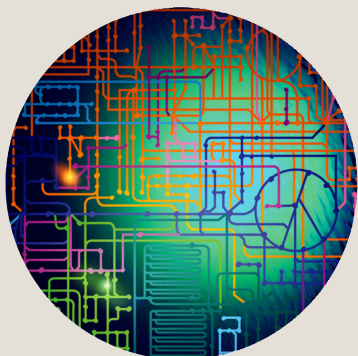


1.0. Genomics:GTL Roadmap Strategy

1.1. Industrial Biotechnology and DOE Missions	2
1.2. Genomics and Systems Biology	4
1.3. Genomes to Life: Achieving a Predictive Understanding of Microbial Function.....	5
1.3.1. GTL User Facilities: Performance, Throughput, and Cost	6
1.3.2. Computing and Information Science	8
1.3.3. Power of the GTL Knowledgebase: Economies Provided by Nature.....	8
1.3.4. Bridging the Gap Between Big and Small Science—The Need for a Third Model.....	9
1.3.5. Department of Energy: Experienced at Large-Scale Projects	9
1.3.6. Implementation: What is the Time Frame, Who will be Involved, and How will Decisions be Made?	10
1.3.6.1. Three-Phase Implementation of the GTL Program.....	10
1.3.6.2. Integrated Management and Development.....	11
1.3.6.2.1. GTL Program and Facility Governance	12
1.3.6.2.2. Facility Development and Acquisition Process	12
1.3.6.2.3. Community Involvement in GTL Technology Development	12
1.3.6.2.4. Communication in a Multidisciplinary Environment	12
1.3.6.2.5. Facility User Access	12
The Microbial World: A Challenging Frontier	13

To accelerate GTL research in the key mission areas of energy, environment, and climate, the Department of Energy Office of Science has revised its planned facilities from technology centers to vertically integrated centers focused on mission problems. The centers will have comprehensive suites of capabilities designed specifically for the mission areas described in this roadmap (pp. 101-196). The first centers will focus on bioenergy research, to overcome the biological barriers to the industrial production of biofuels from biomass and on other potential energy sources. For more information, see Missions Overview (pp. 22-40) and Appendix A. Energy Security (pp. 198-214) in this roadmap. A more detailed plan is in *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda*, DOE/SC-0095, U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy (<http://genomicsgtl.energy.gov/biofuels/>).



Genomics:GTL Roadmap Strategy

GTL aims to understand biological systems well enough to predict their behavior accurately with mechanistic computational models. The ultimate result will be the ability to use the biochemical sophistication of microbes for a broad range of innovative applications, a potential achievable only with huge gains in research performance, productivity, efficiency, cost-effectiveness, and quality.

GTL Hallmarks

- Global, genome-wide view of microbial functions
- Advanced technologies with improved performance to provide comprehensive data sets
- Facilities to dramatically improve throughput, cost, and data quality
- Comparative analysis at all levels enabled by the GTL Knowledgebase (genes → molecules → cell processes → cell and community function)
- Modeling and simulation tools for predictive understanding to enable in silico biology
- Open access to data, protocols, and facilities

*Biotechnology includes applications beyond clinical medicine to those in agriculture and, more recently, industry and the environment. For additional information, see the Biotechnology Industry Organization's web site (www.bio.org).

1.1. Industrial Biotechnology and DOE Missions

Industrial biotechnology* is critical for the future of the nation and will be an engine of economic competitiveness in this century. Biological solutions must contribute to the mix of technologies needed for the key DOE missions of energy security, environmental restoration, and climate change. These missions, shown in Fig. 1. GTL Science and Technology Foundations for DOE Missions, this page, are grand challenges for the development of bioscience and biotechnology, with the potential for trillions of dollars of impact. DOE missions require that we understand biology at every level, from the most detailed molecular processes to vast natural ecosystems. Such knowledge will point the way for countless applications, leading to new industries and economic and social benefits (Brady, Chao, and Clardy 2002; Luengo 2003). Broad interest in the development of the necessary science base will accelerate the pace of discovery and increase the scope of its

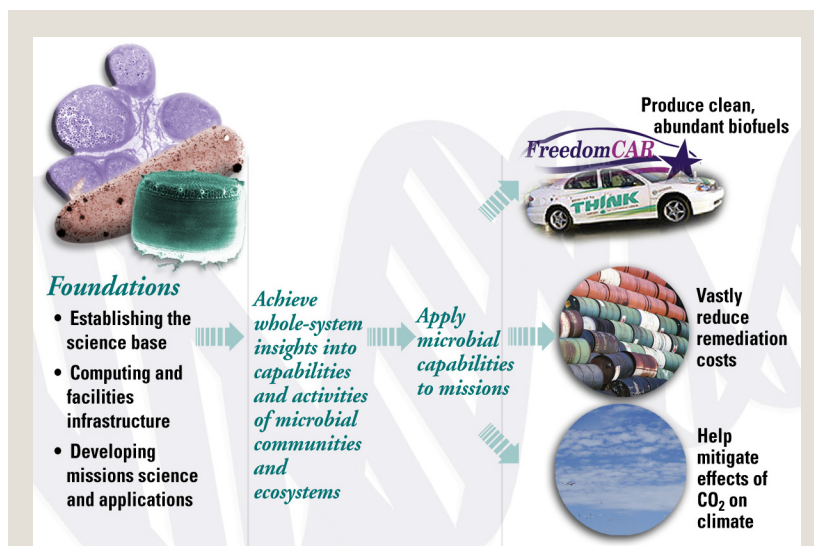


Fig. 1. GTL Science and Technology Foundations for DOE Missions.

GTL ROADMAP STRATEGY

impact. While the Human Genome Project opened the door for improvements in human health, GTL is the gateway to biotechnological solutions to DOE mission problems and to stimulating new generations of industrial biotechnology (Herrera 2004; Littlehales 2004). DOE's mission challenges are the following:

- Develop biofuels as a major secure energy source.
- Develop biological solutions for intractable environmental problems.
- Understand relationships between climate change and earth's microbial systems; assess options for sequestration.

See 2.0. Missions Overview, p. 21, and DOE mission appendices: A. DOE Mission: Energy Security, p. 197; B. DOE Mission: Environmental Remediation, p. 215; and C. DOE Mission: Carbon Cycling and Sequestration, p. 227.

The National Research Council, in its 1999 report, *Biobased Industrial Products: Research and Commercialization Priorities*, acknowledged the great potential economic impact of industrial biotechnology, noting:

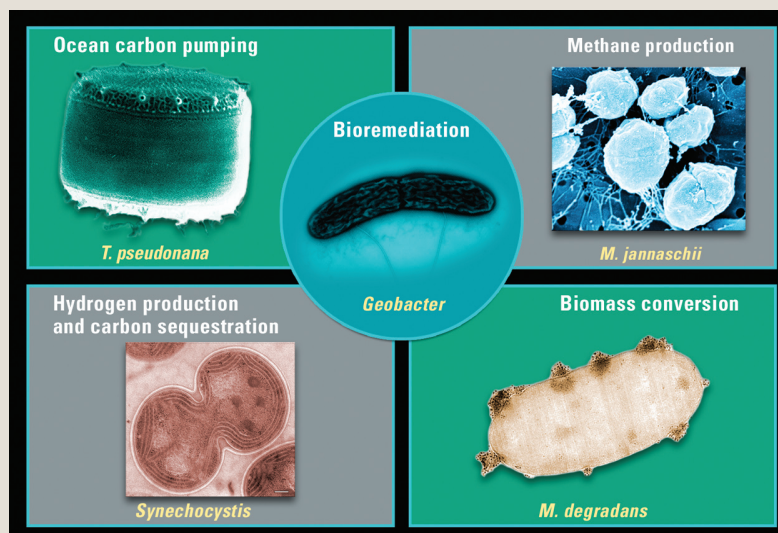
“ Biological sciences are likely to make the same impact on the formation of new industries in the next century as the physical and chemical sciences have had on industrial development throughout the century now coming to a close. ”

The Economist also championed the promise of biotechnology:

“ At the moment, biotech's main uses are in medicine and agriculture. But its biggest long-term impact may be industrial [Editorial, “Saving the World in Comfort” (March 27, 2003)]. ”

Microbes can provide the basis for a revolution in industrial biotechnology. These untapped natural treasures are the foundation of the biosphere and sustain all life on earth. Extreme genetic diversity and the ability to function in complex communities give microbes extraordinary biochemical capabilities and adaptability (see Fig. 2. Using the Natural Diversity of Microbes to Create Biotechnology Solutions for DOE Missions, this page). These single-celled organisms are masters at living in almost every environment and harvesting energy in almost any form, from solar radiation to mineral chemistry, and transforming it into chemical compounds that power life. By understanding how microbes function in their many environments, we can reveal their contributions to earth ecosystems and their relationships to climate change. We also can understand how they can provide the basis for environmental remediation and for creating new sources of renewable, less-polluting energy sources and new generations of processes for industrial application (see Fig. 1, p. 2, and sidebar section, The Microbial World, beginning on p. 13).

