5. GTL Facilities

5.0. Facilities Overview 101

5.0.1. Science and Technology Rationale 102
5.0.2. A New Trajectory for Biology 104
5.0.3. Capsule Facility Descriptions 104
5.0.3.1. Facility for Production and Characterization of Proteins and Molecular Tags 104
5.0.3.2. Facility for Characterization and Imaging of Molecular Machines 105
5.0.3.3. Facility for Whole Proteome Analysis 105
5.0.3.4. Facility for Modeling and Analysis of Cellular Systems 106
5.0.4. Relationships and Interdependencies of Facilities 106
5.0.5. Research Scenarios 107
5.0.6. Facility Development 107

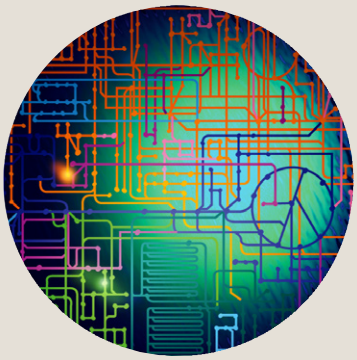
5.1. Facility for Production and Characterization of Proteins and Molecular Tags 111

5.2. Facility for Characterization and Imaging of Molecular Machines 139

5.3. Facility for Whole Proteome Analysis 155

5.4. Facility for Analysis and Modeling of Cellular Systems 173

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, the Bioenergy Research Center, will concentrate on overcoming biological barriers to the industrial production of biofuels from biomass and on other potential energy sources such as biophotolytic decomposition of water to hydrogen. 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://doegenomestolife.org/biofuels/).



To address the analytical and computational capabilities needed to put the GTL research program on track for creating a science foundation for DOE missions, workshops were held between June 2002 and June 2004. Much of the material in this chapter was drawn from the outputs of those workshops, in which nearly 800 different individuals participated. For a list of GTL workshops, meetings, and links to workshop reports, see Appendix D. GTL Meetings, Workshops, and Participating Institutions, p. 239.

Facilities Overview

The proposed GTL user facilities for 21st Century biology and biotechnology will be a major strategic asset in achieving DOE mission goals in industrial biotechnology—a critical arena of national economic competitiveness. The facilities will enable a new era in biology, building on the national investment in genomics.

The research community increasingly is recognizing the need for global analysis of myriad simultaneous cellular activities and is calling for a new research infrastructure. "Progress in microbiology always has been enabled by the technology available, a fact that is still true today. However, many researchers are stymied by lack of access to the expensive instruments that would enable them to make the greatest strides." (Schaechter, Kolter, and Buckley 2004, p. 13; see also Aebersold and Watts 2002; Buckley 2004a; Stahl and Tiedje 2002).

5.0.1. Science and Technology Rationale

In the simplest interpretation, an organism's genome contains encoded information to produce the proteins that are the cell's workhorses. Central to the strategy of GTL and the GTL facilities suite, however, is the fact that information encoded in the genomes of microbes and the metagenomes of microbial communities goes well beyond encoding proteins. Genomes control microbial structure and function at many spatial and temporal scales—molecular, machine, cellular, community, and ecosystem—through an intricate set of interrelated and communicating regulatory and control processes (see Fig. 1. DOE Genomics:GTL High-Throughput User Facilities, p. 103). Microbes display such strong interactions that capabilities are needed to explore these systems in a comprehensive and integrated way at all levels. Proteins and even microbes are thought to function rarely in isolation.

In the Missions Overview, our example mission descriptions demonstrate that the technical challenges of these analyses and the scale of the systems that must be understood exceed any existing capabilities (see Missions Overview, Tables 1–9, beginning on p. 26; 3.2.2. Science and Technology Milestones, p. 44; and sidebar, High-Throughput Model Guides Future Facilities, p. 6). Facilities must be established to dramatically improve research performance, throughput, quality, and cost.

