrosis Foundation, would also encourage
increased research.

• Although databases on microorganisms exist, they are not connected
to each other, so microbiologists cannot move from one database to another
easily. To conduct a comprehensive search, they must be aware of various
databases. Furthermore, information in some databases is outdated, and
nomenclature is not uniform among databases. Effort is required to update
the databases, and software engineers are needed to design technologic
infrastructure that will link databases seamlessly.

EDUCATION AND RESEARCH

• Biologists need to learn more about systems approaches and learn
enough mathematics and computing knowledge to communicate with
systems engineers who design software and computational technologies.
Systems engineers could benefit from learning more about biology and the
interesting questions that could be answered by their discipline.

• Interdisciplinary research projects would be encouraged by increased
long-term funding opportunities. Scientists can facilitate collaboration not
only by learning a common language but by designing projects that are of
interest to scientists working in different fields.

• Training people with enough overlap of various disciplines will allow
them to communicate with one another. New graduate and undergraduate
curricula could be developed to facilitate such training.

• Experienced scientists can take advantage of short courses offered
by various research centers to learn more about other disciplines.

• Focusing systems microbiology too narrowly around microbiology,
modeling, and computational biology could exclude important fields of
knowledge that are essential for the understanding of the complexities of
systems. Biochemistry, organic and inorganic chemistry, physics, economics,
and other disciplines should all play a role in the exploration of microbial
ecosystems.

Appendix A

Progress and Promise in Systems Microbiology

Agenda

Date: August 19, 2003
Location: The National Academies
Keck Building, Room 205
500 Fifth Street NW
Washington, DC

8:30 – 8:45	Welcome and Introductions	
	• Introductions	*Evonne Tang*
	• Format of workshop	*John Doyle*
	Case Study 1: *Geobacter*	
8:45 – 9:15	Introductory talk on *Geobacter*	*Derek Lovley*
9:15 – 9:45	Definition of a system	*Derek Lovley*
9:45 – 10:15	Data availability and high throughput techniques for discovery and hypothesis driven science	*Mike Savageau*
10:15 – 10:45	Break	
10:45 – 11:15	Modeling	*John Doyle*
	Case Study 2: Termite hindgut	
11:15 – 11:45	Introductory talk on termite hindgut	*Jared Leadbetter*
11:45 – 1:00 p.m.	Lunch	

29

1:00 – 1:30	Definition of a system	*Lyle Whyte*
1:30 – 2:00	Data availability and high throughput techniques for discovery and hypothesis driven science	*Alan Hasting*
2:00 – 2:30	Modeling	*Lyle Whyte*
2:30 – 3:00	Break	
3:00 – 3:30	Comparison between systems approaches on single- and multiorganism systems	*Michael Savageau*
3:30 – 4:00	Integration of disciplines	*John Doyle*
4:00 – 4:30	Education and training	*Alan Hastings*
4:30 – 4:40	Wrap-up	*John Doyle*
4:40	Adjourn for day	

Appendix B

Participant Biographies

John C. Doyle (Workshop Chair) is professor of electrical engineering and control and dynamical systems at California Institute of Technology. His research interests are in integrating modeling, analysis, and design of uncertain nonlinear systems with applications throughout the aerospace and process-control industries and for complex networks in both engineering and biology. He received his BS and MS in electrical engineering from Massachusetts Institute of Technology in 1977 and his PhD in mathematics from the University of California, Berkeley in 1984. He has received the IEEE Centennial Outstanding Young Investigator Award (1984), the IEEE Hickernell Award (1976), the American Automatic Control Council Eckman Award (1983), and the Bernard Friedman Award (1984) and was an ONR Presidential Young Investigator and an NSF Presidential Young Investigator. He was a coconvenor of the Second International Conference on Systems Biology in Pasadena, California (2001).

Patrick Dennis is program director of Genes and Genome Systems Cluster in the Division of Molecular and Cellular Biosciences at the National Science Foundation. His research interests have focused on aspects of biochemistry and molecular biology of ribosome synthesis and cell growth. In the last 15 years he has characterized the organization of rRNA and ribosomal protein genes and the pathways for precursor rRNA processing and ribosome assembly and biogenesis in various halophilic and thermophilic archaeal species. Recently, he identified and biochemically characterized a novel class of small archaeal ribonucleoprotein machines that function in the assembly of ribosomal particles. In addition, he has continued a long-term interest in

31

the intimate relationship between ribosome synthesis and growth rate in *Escherichia coli*.