Fig. 2. Using the Natural Diversity of Microbes to Create Biotechnology Solutions for DOE Missions. Microbes, which have been adapting to countless environments for some 3.5 billion years, offer an untapped reservoir of sophisticated chemistries. Harnessing their functionalities can result in novel and highly efficient strategies for producing hydrogen and other fuels, cleaning up toxic waste at contaminated DOE sites, and capturing and storing carbon from the atmosphere to help mitigate global climate change.



T. pseudonana: B. Palenik and D. Lee, Scripps Inst. of Oceanography;
M. jannaschii: R. Weiner, Univ. of Md., College Park; *M. jannaschii*:
B. Boonyaratankomkiet, D. S. Clark, G. Vrdoljak, Univ. of Calif., Berkeley

1.2. Genomics and Systems Biology

Every organism's genome encodes its ability to create and sustain life. Genome data provide the foundation for studying biological processes rather than examining isolated parts. The longstanding successful approach to biological research—variously described as “single gene,” “reductionist,” or “linear”—is piecemeal and, while productive, is less efficient at expeditiously addressing questions of biological complexity and integration.

The new approach for exploration—systems biology—will allow us to envision the microbe as a complete set of intersecting processes and to create models for simulating how microbes operate and respond. This is a major first step toward illuminating the most fundamental principles of living cells and achieving a predictive understanding of the scale and complexity of natural systems.

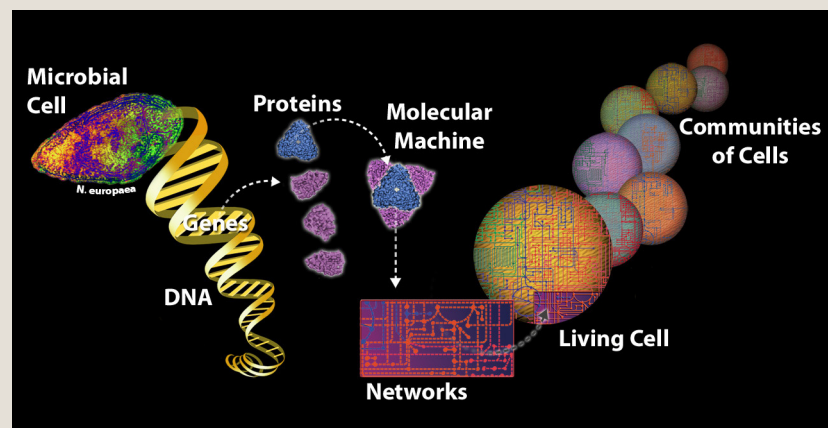
A comprehensive approach to understanding biology encompasses many cellular components. While a genome is a fixed catalogue of information, it dynamically creates the machinery of a cell in response to the changing environment. Thousands of genes encode even greater numbers of proteins that mediate biology in a “just-in-time” strategy by associating in myriad ways (protein “machines”) to form intricate pathways and networks within a cell (see sidebar, The Basics: From DNA to Living Cells to Communities, this page). Demonstrating the power of these finely tuned systems, microbes rapidly respond to environmental cues by adjusting their entire cellular operation (O'Toole 2003). (See 3.2.2. Science and Technology Milestones, p. 44, and 5.0. Facilities Overview, p. 101.)

Microbes, in their adaptability, also have the best of both the single- and multicellular worlds. Using mechanisms we are only beginning to understand, microbes carry on a dialogue that establishes community and environmental awareness and enables individual microbes to function together as multicellular organisms (e.g., “biofilms”) in complex geochemical environments, with the many benefits that can provide. Functions include sensing the environment; assembling appropriate cells or communities of cells as environmental conditions change; regulating and carrying out cellular function, including critical energy capture and manipulation; and providing for reproduction, sporulation, or senescence as conditions dictate (Check 2002; see sidebar, Life in a Biofilm, p. 18).

We can now rapidly and accurately decode the genomes of microbes and microbial communities in complex natural ecosystems (metagenomes). While more than 99% of microbes historically have been hidden from study because they could not be cultured (Handelsman et al. 1998), genomics allows the assessment of these microbial systems to determine who's there and what some of their functionalities are. Other emerging technologies such as imaging will

The Basics: From DNA to Living Cells to Communities

- **Cells** contain DNA—the hereditary material of all living systems.
- The **genome** is an organism's complete set of DNA.
- **DNA** contains genes whose sequences specify how and when to build proteins.
- **Proteins** perform most essential life functions, often working together as molecular machines. In addition, they form most cell structures.
- **Molecular machines** interact through complex, interconnected pathways and **networks** to make the working cell come alive.
- **Communities of cells** are associations of microbes (each a single cell) working together in a particular environmental niche.



N. europaea: M. A. Bruns, Center for Microbial Ecology, Michigan State University

allow us to track molecules and cells in complex living systems to add a functional perspective without the need for classic culturing.

Not only has genomics enabled early insight into these complex systems, it also has revealed a vastly greater microbial diversity and presence than was previously appreciated (Stein et al. 1996; Meyer 2004; Schaechter, Kolter, and Maloy 2005). Recent genomic studies of microbial communities already have led to the discovery of millions of genes and proteins, thousands of species, and innumerable variations in critical functionalities (Venter et al. 2004), establishing the globe's vast microbial communities as a potentially rich resource for understanding biology and for catalyzing industrial biotechnology (Schloss and Handelsman 2003; Riesenfeld et al. 2004).

Functions of about 40% of sequenced genes, however, remain unknown or poorly characterized—a challenge that will be prevalent in most DOE-relevant systems. As a further objective, this program aims to achieve a holistic, mechanistic understanding of biology. Genes provide life's potential list of components, but complete mechanisms of function can be revealed only after additional experiments and analyses are performed. The research community (Buckley 2004b; Roberts 2004) recognizes that we need new approaches for studying microbes and other systems efficiently (Aebersold and Watts 2002; Roberts et al. 2004) to understand gene and systems function.

1.3. Genomes to Life: Achieving a Predictive Understanding of Microbial Function

GTL aims to understand biological systems well enough to predict their behavior accurately with sophisticated computational models. GTL analyzes critical microbial properties and processes on three fundamental systems levels (see 3.0. GTL Research Program, p. 41, and 4.0. Creating an Integrated Computational Environment for Biology, p. 81).

- **Molecular:** Focusing on genes, proteins, multicomponent protein complexes, and other biomolecules that provide structure and perform the cell's functions.
- **Whole cell:** Investigating how molecular processes, networks, and subsystems are controlled and coordinated to enable such complex cellular processes as growth and metabolism.
- **Microbial community:** Exploring how diverse microbes interact to carry out coordinated complex processes enabling microbes to both respond to and alter their environments.

Conceptually, genomes contain all the information needed to deduce function, yet the intricately detailed biology underlying life creates a huge barrier to a facile connection between genome sequence and function. The GTL strategy is to provide the technologies, computing infrastructure, and comprehensive knowledge-base to break through the barrier of complexity that prevents the direct translation of genome sequences into predictive understanding of function. This new opportunity grows out of rapid advances in instrumentation for the biosciences, exponential improvements in computing speeds and modeling capabilities, and a growing interest by physical and information scientists in applying these methods to biological problems. The sequences furnish the blueprint for exploration, technologies can produce the data, and computing can relate these enormous data sets to models of process and function (Ellis et al. 2004; Kitano 2002).

Understanding biological systems at these three levels, however, is a daunting task: While the genome of a microbe is a fixed code that represents megabytes of information, a full description of all the dynamic processes involved in making a living cell operate and respond to its environment will be a mixture of complex data sets potentially in the petabyte range (i.e., a billion megabytes) (see 5.3. Facility for Whole Proteome Analysis, Table 1. GTL Data, p. 159). Gathering these data will require huge gains in performance, data quality, productivity, and cost-efficiency. New generations of computing and information capabilities thus are needed to manage, analyze, and transform the information into accurate models. In addition, for biology to have timely impacts on national needs, the time required to achieve fundamental understanding of a system must be accelerated from many years to months, compounding the challenge (see The Framework for DOE Missions, p. 24).

GTL ROADMAP STRATEGY

A national priority is to achieve an unprecedented understanding of the natural world:

“ The application of DNA sequence and other data allows the development of new biotechnological tools, such as microarrays to decipher the functional implications of gene expression, which are helping to unravel longstanding questions in biology. Agencies should target investments toward the development of a deeper understanding of complex biological systems through collaborations among physical, computational, behavioral, social, and biological researchers and engineers. These collaborations will yield new ways of collecting and analyzing data allowing for the exploration of the living world across all levels of biological organization both spatially and temporally. . . . Federal agencies should continue to invest in obtaining additional sequence data and in the development of genomics tools and resources (Marburger and Bolton 2004, www.ostp.gov/html/m04-23.pdf). ”

GTL aims to bring biologists, physical scientists, and computing scientists together to establish a new biology for meeting energy, environmental, and climate needs (Frazier et al. 2003a; 2003b).