Examples of performance challenges include producing and characterizing complex proteins (e.g., membrane and multidomain); isolating,

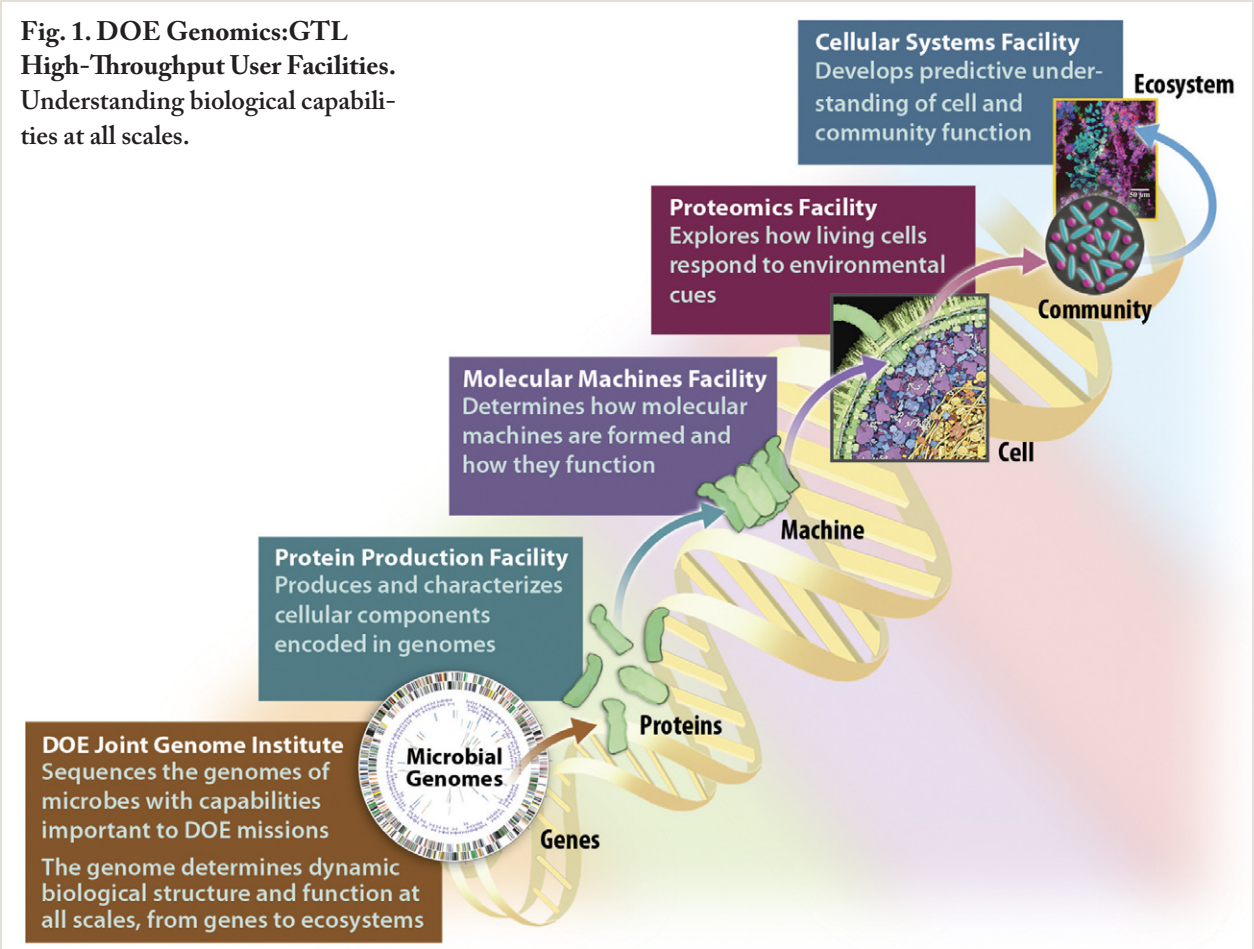


Fig. 1. DOE Genomics:GTL High-Throughput User Facilities. Understanding biological capabilities at all scales.

characterizing, and modeling large or tenuous molecular machines; measuring the full molecular profile of microbial systems; and imaging molecules as they carry out their critical functions in cells in structured communities. Examples of throughput challenges include providing insight into the functions of hundreds of thousands of unknown genes and their modifications; processing thousands of molecular machines; analyzing molecular profiles of thousands of microbial samples under different conditions; and spanning the full range of conditions and processes governing microbial-community behaviors. Quality control includes developing and implementing strict protocols and providing the most sophisticated diagnostics. High-throughput methods and resource sharing among community members will lower the unit cost for production and analyses.