Timothy Donohue is professor of bacteriology at the University of Wisconsin, Madison. His research interest is in deciphering the fundamental problem of biological energy generation. His laboratory analyzes the strategies that cells use to generate energy. Members of his laboratory are using the resulting information to dissect important metabolic activities of bacteria to identify new energy-generating pathways of agricultural, environmental, and medical importance and to design new types of microbial machines that can efficiently degrade toxic compounds, remove greenhouse gases, or synthesize biodegradable polymers.

Frank Doyle is professor of chemical engineering at the University of California, Santa Barbara. His research focuses on nonlinear model-based control of complex nonlinear and distributed processes and the application of systems-engineering tools to problems in biology. The former topic is addressed with advanced simulation, optimization, and model-reduction methods, and the problems of interest span the chemical, pulp and paper, and pharmaceutical sectors. In systems biology, he uses traditional systems-engineering tools (such as model identification, parametric sensitivity, and closed-loop analysis) to analyze complex, hierarchic biological systems.

Daniel Drell is biologist and program manager in the Human and Microbial Genome Program in the Life Sciences Division of the Department of Energy. He also manages the Ethical, Legal, and Social Issues Program and has spoken extensively on this topic.

James F. Fredrickson is Laboratory Fellow in the Environmental Microbiology Group at Pacific Northwest National Laboratory. Since joining Battelle-Northwest in 1985, he has focused his research efforts on the microbial ecology and biogeochemistry subsurface environments, using geochemical and molecular approaches. More recently, he has been investigating the geochemical and microbial factors that control the rate and extent of reduction and dissolution of Fe oxides by dissimilatory iron-reducing bacteria. Integral to this research are investigations of the influence of the processes on the fate of inorganic contaminants, including uranium and technetium.

Steven Gill is associate investigator at the Institute of Genomic Research. He is project leader of the *Staphylococcus aureus* genome project. His research interests include genomics of microbial pathogens, microbial pathogenesis and host interactions, gene regulation and protein networks, and cytoskeletal-based motility.

Alan Hastings is professor in the Department of Environmental Science and Policy and a member of the Institute of Theoretical Dynamics and the Center for Population Biology at the University of California, Davis. His research interests include theoretical ecology (structure populations, and chaos) and theoretical population genetics (multilocus theory and quantitative genetic models). His current research projects focus on metapopulation dynamics, transients in ecological systems, fitting models to data in ecology, spatial dynamics in ecology (tussock moths and marine populations with mero-planktonic larvae), and hybrid zone dynamics. He received his PhD in applied mathematics from Cornell. He was president of the Society for Mathematical Biology in 1999-2001 and organizer of the NSF Workshop on Quantitative Environmental and Integrative Biology (2000). He also is managing editor of the *Journal of Mathematical Biology*, associate editor of *Theoretical Population Biology*, and a member of the editorial board of *Mathematical Biosciences* and served on the editorial board of *Oecologia* in 1997-2001.

John C. Houghton directs programs in computational biology and selected aspects of the Genomes to Life research program and directs the Integrated Assessment of Global Climate Change research program in the Department of Energy (DOE) Office of Science. He is the point of contact in the Office of Science for the DOE hydrogen program.

Matthew D. Kane is program director in the Division of Molecular and Cellular Biosciences at the National Science Foundation (NSF). His research interests range from specific study of the ecology and evolution of insect-microorganism to general aspects of the taxonomic, genomic, and physi-ological diversity of microorganisms. At NSF he manages or comanages a wide variety of programs that support microbial research, including Micro-bial Observatories and Microbial Interactions and Processes, the Interagency (NSF-US Department of Agriculture) Microbial Genome Sequencing Pro-gram, Biocomplexity in the Environment: Genome-Enabled Environmental Science and Engineering, and Assembling the Tree of Life.

Jared Leadbetter is assistant professor of environmental microbiology at the California Institute of Technology. His research seeks to clarify the form, function, and spatial distribution of diverse microorganisms in their environment. His studies have focused on the cultivation of microbial strains possessing unusual, atypical, or previously unrecognized properties and have sought to reveal the impact of these organisms on their environ-ment. He has applied a number of physiological, chemical, and molecular genetic techniques to his studies, underscoring the utility of both classical

and newly developed methods in the pursuit of fundamental questions in environmental microbiology.