1.3.1. GTL User Facilities: Performance, Throughput, and Cost

Attaining the ability of the scientific community to analyze microbial systems on a timetable that supports applications in years rather than decades requires that we begin generating and assimilating materials and data on a scale that far exceeds today's capacities. Genome-sequencing projects have shown that pursuing transformational production goals in dedicated facilities can result in such gains (see sidebar, High-Throughput Model Guides Future Facilities, this page). Consequently, to advance its research goals, DOE plans to develop four cost-effective, high-throughput user facilities for systems microbiology research, in addition to supporting a broad range of research projects in GTL. Achieving economies of scale, the GTL suite of four facilities is designed to provide a phased and integrated set of analytical and production capabilities to determine microbiological structure and function from the genomic through the ecosystem levels. This is depicted in the sidebar, GTL Facilities: Accelerating Scientific Discovery and Applications Research for Energy and Environment, p. 7, and in the Facilities Overview and subsequent descriptions beginning on p. 101.

Each of these facilities is important in its own right, but all are intricately linked in their long-term goals, targets, technologies, capabilities, and capacities. They will provide production and analytical resources for scientists to collect and use the information needed to put microbes and their capabilities to work. While vital for GTL progress, the facilities also will help accelerate biological research supported by other agencies.

High-Throughput Model Guides Future Facilities

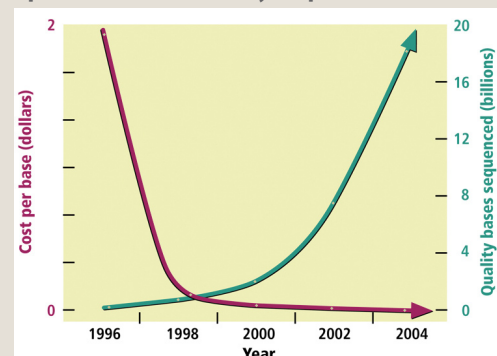
The dramatically increased productivity and reduced costs achieved in the Human Genome Project via high-throughput production environments (e.g., the DOE Joint Genome Institute) provide the paradigm for dedicated industrial-scale facilities envisioned for GTL systems biology research. These resources democratize cutting-edge science, enabling even the smallest research laboratory to participate.

A growing mandate from the scientific community echoes the need for systems biology facilities:

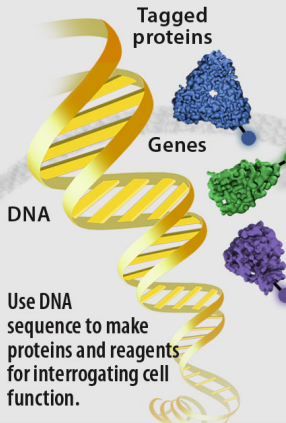
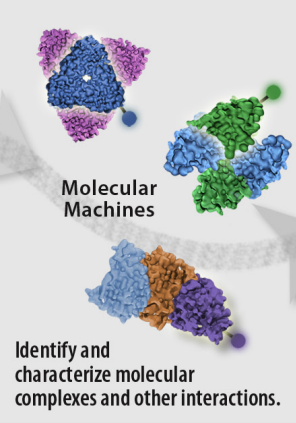
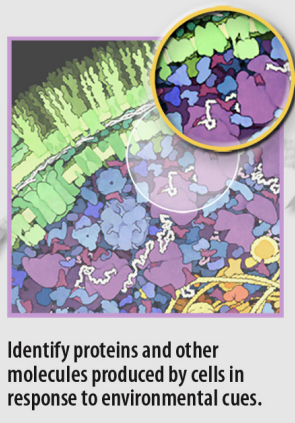
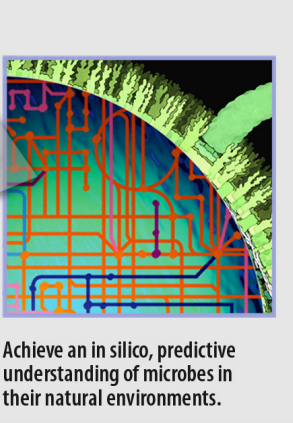
“To make progress, science should not accept the limitations placed on discovery by traditional methods, conventional approaches, or existing infrastructure. Powerful, but expensive, modern equipment should be housed in community facilities, open to researchers who might not otherwise have access to these technologies.”

[Source: *Microbiology in the 21st Century: Where Are We and Where Are We Going?* American Society for Microbiology (2004)]

Large-Scale Genome Sequencing Facilities Spur Cost, Productivity Improvements

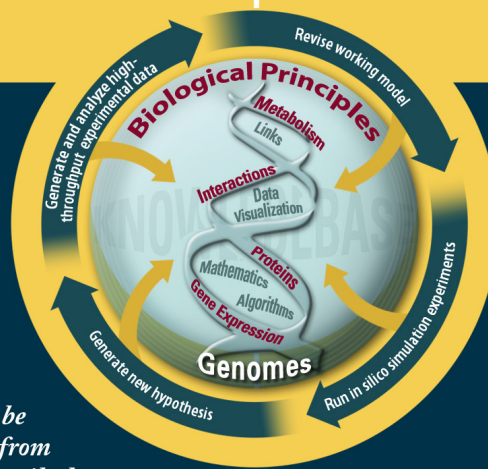


GTL Facilities: Accelerating Scientific Discovery and Applications Research for Energy and Environment

Protein Production and Characterization	Molecular Machines	Proteomics	Cellular Systems
 <p>Use DNA sequence to make proteins and reagents for interrogating cell function.</p>	 <p>Identify and characterize molecular complexes and other interactions.</p>	 <p>Identify proteins and other molecules produced by cells in response to environmental cues.</p>	 <p>Achieve an in silico, predictive understanding of microbes in their natural environments.</p>
Production and Characterization of Proteins and Molecular Tags	Characterization and Imaging of Molecular Machines	Whole Proteome Analysis	Modeling and Analysis of Cellular Systems
<ul style="list-style-type: none"> ▶ Produce proteins encoded in the genome. ▶ Create affinity reagents that allow each protein to be identified, located, and manipulated in living cells. ▶ Perform biophysical and biochemical characterizations of proteins produced to gain insights into function. 	<ul style="list-style-type: none"> ▶ Isolate and analyze molecular machines from microbial cells. ▶ Image structure and cellular location of molecular machines. ▶ Generate dynamic models and simulations of molecular machines. 	<ul style="list-style-type: none"> ▶ Measure molecular profiles and their temporal relationships. ▶ Identify and model key pathways and other processes to gain insights into functions of cellular systems. 	<ul style="list-style-type: none"> ▶ Integrate knowledge and models to understand the structure and functions of cellular systems, from single cells to complex communities. ▶ Integrate imaging and other technologies to analyze molecular species from subcellular to ecosystem levels as they perform their functions.

Understanding how the information in a genome dictates cellular functions requires knowledge of a cell's molecular complement, interactions, and regulation. These studies must be carried out on a scale far exceeding today's capacities.

Microbial genome sequences will be the foundation on which all data from the large-scale GTL facilities (described above) are related.



- ▶ Comprehensive integration of GTL and research community databases
- ▶ Transparent and intuitive access to computational tools
- ▶ Simulations of microbial behavior using genome sequences as input
- ▶ Information and support for research, policy, and applications

Systems Microbiology Knowledgebase to Enable a Predictive Understanding of Microbes and Communities

YGG 04-0172R4

1.3.2. Computing and Information Science

GTL is as much a computing program as a biology program. Computational modeling is at the heart of the research described in this roadmap. The new biology must tightly integrate computational analysis and experimental characterization of biological systems and combine linked measurements using dozens of sophisticated new analytical instruments to create huge multivariate data sets and relate them to models at the appropriate systems level.

The vision for the GTL computing environment encompasses the creation of a seamless enterprise with support mechanisms for collaboration via transparent scientific-community access to data and tools. Features include the following.

- **Open access to data.** Furnish easily accessible, large-scale data archives and community databases containing enabling data, knowledge, and models of biological systems.
- **Open access to tools.** Provide powerful suites of analysis, data-mining, modeling, and simulation tools that enable GTL facilities, projects, and end users to interpret, understand, and predict the behavior of biological systems.
- **Accessible infrastructure.** Provide computing hardware, operating software, data storage, and network capabilities to support large-scale systems biology conducted by a diverse research community.

The GTL vision puts biology on the same path that much of research and industry have followed, with computation, modeling, and simulation as an integral part of the research process. As the performance of computing increases and costs decline, modeling and simulation are critical to experimentation that is becoming ever more complex, time consuming, and costly. Computation provides the insights needed to make experimentation more focused and informative. Transforming biology into a quantitative and predictive science based on models and data will accelerate discovery and ultimately shorten the technology-development cycle, more rapidly yielding practical national benefits (see 3.2.2.3. Milestone 3, p. 51; 4.0. Computing, p. 81; and computing roadmaps for each facility in 5.0. Facilities Overview, p. 101).

1.3.3. Power of the GTL Knowledgebase: Economies Provided by Nature

Analysis of each new microbe benefits from knowledge about all other microbes and life forms because of nature's simplifying principles. Just as a finite number of rules determine the structure and function of proteins, so the higher-order functions of cells seem to emanate from another finite set of rules. Once successful machines, pathways, and networks arise, they tend to be preserved, subtly modified and optimized, and then reused as variations on enduring themes throughout many species. Thus, accumulating detailed information on numerous microbes across a wide range of functionalities will provide the insight needed to interpret these principles. To take full advantage of this phenomenon, relating all known information generated computationally or experimentally to the genome (annotation) is an important task that must be performed continually as new genomes are sequenced or new experiments performed.

