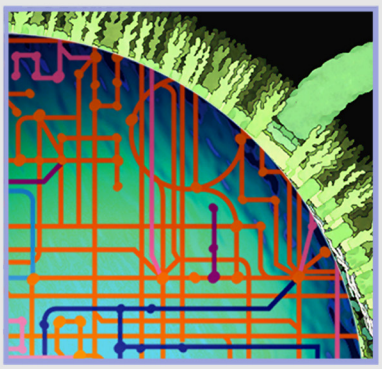
5.4. Facility for Analysis and Modeling of Cellular Systems

5.4.1. Scientific and Technological Rationale 174
5.4.1.1. Probing Mixed Microbial Populations and Communities 175
5.4.1.2. Foundations for Community Analyses 177
5.4.2. Facility Description 178
5.4.2.1. Laboratories, Instrumentation, Quality Control, Computing, and Support 178
5.4.2.2. Performance and Production Targets 178
5.4.3. Technology Development for Cultivation of Microbial Communities 179
5.4.3.1. Requirement Examples 180
5.4.4. Development of Genomic Capabilities 181
5.4.5. Technology Development for Imaging and Spectroscopy 181
5.4.5.1. Analytical Characterization of Cellular Systems 181
5.4.5.1.1. Examples of Analytical Requirements 184
5.4.5.1.2. Monitoring and Interacting with Cellular Systems 185
5.4.5.1.3. Technology Development Progress and Benefits 185
5.4.5.2. Imaging Macromolecular Complexes 186
5.4.5.3. Development Options 187
5.4.6. Development of Computing Capabilities 187

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/).

Facility for Analysis and Modeling of Cellular Systems



Achieve an in silico, predictive understanding of microbes in their natural environments.

Cellular Systems

- 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.

The Facility for the Analysis and Modeling of Cellular Systems will be a user facility to provide scientists with insight into the responses and functionality of microbes and microbial communities in complex environments. Modeling and real-time functional mapping of processes from the molecular through the ecosystem levels will be used.

5.4.1. Scientific and Technological Rationale

The Facility for Analysis and Modeling of Cellular Systems will be the GTL capstone needed to provide the ultimate integration of analytical capabilities and knowledge synthesis critical for systems biology. Users of this facility will investigate how microbial communities and their subsystems of cells function together to sense, respond to, and modify their environment. They will accomplish this by dynamically identifying, localizing, and quantifying molecular machines and all other important biomolecules and their interactions as they carry out their critical roles throughout microbial and community life-cycles. This grand challenge for biology ultimately must be addressed before scientists can develop and test models to predict the behavior of microbes and take advantage of their functional capabilities.

This facility will provide the ultimate testing ground for fully integrated models developed from component models created from ongoing research and from previous facility data, modeling, and experimentation. The experimental capabilities of the facility will drive a new generation of systems models. Essential aspects of the computational challenges and conceptual roadmaps are described in 4.2.1. Theory, Modeling, and Simulation Coupled to Experimentation of Complex Biological Systems, p. 85; and roadmap tables beginning on p. 91 (see Fig. 1. Probing Microbial Communities, p. 176).

The other three GTL facilities will provide new high-throughput production and analysis capabilities to define and understand component parts and processes of microbial systems and analyze physiological and molecular conditions on a global level (i.e., measure the properties of samples comprising large numbers of cells). One of the key insights from recent research, however, is that microbial communities are dynamic and highly structured physically and functionally, suggesting that ensemble measurements that look at properties averaged across many cells can reveal only part of the picture.

To address this challenge, the Cellular Systems Facility will focus on dynamic systems-level studies, ranging from molecular processes within individual living cells to complex, structured microbial communities. Microorganisms in such communities—microbial mats and biofilms, for example—occupy various microniches established as a result of coupled biological, chemical, and physical interactions. Each member of the community carries out unique functions that can vary in space and time but are integral to community stability and overall function. The Cellular Systems Facility will provide the underlying capabilities to allow the spatial and temporal analysis of these complex microbial systems in a concerted and integrated way, from molecular processes to ecosystem functionality. This is a daunting challenge, partly because the complex multicellular drama is playing out at submicron scales. Nonetheless, we will need to dynamically image and functionally analyze the critical substructures and molecular species within microbes and their communities and develop models that describe and predict their behaviors. This capability builds on the Molecular Machines Facility, which will focus on intracomplex imaging to determine molecular makeup and structure and on intracellular imaging to localize machines within the cell. (See box, Cellular Systems Facility Objectives, this page; Fig. 1, p. 176; sidebar, Group Living and Communicating, p. 18; Fig. 1. DOE Genomics:GTL High-Throughput User Facilities, p. 103; and Fig. 3. GTL Facilities: Core Functions and Key Interactions, p. 108, from 5.0. Facilities Overview).