Derek R. Lovley is Distinguished University Professor and head of the Department of Microbiology at the University of Massachusetts, Amherst. He received his BA from the University of Connecticut and his MA from Clark University in biological sciences and his PhD from Michigan State University in microbiology. His research focuses on the physiology and ecology of novel anaerobic microorganisms. Current topics of investigation include microbial metabolism and community structure in the deep subsurface and hydrothermal zones, evolution of anaerobic respiration, mechanisms of electron transport to Fe(III) and humic acids, anaerobic bioremediation of petroleum-contaminated subsurface and aquatic habitats, and bioremediation of metal contamination. These studies are being approached at the genetic, biochemical, and ecological levels.

Michael Savageau is professor of bioengineering at the University of California, Davis. He received his PhD in cell physiology and systems science from Stanford University. His research interests are in quantitative theory of organizationally complex biological systems, generic nonlinear methods for mathematical and computer analysis, application of theory to specific classes of cellular and molecular networks, elucidation of biological design principles, and design and construction of novel gene circuits. He was a Guggenheim Fellow and Fulbright Senior Research Fellow at the Max-Planck Institut für Biophysikalische Chemie in Gottingen and is a Fellow of the American Association for the Advancement of Science. He has served or serves in an editorial role for *Mathematical Biosciences, Journal of Theoretical Biology, Mathematical Ecology and Evolution Series, Biocomplexity,* and *Nonlinear World.*

Gary S. Sayler is professor of microbiology and ecology and evolutionary biology and director of the Center for Environmental Biotechnology at the University of Tennessee, Knoxville. His major research interests include ecological and toxicological impact of environmental contaminants on microbial communities, biodegradative mechanisms, plasmids and transposons, the application of molecular methods in analysis of biodegradative microbial community structure and function, and genetic-engineering strategies for biodegradative and biosensing organisms.

Thomas M. Schmidt is associate professor of microbiology at Michigan State University. His interests and expertise are in microbial physiology and ecology, organization and activity of microbial communities, relationships between genes, genomes, and ecologic strategies of microorganisms.

Research in his laboratory focuses on microbial physiology and ecology. Members of his laboratory routinely develop and apply techniques of molecular biology to explore the diversity of microbial communities without the bias potentially introduced by cultivation methods.

Hal Smith is professor of mathematics at Arizona State University. His research interests include differential equations, dynamical systems, and mathematical biology. He is studying dynamical systems in biology.

Eberhard O. Voit is professor in the Department of Biometry and Epidemiology and the Department of Biochemistry and Molecular Biology at the Medical University of South Carolina. He is in charge of bioinformatics and systems biology. His research interests focus on the analysis of organizationally complex biomedical systems, particularly metabolic pathways. He uses biochemical systems theory—defined by highly structured sets of nonlinear ordinary differential equations—to study complex systems that are characterized by numerous components that interact in a nonlinear fashion and can exhibit emerging responses not possible in smaller subsystems. Their function and design can be fully understood only if they are represented and analyzed mathematically and computationally.

Lyle Whyte is associate professor in the Department of Natural Resource Sciences at McGill University, Canada. Prior to joining McGill University, he was research officer at the National Research Council of Canada. He received his BSc from the University of Regina, Canada, and his PhD from the University of Waterloo, Canada. His research interests include genetics, physiology and ecology of cold-adapted microorganisms able to degrade a variety of pollutants at cold temperatures, polar microbial ecology and biodiversity, and development of genome-based tools and their applications for studying microbial ecology.

Jizhong Zhou is distinguished scientist and director for the Environmental Microbial Genomics Program at Oak Ridge National Laboratory. He received a BS in plant pathology and entomology, MS in mathematical ecology, and PhD in molecular genetics and cell biology. His research focuses on microbial functional genomics, genomics technology, and microbial ecology. He received the Presidential Early Career Award for Scientists and Engineers in 2001, and the Environmental Sciences Division Distinguished Scientific Achievement Award in 2001. He has authored more than 80 publications on molecular biology, molecular evolution, microbial ecology, bioremediation, and theoretical ecology. He chaired the 7th, 9th and 11th International Conference on Microbial Genomes.

Observer

Karl Koehler, National Science Foundation

Staff

Evonne Tang
Robin Schoen
Fran Sharples
Bridget Avila
Danielle Greene

Science writer

Patricia McAdams