In this new era of systems biology, all-against-all comparisons of extensive microbial data amassed in the GTL Knowledgebase will accelerate and sharpen our research strategies. Along with high-throughput facilities and computing, this capability is a key element of our approach to reducing the analysis time for a microbial system. Given a knowledgebase with many genes from organisms highly annotated with functional data (cross-referenced to each other), much information about a newly sequenced genome will be at scientists' fingertips. Comparative genomics, founded on these principles, ultimately will allow us to predict the functions of unknown microbes by deriving a working model of a cell from its genetic code.

This paradigm combines (1) discovery science as we navigate huge unexplored data sets that can reveal unforeseen properties and phenomena and (2) computationally driven hypothesis science to derive insights into previously unfathomable complexity (see conceptual diagram of GTL Knowledgebase in sidebar at bottom of p. 7; 3.2.2.3.2. GTL Knowledgebase, p. 52; and 4.2.1. Theory, Modeling, and Simulation Coupled to Experimentation of Complex Biological Systems, p. 85).

1.3.4. Bridging the Gap Between Big and Small Science—The Need for a Third Model

The biology community faces a complex, yet critical, challenge: Preserve the creativity and entrepreneurial spirit of the single investigator in light of the increasingly sophisticated and costly resource requirements of leading-edge biological research. Making the most advanced technologies and computing resources available to the research community will democratize access to the tools needed for systems biology. The GTL research program, its integrated computational environment, and user facilities are designed for DOE science needs but also will help bridge the capability gap between large and small labs (Relman and Strauss 2000). Already, only a minority of even large laboratories can afford to be adequately equipped. Individual investigators need the capabilities of big science, and scientific progress requires diverse contributions from the whole scientific community. One benefit to DOE in providing these capabilities will be the involvement of a larger scientific community in the important mission challenges we face. Another benefit to the research community is the availability from industrial vendors of instrumentation and processes developed to meet GTL's requirements. A key point is that GTL facilities will not supplant or compete with investigator-initiated science but rather will complement and enhance it, just as high-throughput sequencing facilities (e.g., the DOE Joint Genome Institute) do today.

The facilities will open new avenues of inquiry, fundamentally changing the course of biological research and greatly accelerating the pace of discovery. Daunting technical, time, and cost barriers to a concerted and comprehensive approach to biology will be removed, allowing scientists to aspire to a higher-level perspective and higher-value research.

“ The ecosystem predictive capability will come from detailed work that links an understanding of the genome to an understanding of gene expression, protein function, and complex metabolic networks. We must create centers that facilitate research community access to postgenomic analytical capabilities. [The Global Genome Question: Microbes as the Key to Understanding Evolution and Ecology, American Society for Microbiology, 2004]

1.3.5. Department of Energy: Experienced at Large-Scale Projects

The Department of Energy's Office of Science (SC), in addition to providing major facilities for the scientific community, conducts research on problems in fundamental science, energy, the environment, and climate that involve large interdisciplinary teams from many institutions pursuing strategic science goals. To make the necessary advances in biology over the coming decades, this research model will be critical to the comprehensive study of biology, ultimately impacting DOE mission problems (see sidebar, GTL User Facilities Leverage DOE Experience and Skills, this page).

Multidisciplinary Teams. Organized and focused teams are needed because meeting these great challenges will require many skills and capabilities in complementary and supportive roles to foster new thinking and approaches (Nass and Stillman 2003). The teams will take advantage of the unique convergence of disciplines that has occurred over the past decade;

GTL User Facilities Leverage DOE Experience and Skills

DOE is the leading funder of physical sciences in the nation. For more than half a century, its Office of Science (SC) has envisioned, designed, constructed, and operated many of the world's premier research facilities. These facilities continue to grow in importance to biology and today serve 20 times as many users from the life sciences community as in 1990.

DOE now seeks to bring its experience and skills in physical sciences to finding biological solutions to mission challenges. GTL facilities are among those featured in SC's 20-year plan (*Facilities for the Future of Science: A Twenty-Year Outlook*, 2003, www.science.doe.gov), with the director noting that “Investment in these [GTL] facilities will yield extraordinary scientific breakthroughs and vital societal and economic benefits.” This plan was developed through discussions with and assistance from the scientific community.

GTL ROADMAP STRATEGY

several fields have arrived simultaneously at the same scale and complexity levels from different directions. This meeting of biology, computing, computational chemistry, materials science, synthetic and analytical sciences, microtechnologies, and, most recently, nanoscience and nanotechnology allows us to explore the amazing nanoworld of microbes with rich and powerful probes and methods.

Through this strategic approach to research and in the proposed GTL facilities, DOE's strengths in the biological, physical, and computational sciences will bring breakthrough technologies to bear on biology. GTL program sponsors, the Office of Biological and Environmental Research (BER) and the Office of Advanced Scientific Computing Research (OASCR), have made this commitment.

DOE supports a wide range of applied research and technology development at the national laboratories and in academia through its offices of Energy Efficiency and Renewable Energy (EERE), Fossil Energy (FE), and Nuclear Energy (NE). Through advanced technology development and commercialization programs, DOE works with industry to bring new tools and processes to the marketplace. GTL will be coordinated with these programs and will provide its research resources, supporting facilities, and computing infrastructure to enable the timely migration of new science and scientific capabilities into impactful technologies for DOE missions.

This coordinated approach is consistent with the DOE research tradition.

“ This capacity to deal with both the scale and complexity of these efforts is especially important in today's rapidly changing world. As we continue to gain the ability to work at very small scales and to probe the dynamic, three-dimensional structure of molecules, the interface between physical science and life science is of critical importance. Work at the interface of frontier disciplines like bioinformatics, genomics, proteomics, and nanotechnology is greatly enhanced by DOE's capacity in large-scale computation and research tools based upon physical science. [*Critical Choices: Science, Energy, and Security: Final Report of the Secretary of Energy Advisory Board's Task Force on the Future of Science Programs at the Department of Energy*, October 13, 2003] ”

1.3.6. Implementation: What is the Time Frame, Who will be Involved, and How will Decisions be Made?

1.3.6.1. Three-Phase Implementation of the GTL Program

GTL is following a roadmap (see Table 1. GTL Science, Technologies, and Applications Roadmap, p. 11) to do the critical science to establish systems biology; to develop the necessary tools, resources, and facilities; and to complete the transformation to true systems biology aimed at mission applications. The strategy is divided into three distinct phases:

- **Phase I (First 8 years): Genomics to Systems Biology.** Transition from genomics to systems biology will include key proof-of-principle experiments in systems biology and technology prototyping and piloting that will create the science and technology base and begin to train a community of scientists in systems biology. The design, R&D, and early deployment of critical computing and information tools and infrastructure will provide the necessary foundation for data management, analysis, and modeling. Intense planning activities involving scientists, technologists, and mission programs will support the conceptualization, design, R&D, and construction of GTL production facilities.
- **Phase II (9 to 16 years): Technology Integration and Scaleup.** Production facilities will provide an engine for dramatically accelerating the study of microbial systems and for discovering and developing predictive systems understanding—ultimately reducing the time for analysis of a microbial system from years to months.
- **Phase III (16+ years): Biological Systems Knowledge for DOE Applications.** Knowledge and capabilities developed in GTL will be provided for rapid and complete systems studies of important problems in science and for useful applications. This knowledge and these capabilities will position GTL to rapidly transform new science into revolutionary new processes and products to help meet critical DOE national needs.

1.3.6.2. Integrated Management and Development

A concerted planning and management process will allow a more aggressive technical and scientific strategy and optimize the use of research program resources and R&D investments. Key issues include:

- Establishing program and facilities governance to ensure the best science and most useful technologies.
- Coordinating GTL program needs with facility design and development.
- Developing an integrated computational infrastructure for the facilities and program.
- Coordinating development activities across facilities and within the program to accommodate interdependencies and to optimize the use of resources.
- Establishing and coordinating workshops and working groups for solving problems and developing technologies.
- Understanding the needs of DOE mission applications groups to support use of GTL facilities for biotechnology development.
- Identifying promising technological areas that have global applicability or require long lead times.
- Building a communication strategy for public and community outreach and information gathering.
- Making GTL facilities accessible to the broader research community on a peer-reviewed basis.

Specific examples of development needs are outlined in 6.0. GTL Development Summary, p. 191.