Facility Objectives

- **Relate community composition, structures, and functions to environmental physicochemical conditions measured at the scale of microbial communities—an overlay of community physical and functional maps.**
- **Determine community composition, relative positioning of members, and phenotypes.**
- **Analyze overall community functionality and distributions and fluxes of molecular species.**
- **Dynamically image critical molecules and substructures as they function intra- and intercellularly.**
- **Develop models of microbial function at the molecular, cellular, community, and ecosystems levels.**
- **Provide protocols, data, models, and tools to the community.**

The analytical and conceptual challenges of this ultimate step in systems microbiology will require unprecedented technical and computational resources and infrastructure far beyond the reach of individual investigators.

5.4.1.1. Probing Mixed Microbial Populations and Communities

In microbial communities, the complex and dynamic set of interactions that we seek to analyze is taking place in an area far smaller than the period at the end of this sentence. To understand community function, we first must be able to analyze environmental and community structure and composition at high resolution (Fig. 1, p. 176). On an even-smaller scale (roughly 1/500th the size of a period), we must be able to peer inside an individual microbe in a nondestructive way to locate and continually track essential biomolecules that reveal the cell's inner workings; these biomolecules include DNA, RNAs, proteins, protein complexes, lipids, carbohydrates, and metabolites (Riesenfeld, Schloss, and Handelsman 2004; Johnston et al. 2004; Schaechter, Kolter, and Buckley 2004). Models then must be developed to describe key features of these biological interactions within the physicochemical environment and predict how the system will evolve structurally and functionally. Robust models are required to conceptualize these intricate systems, formulate meaningful hypotheses, manage the ensuing complexity and sheer volume of information that comes from experiments, and, ultimately, allow incorporation of the resulting science into applications.

Key information we seek about microbial-community function includes:

- Community arrangement and its physical environment
- Members and where they reside in relation to each other and their environment

- Phylogenetic, phenotypic, and functional properties of members. Do the genes in one organism regulate gene expression in another organism?
- Microbial interactions among themselves and with their environment
  - Information flow (e.g., genes, signaling molecules) within the community
  - Ecological interactions (e.g., predation, symbioses) among the various members
  - Food webs and communications
  - Excretions, secretions, and consumptions
  - Interactions of secreted components and nutrients with the environment
- Energy and element flow through communities and cells and its regulation
- Intrinsic biological (genetic) and environmental factors that control the structure, stability, and functioning of communities