Figure 1, this page, depicts facilities focused on building an integrated body of knowledge about behavior, from genomic interactions through ecosystem changes. Simultaneously studying multiple microbial systems related to various mission problems is powerfully synergistic because enduring biological themes are shared and general principles governing response, structure, and function apply throughout. The biology underlying the challenges of one mission will inform those of the others. Accumulating the data as it is produced, the GTL Knowledgebase and the computational environment that GTL will create will act as the central nervous system of the facilities and program, allowing this information to be integrated into a predictive understanding.

The Office of Science has a tradition of strategic basic research in a multidisciplinary team environment for national missions. These facilities will bring together the biological, physical, computational, and engineering sciences to create a new infrastructure for biology and the industrial biotechnology needed for the 21st Century. DOE's technology programs can work with industry to apply such capabilities and knowledge to a new generation of processes, products, and industries.

5.0.2. A New Trajectory for Biology

As we have learned from the genome projects, consolidating capabilities and focusing on aggressive goals will drive dramatic improvements in performance and cost (Fig. 2. Putting Biology on a New Trajectory, this page). As depicted in Fig. 2, GTL facilities will accelerate discovery and reduce the time for useful applications. With this higher level of performance, microbial systems biology is tractable and affordable to support the next generation of industrial biotechnology for the coming decade and beyond.

5.0.3. Capsule Facility Descriptions

The GTL facilities provide a complementary set of technologies and products. Two facilities are focused on analysis of properties and functions of cellular components, proteins, and molecular machines:

- Facility for Production and Characterization of Proteins and Molecular Tags
- Facility for Characterization and Imaging of Molecular Machines

Two are focused on analysis of microbial-system responses and functions at the molecular, cellular, and community levels:

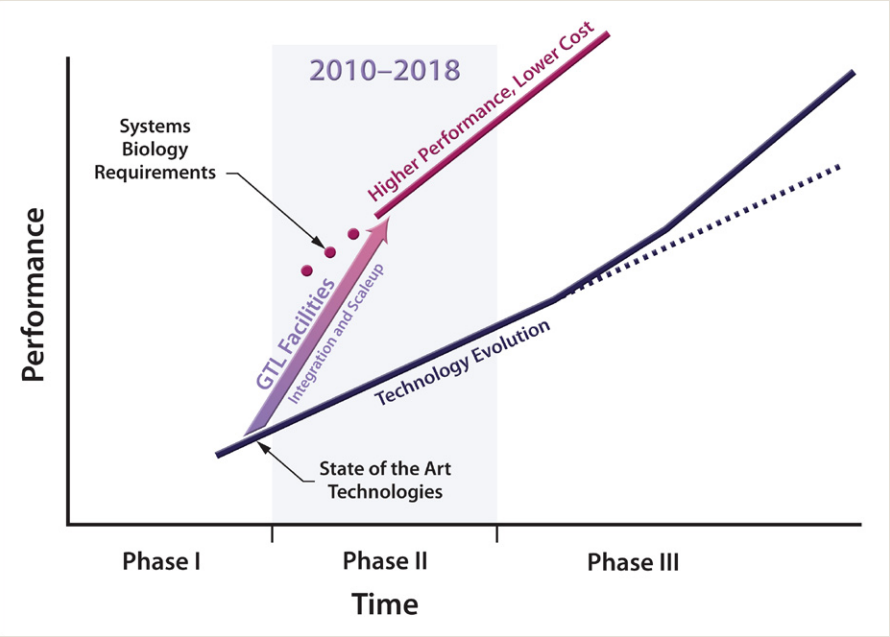
- Facility for Whole Proteome Analysis
- Facility for Modeling and Analysis of Cells and Communities

The ensuing chapters of this roadmap discuss each facility in further detail. Capsule facility descriptions follow.

5.0.3.1. Facility for Production and Characterization of Proteins and Molecular Tags

The Protein Production and Characterization Facility will use DNA sequence to make proteins and reagents for interrogating cell function. Specifically, this facility will have the capability to produce all proteins encoded in any genome on demand; create molecular tags that allow each protein to be identified, located, and manipulated in living cells; and, to gain insights into function, perform biophysical and biochemical characterizations of proteins produced. Using high-throughput in vitro and in vivo techniques will lower the

Fig. 2. Putting Biology on a New Trajectory.



cost of producing proteins to levels that will allow comprehensive analysis of all proteins within the cell. The facility's products and analysis capabilities will be made available to all scientists.