Table 1. GTL Science, Technologies, and Applications Roadmap

Science Base		
Genomics to systems biology Molecular, cellular, and community studies Key insights and strategies for study	High-throughput study of key systems and processes Comparative analyses, systems modeling, and simulation Fully integrated systems for experimentation and computing	Integrated knowledgebase for designing mission solutions Science and technology development and application Science of innovation and next-generation concepts
Technologies, Computing, and Facilities		
Advanced technology development and testing Pilot studies, computing, and technology scaleup Facilities researched, developed, designed, and built	Facilities' operations: Comprehensive data Integrated data and computing capabilities operational Full biological systems data available in months	Facilities applied to engineered systems Tested, evaluated, monitored, and verified New functions and advanced concepts engineered
Applications of Missions Science		
Key systems chosen for mission interest Insights into cellular and community processes and interactions Systems data needs and strategies	Mission model systems analysis begun Understanding leading to engineering strategies for missions Application-specific strategies set	Fully engineered systems designed and developed Engineered systems tested and evaluated First generation fielded, next generation developed

2002

Genomics to Systems Biology

Technology Integration and Scaleup

Biological Systems Knowledge for DOE Applications

Phase I

8 Years

Phase II

16 Years

Phase III

1.3.6.2.1. GTL Program and Facility Governance

DOE will establish a governance process to ensure advancement of DOE, GTL, and research-community objectives; excellence in science; optimized facility access and operations; and continuous facility and equipment enhancement. Governance will include appropriate scientific and technological advisory groups and peer review for access and resource allocations (see 3.4. GTL Program and Facility Governance, p. 77, and 5.0. Facilities Overview, p. 101).

1.3.6.2.2. Facility Development and Acquisition Process

The GTL user facilities will be developed and acquired using a process based on the robust criteria described in DOE Order 413.3 (www.science.doe.gov/opa/PDF/O4133.pdf), which ensures the successful design, implementation, and management of DOE facilities. This construction-project management process includes a rigorous assessment of the science objectives, technology requirements and development needs, facility requirements, and staffing. The process encourages a sound technology-baseline definition and supports research, design, development, testing, and evaluation as needed to ensure that the facilities and all associated equipment and information systems perform at specifications. The scientific and technological community is involved extensively through workshops, working groups, and other mechanisms to refine objectives and designs and to ensure that new developments are incorporated into eventual facility technical capabilities and that scientific goals and functions take advantage of progress.

1.3.6.2.3. Community Involvement in GTL Technology Development

The GTL program and supporting facilities require extensive and continuing technology development. While GTL facilities will have a mission to keep capabilities fresh and relevant to emerging science and progress, a major element will be a distributed model for development of new methods and instruments that can be incorporated into the facilities. GTL will place key technology development where it can best take advantage of skills, capabilities, and infrastructure in universities, national laboratories, and industry.

1.3.6.2.4. Communication in a Multidisciplinary Environment

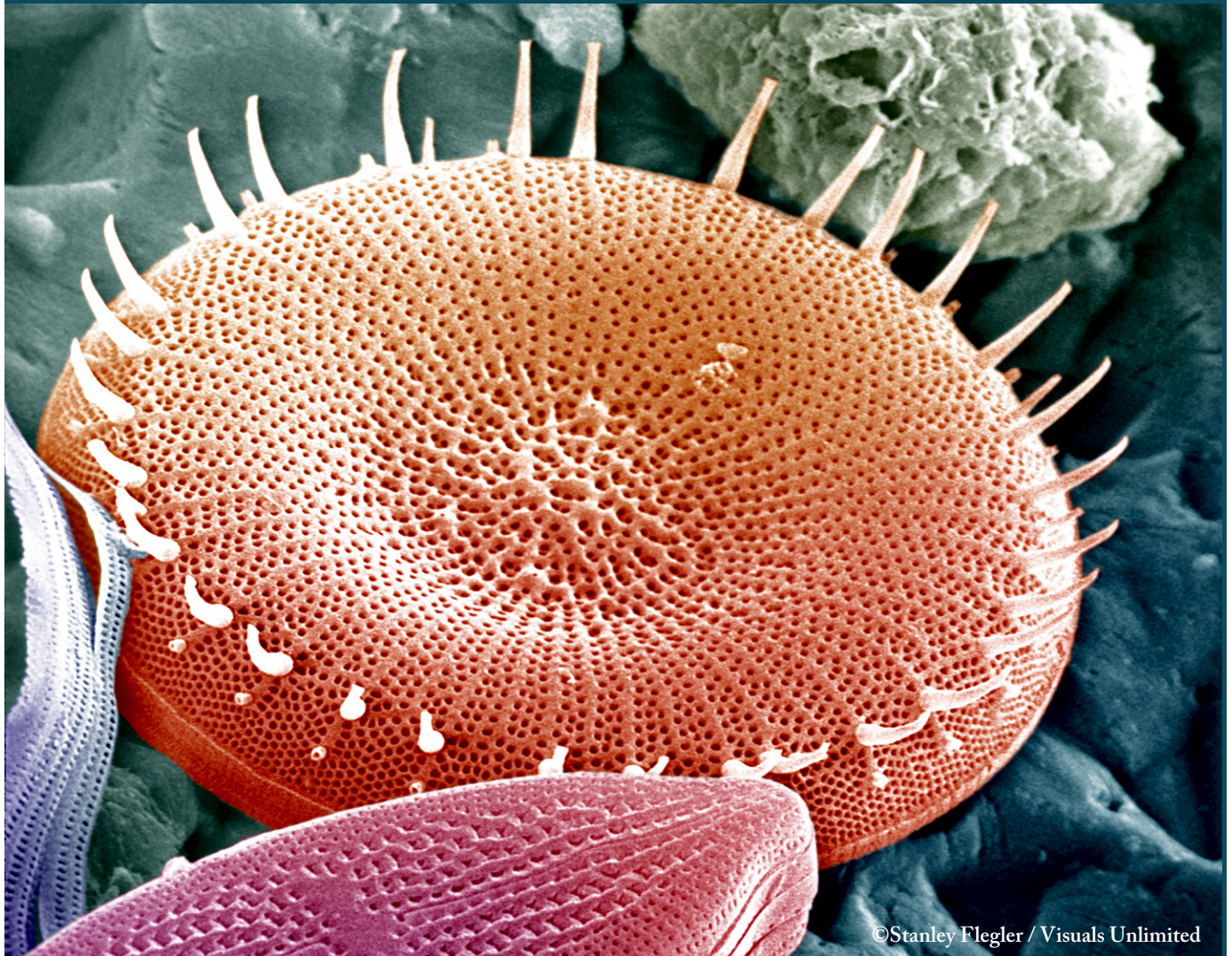
GTL communication strategies are to provide information to stimulate discourse and consensus, resulting in faster evolution of program design and content and more effective plans for achieving program aims. This communication will aid coordination and integration, particularly needed by a program requiring contributions from many disciplines, each with its own culture and vernacular. GTL also will foster cooperation between fundamental research and the development of technologies and applications. Communication will be prospective in conveying strategies and plans to generate new ideas. It will be retrospective in relating progress, results, and impacts. Effective communication, community involvement, and outreach will be prerequisites to establishing the GTL user facilities and a constituency to make best use of them. GTL will involve scientific societies and their ongoing meetings and other functions in these activities.

1.3.6.2.5. Facility User Access

GTL's dedicated user facilities will provide the broader scientific community with technologies, research resources, data and data-analysis tools, and computing and information infrastructure to perform systems microbiology studies (see 5.0. Facilities Overview, p. 101).

Access to GTL user facilities will be based on a peer-review process that will judge science quality and relevance and the need to use these valuable national assets. Factors in judging proposals will include science inventiveness, relevance to solving energy- and environmental-mission problems, quality and breadth of interdisciplinary teams, institutional capabilities to execute the science, performance records of investigators, and quality of the plan to use facility outputs. This formula allows for the study not only of systems with direct relevance to DOE missions but also of model systems that could shed light on DOE missions.

The Microbial World: A Challenging Frontier



The physical diversity of microbes reflects a commensurate underlying genetic and functional diversity—yielding a broad range of biochemical capabilities that sustain the planet. Diatoms (pictured above) are photosynthetic microorganisms that play a role in global carbon cycling and sequestration. Famous for their wide variety of intricately shaped silica walls, these organisms are abundant in plankton and in marine and freshwater sediments, often being found in fossil deposits.

THE MICROBIAL WORLD:

A Vast and Genetically Rich Resource

Microbes and their communities make up the foundation of the biosphere and sustain all life on earth. These single-celled organisms are masters at living in almost every environment and harvesting energy in almost any form, from solar radiation to photosynthesis-generated organic chemicals to minerals in the deep subsurface.

Microbes have evolved over 3.5 billion years, transforming the atmosphere with oxygen (a by-product of photosynthesis) more than a billion years ago to create the environment for life as we know it. Some microbes can thrive in either aerobic (with oxygen) or anaerobic (without oxygen) conditions. Microbes also capture nitrogen from the atmosphere, make it available to plants (and other life forms), and carry out processes responsible for soil fertility. Most do not cause disease. The unique microbial biochemistries amassed over eons in every niche on the planet now offer a deep and virtually limitless resource of capabilities that can be applied to national needs, including DOE energy and environmental missions.

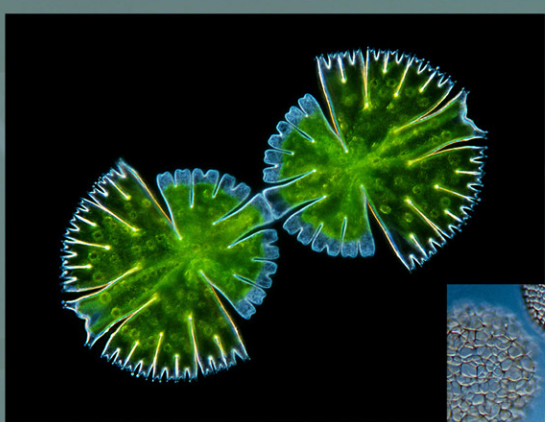
Although immense, the microbial world remains largely unexplored, a frontier of truly astronomical dimensions: The estimated nonillion or 10^{30} individual bacteria on earth are 10^9 times more than the number of stars in the universe. The vast majority, however, cannot be studied using standard techniques. While 2000 to 3000 species are estimated to be present in a single gram of soil, we can cultivate for study only some 0.1 to 1% of the species in that or any other environment. About 5700 species have been described thus far.¹⁻³

Investigators now are beginning to apply the tools of genomics to studying this enormous untapped natural treasure. Because microbes have modest-sized genomes (averaging 4 to 5 million bases compared with 3 billion bases in the human and other mammalian genomes), they represent a tractable life form we can use to explore and understand life processes at a whole-system level. Already, limited environmental sampling of microbes and their communities has led to the

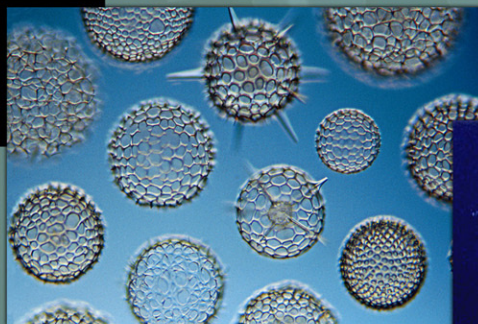
discovery of millions of previously unknown genes and proteins, thousands of species, and innumerable variations in critical functionalities. As scientists begin to scratch the surface of the microbial world, they are finding analysis an enormous challenge.

Recent discoveries from projects funded by DOE's Biological and Environmental Research program highlight the ubiquitous presence and critical importance of microbes in all ecosystems. For example:

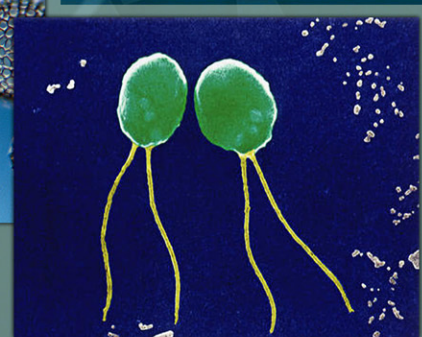
- The cyanobacteria *Prochlorococcus* and *Synechococcus*, along with other ocean phytoplankton, account for about half of global photosynthesis.⁴



***Micrasterias rotata*, a Desmid Undergoing Cell Division or Cytokinesis.** Desmids are attractive unicellular freshwater green algae that have a distinct basic symmetry consisting of two semicells with the nucleus situated in the narrow center. When they divide, two new semicells are formed.
[© Wim van Egmond / Visuals Unlimited]



***Spumellarian radiolarian*, Skeletons from the Ocean Bottom.** Radiolarians are unicellular protists with strikingly beautiful siliceous skeletons showing radial symmetry.
[© Wim van Egmond / Visuals Unlimited]



***Chlamydomonas*, Green Algae with Two Flagella for Movement.** These microbes can generate hydrogen from light, water, and basic nutrients.
[Elias Greenbaum, Oak Ridge National Laboratory]

- Diatoms, ancient and intricately shaped ocean microbes, store an amount of carbon comparable to that in all the earth's rainforests combined. Over geological time, diatoms may have influenced the earth's climate.⁵

- More than a million previously undiscovered genes, possibly representing new biochemical functions, were the surprising find in sequencing DNA fragments from the Sargasso Sea—a region heretofore thought to sustain little life.⁶ This discovery also was named one of *Science* magazine's "Breakthroughs of the Year."⁷

A CHALLENGING FRONTIER

- Microbes thrive deep within the earth's subsurface and at extremes previously thought to extinguish life.⁸

Growing recognition of microbial capabilities and potential applications has made a compelling case for further investigations by DOE and other agencies and institutions.

Before we can harness their capabilities, microbes must be understood in far greater detail and in the realistic context of whole living systems—whether as individuals or communities of interacting microbes—rather than as isolated components such as single genes and proteins. Microbes already can be manipulated at the molecular, cellular, and system levels, but understanding and taking advantage of their complexities and surmounting the technical challenges of whole-systems biology is a daunting prospect.

Understanding MICROBES and Their Communities

Most microbes live in highly organized and interactive communities that are versatile, complex, and difficult to analyze from many perspectives. Some of these challenges are outlined below.

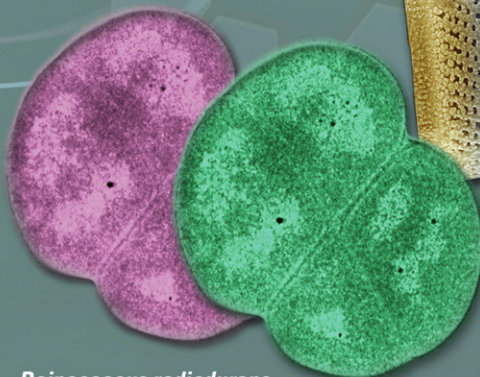
- Microbes are exceedingly small—only 1/8000th the volume of a human cell and spanning about 1/100th the diameter of a human hair. Investigating processes within this size range is challenging.
- The microbial world encompasses millions of genes from thousands of species, with hundreds of thousands of proteins and multimolecular machines operating in a web of hundreds of interacting processes in response to numerous physical and chemical environmental variables. Gene control is complex, with groups or “cassettes” of genes (operons) directing coordinated transcription and translation of genes into interacting proteins.
- Microbes adapt rapidly in response to environmental change, an ability that underlies their survival for billions of years. For example, various species of “extremophile” microbes have adapted to great extremes of pressure, temperature, pH, salinity, and radiation. Their high surface-to-volume ratio enhances interactions and supports adaptation. Unlike animal cells, they have no protective nucleus for their DNA, which leaves it more vulnerable to alteration. Genes move easily among species. Moreover, microbial communities are awash in genetic material from viruses that confer additional genetic properties and expand their range of adaptability.



Rod-Shaped (Bacilli) and Spherical (Cocci) Bacteria Found in Compost.

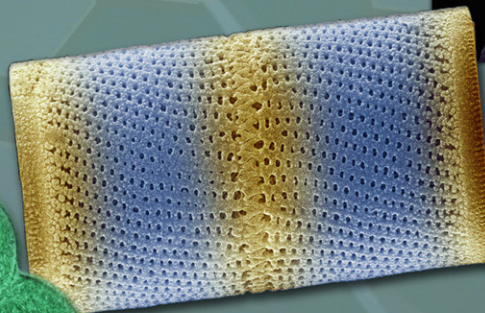
Decomposition of organic matter is an extremely important process in nature and a part of the global carbon and nutrient cycle.

[© Simko / Visuals Unlimited]



***Deinococcus radiodurans*, the Most Radiation-Resistant Microbe Known.**

[Michael Daly, Uniformed Services University of the Health Sciences]



Diatom, a Unicellular Algae.

The cell walls of diatoms are made of silica and come in a variety of shapes. These microscopic algae may be either fresh or saltwater, are photosynthetic, and play a role in carbon cycling.

[© Stanley Flegler / Visuals Unlimited]

other natural ecosystems. These “metagenomics” studies have given us a glimpse into the intricacies of these natural ecosystems and their diverse functions.

References noted on these pages are listed on the last page of this section.

- Microbial communities can extend in size from cubic millimeters (or smaller) to cubic kilometers. Even relatively simple communities can have millions of genes, giving them a genetic diversity substantially greater than that of higher life forms, even humans. Recent investigations have focused on collecting DNA fragments from environmental samples in the sea and

THE MICROBIAL WORLD

Microbes on the Move

Chemotaxis: Sensing and Moving in a Chemical Gradient

Motile bacteria use sophisticated information-processing devices to detect and respond to changes in their chemical environments. *Escherichia coli* cells, for example, use a signaling cascade of protein phosphorylation and dephosphorylation reactions to control the stiff flagellar filaments responsible for cell motility. The filaments, made up of several individual molecular motors, rotate to propel the cell in favorable directions (called chemotaxis) or allow it to tumble randomly. Cells make motility decisions by comparing current conditions to those occurring previously. Chemical changes as minute as 1 part per 1000 can be detected.

The methylation state of transmembrane chemoreceptors (methyl-accepting chemotaxis proteins) encodes the memory of its chemical environment and controls the flux of phosphates through a signaling cascade. Two signals are produced: A feed-forward signal that alters the motor rotation and a feedback signal that updates the methylation record. Motor responses occur in a few hundred milliseconds, whereas sensory-adaptation machinery updating the methylation record takes several seconds. New in vivo experimental approaches are needed to better understand the functional anatomy of bacterial receptor clusters.