In parallel with protein production will be the generation of "affinity reagents." These small proteins or nucleic acids will permit the detection and tracking of individual proteins in living systems, including complex molecular assemblages; the intracellular position of all proteins and their spatial dynamics; if exported, the extracellular localization and interaction with other community members; and techniques for manipulating protein activity in the environment.

Core facility instrumentation:

- Gene synthesis and manipulation techniques
- High-throughput microtechnologies for protein-production screening
- Robotic systems for protein and affinity-reagent production and characterization
- Computing for data capture and management, genomic comparative analyses, control of high-throughput system and robotics, and production-strategy determination

5.0.3.2. Facility for Characterization and Imaging of Molecular Machines

The Molecular Machines Facility will identify and characterize molecular assemblies and interaction networks. It will have capabilities to isolate and analyze molecular machines from microbial cells; image and localize molecular machines in cells; and generate dynamic models and simulations of the structure, function, assembly, and disassembly of these complexes. The facility will identify molecular machine components, characterize their interactions, validate their occurrence and determine their locations within the cell, and allow researchers to analyze the thousands of molecular machines that perform essential functions inside a cell. It will provide a key step in determining how the network of cellular molecular processes works on a whole-systems basis by completely understanding individual molecular machines, how each machine is assembled in 3D, and how it is positioned in the cell with respect to other components of cellular architecture.

Core facility instrumentation:

- Robotic culturing technologies to induce target molecular machines in microbial systems and supporting robotic techniques for molecular complex isolation
- Numerous sophisticated mass-spectroscopy and other techniques specially configured to analyze samples of purified molecular machines for identification and characterization of complexes
- Various advanced microscopies for intracomplex imaging and structure determination
- Imaging techniques for intracellular and intercellular localization of molecular complexes
- Computing and information systems for modeling and simulation of molecular interactions that lead to complex structure and function

5.0.3.3. Facility for Whole Proteome Analysis

The Proteomics Facility will be capable of gaining insight into microbial functions by examining samples to identify (1) all proteins and other molecules that a microbe (or microbial community) creates under controlled conditions and (2) key pathways and other processes. An organism selectively produces portions of its proteome in response to specific environmental or intracellular cues. Studying its constantly changing protein expression thus leads to a better understanding of how and why an organism turns portions of its genome "on" and "off." Facility users will achieve a comprehensive understanding of microbial responses to environmental cues by identifying, quantifying, and measuring changes in the global collections of proteins, RNA, metabolites, and other biologically significant molecules. These molecules, including lipids, carbohydrates, and enzyme cofactors, are important in understanding biological processes mediated by proteins. Integrating diverse global

data sets, the facility will develop computational models to predict microbial functions and responses, inferring the nature and makeup of metabolic and regulatory processes and structures.

Core facility instrumentation:

- Large farms of chemostats to prepare samples from highly monitored and controlled microbial systems under a wide variety of conditions
- Numerous specialized mass and NMR spectrometers and other instrumentation capable of analyzing the molecular makeup of ensemble samples with thousands of diverse molecular species
- High-performance computing and information capabilities for modeling and simulation experiments of microbial-system functionalities under different scenarios to inform the design of experimental campaigns focused on systems-level goals and to infer microbial-system molecular processes from ensuing data

5.0.3.4. Facility for Modeling and Analysis of Cellular Systems

The Cellular Systems Facility will be the capstone for the ultimate analytical capabilities and knowledge synthesis to enable a predictive understanding of cell and community function critical for systems biology. The facility will concentrate on the systems-level study of living cells in complex and dynamic structured communities. Imaging methods will monitor proteins, machines, and other molecules spatially and temporally as they perform their critical functions in living cells and communities. Microbial communities contain numerous microniches within their structures that elicit unique phenotypic and physiological responses from individual species of microbes. We need to be able to analyze these niches and the microbial inhabitants within. This grand challenge for biology must be addressed before scientists can predict the behavior of microbes and take advantage of their functional capabilities. Modeling in the facility will describe essential features of these biological interactions with the physicochemical environment and predict how the system will evolve in structure and function.