Reference

J. S. Parkinson, "Signal Amplification in Bacterial Chemotaxis Through Receptor Teamwork," *ASM News* (2004).

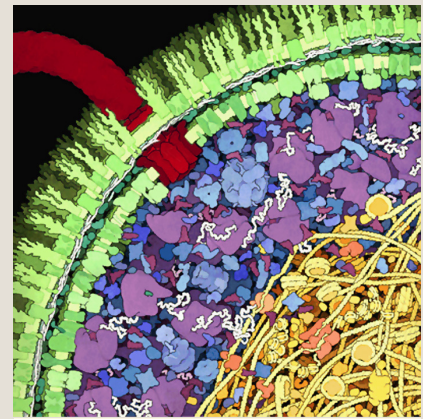


Illustration of *E. coli* Cell with Protruding Flagellum.

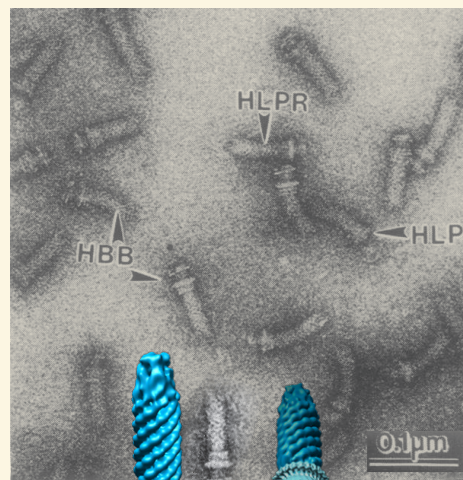
© 1999 D. Goodsell, Scripps Research Inst.

Built for Motility: A Complex Molecular Motor

The bacterial flagellum consists of multiple copies of at least 13 different proteins. This multiprotein machine contains an axial structure running through the length and a set of ring structures within the basal body embedded in the cell envelope. At right is a conventional electron microscopic image of uranyl acetate-stained flagellar complexes and subcomplexes prepared from *Salmonella typhimurium*. The hook-basal body (HBB in photo) contains a hook and rings of proteins; HL-PR refers to the hook and L and P rings, while HLP refers to the complex containing hook, L and P rings, and distal portion of the rod.

Reference

G. E. Sosinsky et al., "Mass Determination and Estimation of Subunit Stoichiometry of the Bacterial Hook-Basal Body Flagellar Complex of *Salmonella typhimurium* by Scanning Transmission Electron Microscopy." *Proc. Natl. Acad. Sci. USA* 89(11), 4801-5 (1992).



Flagellar Complexes from *S. typhimurium*.

Inset: Surface Representation of a 3D Map of the Bacterial Hook-Basal Body Flagellar Complex of *S. typhimurium*.

Inset illustration: D. Thomas, N. Francis, and D. DeRosier, Brandeis Univ. Photomicrograph: C. Anderson, Brookhaven National Laboratory

Growing Flagella in a Pinch

Some microbes grow flagella only when they need to find their way to nutrients essential to metabolic processes. Such is the case with *Geobacter metallireducens*, which produces energy for biochemical reactions by transferring electrons to metals. Although originally thought to be nonmotile, *G. metallireducens* genome analysis turned up genes encoding flagella. Further investigations showed that this microbe produces flagella when faced with insoluble sources of iron or manganese [Fe(III) or Mn(IV)]. Genes for pili (fine hair-like structures on the microbe's surface) also are present and expressed during growth on insoluble oxides; studies indicate their role as facilitating movement toward and aiding attachment to iron oxides (see sidebar, Bacteria Use "Nanowires" to Facilitate Extracellular Electron Transfer, p. 73). Additional genes for chemotaxis also were apparent in the genome, leading to the discovery of a novel mechanism for chemotaxis to iron. Understanding *Geobacter's* physiology is important for optimizing strategies to use this organism to bioremediate metals such as uranium in contaminated subsurface environments. Global gene-expression studies are helping to identify regulatory circuits, specifically those involved in bioremediation pathways and electricity production. [Source: Derek Lovley, University of Massachusetts, Amherst]

References

S. E. Childers, S. Ciuffo, and D. R. Lovley, "Geobacter metallireducens Accesses Insoluble Fe (III) Oxide by Chemotaxis," *Nature* 416, 767–69 (2002).

D. R. Lovley, "Cleaning Up with Genomics: Applying Molecular Biology to Bioremediation," *Natl. Rev. Microbiol.* 1, 35–44 (2003).

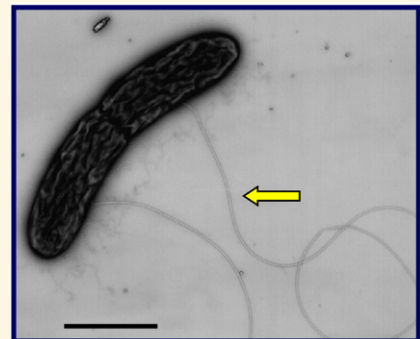
Siderophores: Sending out Shuttles to Scout for Iron

Another efficient mechanism evolved by many microbes to obtain iron in limited environments such as marine surfaces is the production and secretion of siderophores. These low-molecular-weight chelating agents act as shuttles to bind insoluble iron [Fe(III)] and transport it back to the microbe, where it enters the cell by recognizing specific membrane receptor proteins and transport systems. More than 500 types of siderophores are known to exist. In addition to supplying essential nutrients, siderophores of one organism can lock up iron to achieve an advantage over their competitors. However, some microbes were discovered recently to have receptors for the siderophores of other organisms.

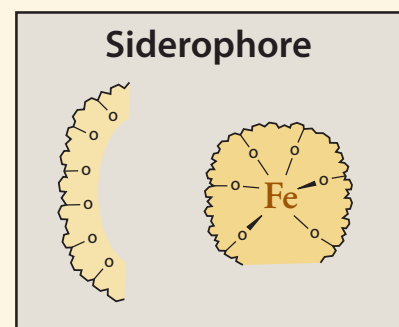
Until the genetic sequence of the ocean diatom *T. pseudonana* was determined in the DOE Microbial Genome Program and compared with sequences of other organisms, researchers were unaware that these organisms possessed siderophores. Diatoms, along with other ocean microbes, contribute to absorbing CO₂ in amounts comparable to that absorbed by all the world's tropical rain forests combined. Obtaining more detailed knowledge of their life processes will help us better understand their vital role in global carbon cycling.

Reference

E. V. Armbrust et al., "The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism," *Science* 306, 79–86 (2004).



Flagella (arrow) Produced by *Geobacter* in the Presence of Insoluble Sources of Iron or Manganese.



Schematic Representation of a Siderophore Before and After Iron Acquisition. [Adapted from H. Boukhalfa and A. L. Crumbliss, "Chemical Aspects of Siderophore-Mediated Iron Transport," *Biometals* 15, 25–39 (2002).]

THE MICROBIAL WORLD

Group Living and Communicating

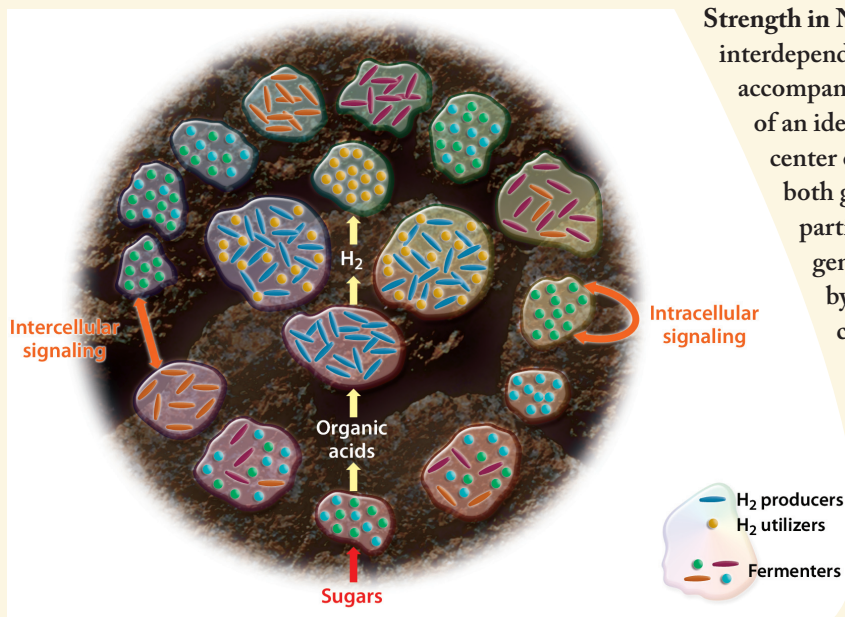
Life in a Biofilm

Most microbes live attached to solid surfaces (biotic or abiotic) within highly organized and functionally interactive communities called biofilms. These biofilms can be composed of populations that developed from a single species or a community derived from multiple species. All exhibit collective and interdependent behavior, with different genes rapidly brought into play as conditions dictate (see figure below). Among the many advantages of biofilm living are nutrient availability with metabolic cooperation, acquisition of new genetic traits, and protection from the environment.

Researchers are only beginning to realize the prevalence and significance of biofilms. These communities probably play major roles in such complex natural processes as the cycling of nitrogen and sulfur and the degradation of environmental pollutants and organic matter, activities that require a range of metabolic capabilities. Recent metagenomic and metaproteomic studies focused on biofilm members in an acid mine drainage environment (see sidebar, Metagenomics, p. 62).

Reference

M. E. Davey and G. A. O'Toole, "Microbial Biofilms: From Ecology to Molecular Genetics," *Microbiol. Mol. Biol.* 64(4), 847–67 (2000).



Strength in Numbers. Microbes in biofilms live an interdependent, community-based existence (see accompanying text above). In this overhead view of an idealized biofilm, four microcolonies in the center of the figure represent organisms that both generate and consume hydrogen. Two participate in syntrophism, in which hydrogen producers use organic acids generated by fermenting organisms that gain their carbon and energy by using various sugars. In addition to potential metabolic interactions, signaling molecules may aid in inter- and intraspecies communication. These genetic factors and environmental influences contribute to the biofilms' spatial organization. [Figure and caption adapted from Davey and O'Toole, 2000.]

Quorum Sensing

Microbes communicate with each other by sending and detecting a wide variety of chemical signals (autoinducers). These molecules trigger group behaviors, including the formation and persistence of biofilms, symbiosis, and other processes. Many of these activities are density dependent, that is, when a threshold concentration of chemicals is detected (reflecting a certain number of cells), microbes respond with a change in gene expression. This process, called quorum sensing, facilitates coordination of gene expression by the entire community, in essence enabling it to behave like a multicellular organism. Quorum sensing allows microbial communities to adapt rapidly to environmental changes and reap benefits that would be unattainable as individuals.

Quorum sensing first was described in the bioluminescent marine bacterium *Vibrio fischeri*. This microbe lives in symbiotic association with several marine animal hosts, who use the light it produces to attract prey, avoid predators, or find a mate. In exchange, *V. fischeri* obtains a nutrient-rich home environment. *V. fischeri* emits light only inside a specialized light organ of the host, where the concentration of these organisms becomes dense; it does not give off light when free living in the ocean. Light production depends on producing, accumulating, and responding to a minimum-threshold concentration of an autoinducer (acylated homoserine lactone). Only under the nutrient-rich conditions of the light organ can *V. fischeri* grow to such high populations. Also, trapping the diffusible autoinducer molecule in the light organ with the bacterial cells allows it to accumulate to a concentration sufficient for *V. fischeri* to detect it.

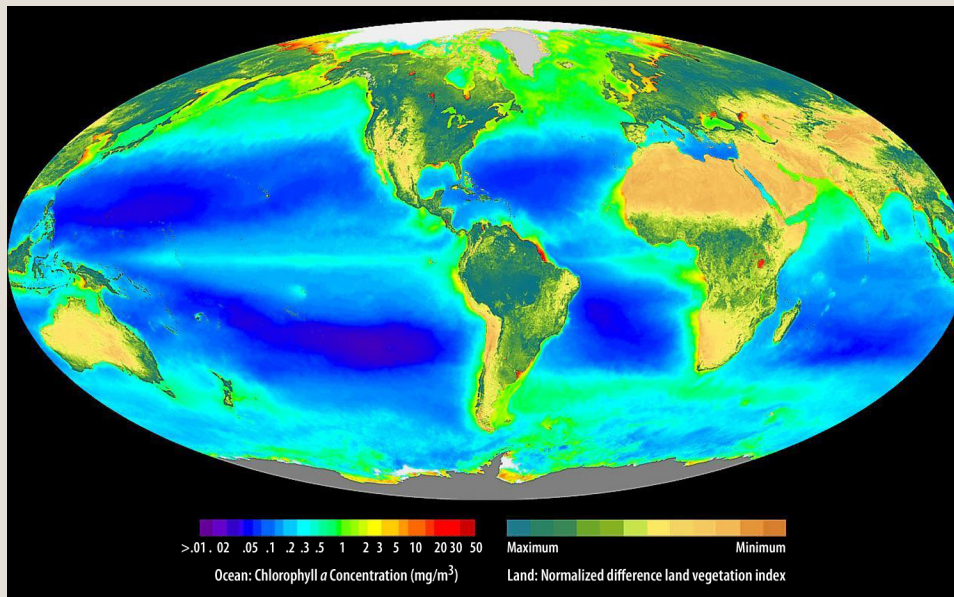
Recent studies have revealed diverse chemical languages that enable bacterial communication both within and between species (the latter called cross talk). The extracellular matrix surrounding mature biofilms (composed of glycans and other components) plays a crucial role in transmitting these chemical signals into and between cells. Biotechnological researchers are developing molecules structurally related to autoinducers to exploit quorum-sensing capabilities and possibly improve industrial production of natural products.

References

- J. W. Hastings and K. H. Nealson, "Bacterial Bioluminescence," *Annu. Rev. Microbiol.* 31, 549–95 (1977).
 S. Schauder and B. L. Bassler, "The Language of Bacteria," *Genes Devel.* 15, 1468–80 (2001).

Photosynthetic Microbes—Major Contributors to Earth’s Life-Support System

When photosynthetic microbes, fungi, and plants convert light energy from the sun into glucose, they establish the foundation for the food chain on which all life, including human, depends. This NASA SeaWiFS image of the global biosphere shows the density of photosynthetic organisms on land and in the oceans. On land, the dark greens represent areas of abundant vegetation, with tans showing relatively sparse plant cover. In the oceans, red, yellow, and green pixels depict dense blooms of phytoplankton (photosynthetic microbes), while blues and purples show regions of lower productivity.



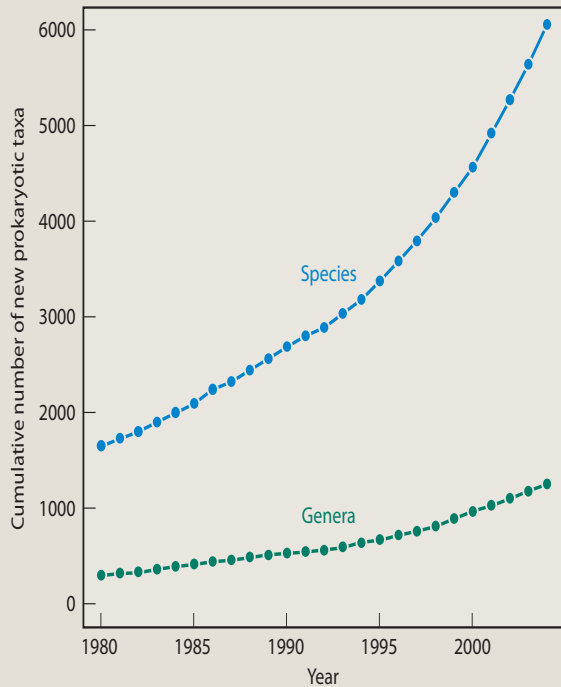
blooms of phytoplankton (photosynthetic microbes), while blues and purples show regions of lower productivity.

Ocean’s Long-Term Average Phytoplankton Chlorophyll Concentration. Image combines September 1997 through August 2000 concentration data with the SeaWiFS-derived Normalized Difference Vegetation Index over land; <http://oceancolor.gsfc.nasa.gov/SeaWiFS/>.

THE MICROBIAL WORLD

Cataloging Microbial Diversity

What's in a Name? The Challenges of Tracking Microbial Species



Prokaryotic systematics is a dynamic field. The rate at which new species, genera, and higher taxa are described in the literature has increased dramatically since the 1990s (see figure at left), driven largely by advances in sequencing technology. Even with all this progress, scientists believe that 99% of the microbial world has yet to be discovered.

Just how many species (and genera) of *Bacteria* and *Archaea* are listed in *Bergey's Manual*, a widely used international reference for taxonomy? It seems a simple question, but, at present, the number of named species actually exceeds the true number of species having official standing in the nomenclatural record by about 22%. An explanation follows.

Many taxa bear two or more names, because when species or higher taxa are reassigned to existing or newly created taxa, both the new name and the old name are valid (in the original published context). Other types of nomenclatural synonyms exist as well. As a result, about 6900 species are listed in *Bergey's*, with information and data published in accordance with *The International Code of Bacterial Nomenclature*. The true number of named prokaryotes, however, is closer to 5700. [Source: George Garrity, *Bergey's Manual*]

References for pp. 14–15

1. T. P. Curtis and W. T. Sloan, "Prokaryotic Diversity and Its Limits: Microbial Community Structure in Nature and Implications for Microbial Ecology," *Curr. Opin. Microbiol.* 7, 221–26 (2004).
2. V. Torsvik and L. Øvreås, "Microbial Diversity and Function in Soil: From Genes to Ecosystems," *Curr. Opin. Microbiol.* 5(3), 240–45 (2002).
3. *Bergey's Manual*, 2nd edition (2001).
4. T. S. Bibby et al., "Low-Light-Adapted *Prochlorococcus* Species Possess Specific Antennae for each Photosystem," *Nature* 424, 1051–4 (2003).
5. E. V. Armbrust et al., "The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism," *Science* 306, 79–86 (2004).
6. J. C. Venter et al., "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* 304, 66–74 (2004).
7. The News Staff, "Breakthrough of the Year: The Runners Up," *Science* 306, 2013–17 (2004).
8. T. C. Onstott et al., "Indigenous and Contaminant Microbes in Ultradeep Mines," *Environ. Microb.* 5(11), 1168–91 (2003); K. Kashefi and D. R. Lovley, "Extending the Upper Temperature for Life," *Science* 301, 934 (2003).