Core facility instrumentation:

- Highly instrumented cultivation technologies to prepare structured microbial communities to simulate natural conditions under highly controlled conditions
- Instruments integrating numerous analytical imaging techniques that can spatially and temporally determine, in a nondestructive way, the relevant molecular makeup and dynamics of the community environment, community, and microbes that comprise it
- Computing and information capabilities to model and simulate complex microbial systems, design experiments, and incorporate data

5.0.4. Relationships and Interdependencies of Facilities

Each of the facilities is technically distinct in the nature of its instrumentation, methods, and overall goals. All will be centered around either production lines designed to maximize quality and throughput and reduce unit costs, the development and operation of frontier instrumentation or unique suites of instrumentation to reach new levels of performance, or combinations of both. While each can serve a user community for a wide range of independent studies, the suite of facilities has complementary strengths and core technologies that together can help provide complete systems knowledge. Figure 3. GTL Facilities: Core Functions and Key Interactions, p. 108, displays how each facility's core functions are complementary to those of the other facilities. The key interactions shown demonstrate their interdependencies and necessary exchange of all information through the GTL Knowledgebase and the program's communication and computing infrastructure.

5.0.5. Research Scenarios

As described in the Missions Overview and related appendices, each mission example has a unique endpoint and research strategy for developing the needed understanding, predictive models, and research capabilities. Table 1. Research Scenarios on Microbial Processes, p. 109, and Table 2. Science Roadmaps for Natural Systems, p. 110, present conceptual research-scenario roadmaps for six cases as illustrations related to Science Milestones and GTL Facilities. Although these systems and problems cover a breadth of microbial phenomenology and system behaviors, they can be studied using the same foundational capabilities. Each of the GTL milestones, as denoted in the left column of Tables 1 and 2, drives the technical core of the facilities, where capabilities resulting from milestone R&D can be scaled up and integrated.

5.0.6. Facility Development

The facility acquisition process will employ project-management practices similar to DOE Order 413.3 Facilities Project Management. The facilities budget will include all costs for the conceptualization, design, R&D and testing, and acquisition of the necessary conventional facilities, instrumentation, computers and software, and supporting technologies, training, and installation of fully operational production lines and analytical facilities upon completion of the project. The process will involve participants from national laboratories, academia, and industry in the necessary workshops and working groups to determine the technical scope and scale of the facilities, technical priorities, and technology development. Many of the long-lead and crosscutting development needs are outlined in the GTL Development Summary chapter. This roadmap is meant to be a starting point for the intensive conceptualization and planning that must occur for successful design, acquisition, and operation of these facilities.

Fig. 3. GTL Facilities: Core Functions and Key Interactions.

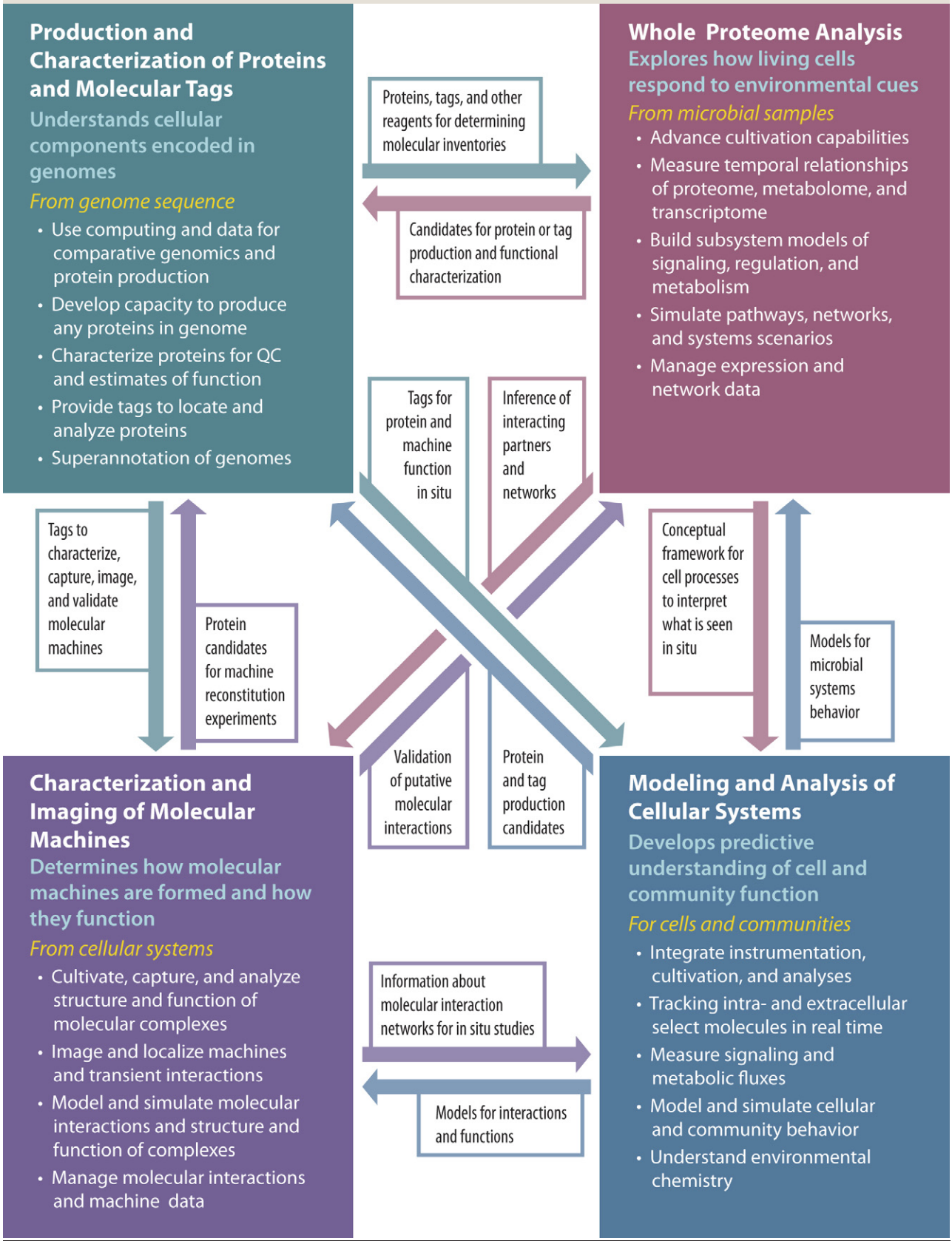


Table 1. Research Scenarios on Microbial Processes: Relationship to Science Milestones and Facilities

Progression of GTL Science Milestones and Facilities	*Conceptual Science Roadmaps for Microbial Energy and Environmental Processes*		
	Convert Sunlight to Hydrogen and High-Hydrogen Fuels	**Convert Cellulose to Fuels**	**Reduce Toxic Metals in Subsurface Environments**
Milestone 1: Determine the Genome Structure and Potential of Microbes and Microbial Communities Facility for the Production and Characterization of Proteins and Molecular Tags Facility for Characterization and Imaging of Molecular Machines	Analysis of hydrogenase families across microbial species: Screen nature for new variants Range of hydrogenase properties Suite of heterologous expression hosts Characterization of partners, energetics, structures, post-translational modifications Wide range of mutations, variations created and screened Functional and structural analysis of machines	Wide range of microbes surveyed for cellulases, ligninases, and other glucosyl hydrolases Partners and structural information established Structure and imaging of interactions important to efficient function	Survey of subsurface species and genomic potential Comparative genomics and superannotation Generation of knockouts, mutations, transmembrane structures to understand function
Milestone 2: Develop a Systems-Level Understanding of Microbial and Community Function and Regulation Facility for Whole Proteome Analysis Facility for Analysis and Modeling of Cellular Systems	Oxygen sensitivity of hydrogenases Electron-transfer reactions and limitations Reverse-reaction mitigation Partitioning of electrons between hydrogenases and competing pathways Light capture Biophotoelectric antenna	Proteome analysis of expression and regulation Fundamental mechanisms of cellulose deconstruction Transport of sugars Measurement of electron transport chains' redox state, control of electron fluxes Carbon partitioning in cells: Carbon, NAD, NADPH, ATP, ADP	Cellular response to environmental stimuli Proteomics, transcriptomics, and metabolomics to elucidate regulation and responses Intra- and intercellular communications Cells in structures such as biofilms Growth processes, toxicity responses, energy transfer, metabolic responses Microbe-mineral interactions
Milestone 3: Develop the Knowledgebase, Computational Methods, and Capabilities to Advance Understanding of Complex Biological Systems and Predict Their Behavior GTL Integrated Computational Environment for Biology	Pathway models—energetics, electropotential, docking, proton fluxes, cofactors Computational tools for rational design Suites of hosts, pathway cassettes Modeling and measurement of pathways, fluxes, regulation	Design of organisms capable of utilizing all sugars Optimization of sugar transport, regulation Redesign of cellulose structure pH- and temperature-tolerant microbes Principles for enzyme redesign	Modeling capable of visualizing realistic biochemical pathways in cells Interactions of membrane proteins with contaminants and solid-phase electron acceptors Design of experiments in cultured and natural systems
Missions Outputs Systems Engineering	In vivo systems Processes captured in nanostructures, biomimetic systems System design: Light harvesting, conversion to hydrogen or fuel, robustness to oxygen, regulation Transgenic approaches	Improved cellulases and production methods to reduce costs, improve stability Modularized processing to reduce transportation of feedstock Sensors for biomarkers and chemical intermediates	Assessment of long-term cellular and system behavior Remediation strategies Sensors for coupled biochemical and geochemical measurements in situ

Table 2. Science Roadmaps for Natural Systems: Relationship to Science Milestones and Facilities

Progression of GTL Science Milestones and Facilities	*Conceptual Science Roadmaps for Natural Systems*		
	Oceans: Photosynthetically Driven Biological Pumps for Carbon and Energy in Aquatic Systems	**Terrestrial: Microbes in Ecological Communities, Carbon and Nutrient Cycles**	**Deep Subsurface: Microbial Community Processes for Mitigation of Toxic Chemicals and Metals**
Milestone 1: Determine the Genome Structure and Potential of Microbes and Microbial Communities Facility for the Production and Characterization of Proteins and Molecular Tags Facility for Characterization and Imaging of Molecular Machines	Single-cell and environmental community sequence Heterotrophs, autotrophs, viruses, and "twilight zone" organisms Comparative analyses of rhodopsin, hydrogenase genomes Gene synthesis and manipulation	Single-cell and community sequence in situ and in vitro Organisms related to processes in soils Genome annotation	Single-cell and community sequence in situ and in vitro to identify members, functions Superannotation, genome plasticity effects Metagenomics, gene transfer Tags to ID microbes, proteins, metabolites
Milestone 2: Develop a Systems-Level Understanding of Microbial and Community Function and Regulation Facility for Whole Proteome Analysis Facility for Analysis and Modeling of Cellular Systems	Photosynthesis, transporters, biomineralization Proteins, machines, metabolites, and functional assays Systems responses Imaging	All GHGs: CO_2, methane, nitrous oxide, dimethyl sulfide Molecular inventories vs cues Systems interactions with soil, rhizosphere, plants: Inputs and outputs (e.g., stable isotope probes) Proteome and metabolome imaging at cellular and community levels	Community structure and relationship to function Pathways and networks: Mechanisms of intercellular communication and function Stoichiometry and kinetics of intercellular fluxes
Milestone 3: Develop the Knowledgebase, Computational Methods, and Capabilities to Advance Understanding of Complex Biological Systems and Predict Their Behavior GTL integrated computational environment for biology	Modeling of climate-based and mitigational perturbations Individual and multiple life-scale models (cellular, community, ecosystem): Metabolic budgets Multiple photosynthetic processes	Modeling of microbial responses to manipulation of plant inputs into carbon cycle Human inputs directed to soils Response to environmental change understood	Four-dimensional reactive transport models based on genomic, geochemical, and hydrological data Scaling of processes through molecular, cellular, community, and environmental levels; and molecular to long time scales
Missions Outputs Measure environmental responses via sensors	Ecogenomics of sentinel organisms Cellular, community, and ecosystem biochemical assays Accompanying environmental assays	Biomarkers: RNAs, proteins, metabolites, signaling Ecogenomics, functional assays, environmental conditions Carbon and nutrient inventories	Biology and geochemistry: DNA, RNA, proteins, metabolites, geochemical from single-cell to field scales Mesoscale simulation of field conditions Regulatory levels of contaminants
Robust Science Base for Policy and Engineering	Natural behaviors of ocean ecosystems, impact on and of climate change scenarios incorporated into IA models Assessment of efficacy and impacts of intervention strategies	Biological processes for carbon and nitrogen cycling, impact on and of climate change scenarios incorporated into IA models Assessment of potential and strategy for terrestrial carbon sequestration	Predictions of transport and fate Assessment of need for remediation Remediation strategies, designs, and tests