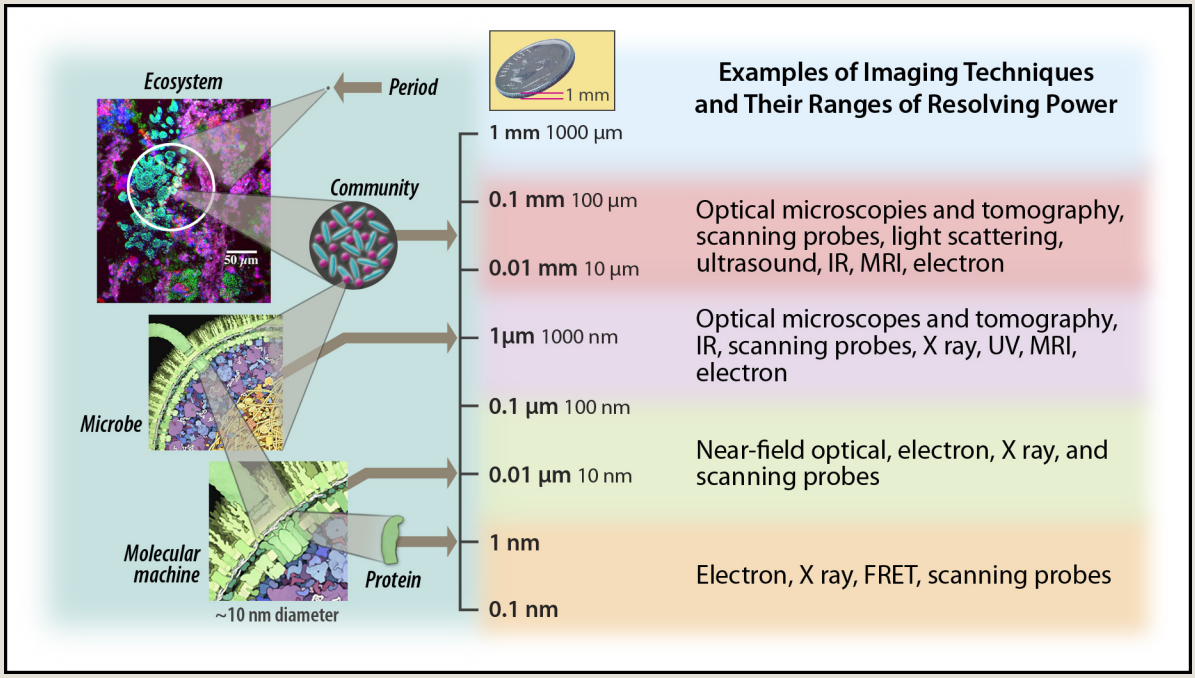


Fig. 1. Probing Microbial Communities. Microbial communities and ecosystems must be probed at the environmental, community, cellular, subcellular, and molecular levels. The environmental structure of a community will be examined to define members and their locations, community dynamics, and structure-function links. Cells will be explored to detect and track both extra- and intercellular states and to determine the dynamics of molecules involved in intercellular communications. Probing must be done at the subcellular level to detect, localize, and track individual molecules. Preferably, measurements will be made in living systems over extended time scales and at the highest resolution. A number of techniques are emerging to address these demanding requirements; a brief listing is on the right side of the figure. These and other techniques are discussed in section 5.4.5, beginning on 181.

- Trajectory of community evolution
  - Life strategies of each population in the community
  - Role of lateral gene-transfer processes in microbial evolution and community metagenome
  - Senescence, death, and turnover rate (consequence of death?)
  - Community resilience (biodiversity, stability)

5.4.1.2. Foundations for Community Analyses

Before undertaking these analyses, we will have a growing body of knowledge (incorporated in the GTL Knowledgebase) and capabilities from work funded by other agencies and within the GTL program and facilities. These resources will include:

- Cooperative analyses of comprehensively annotated genomes of individual microbes and the community metagenome to estimate the genetic potential of individuals and the community.
- Many critical proteins encoded in the community's genome expressed and characterized to produce a substantial body of functional characterization data incorporated into gene-annotation data sets. These data will provide significant insights into "interesting" processes that need to be pinpointed and analyzed in the context of a complete system. Since studies can be performed on the basis of sequence alone, we will have circumvented the fact that these microbes are largely unculturable.
- Ability to produce multiple affinity reagents for any produced proteins and other such biomolecules as RNA and some metabolites that can be used to locate, track, and manipulate these entities within living systems. Fusion tags can be incorporated in a variety of ways.
- Extensive measurements at the global level on bulk and ensemble samples to ascertain the phylogenetic and physiological state of member microbes or the entire community under relevant conditions. Temporal relationships will be revealed through repeated sampling and process interplay via extensive linked measurements of the transcriptome, proteome, metabolome, and other biological and physicochemical variables.
- Analysis of the structure and function of critical molecular complexes in vitro and insights to determine where and in what context they carry out their cellular functions.
- Extensive databases and exploratory tools to begin deriving underlying principles at the molecular, cellular, and community levels and the ability to begin encompassing complexities in detail as processes play out in real, nonlinear, coupled systems.
- Extensive models at molecular, cellular, and community levels to support creating and simulating hypotheses in a systems context. These models and simulations will be used to design and gain insight into experimental campaigns and protocols and provide advanced knowledge of key experimental variables that must be captured in ensuing research.

Even when all this information is at hand, unraveling how all these entities and processes act together in a continuous, concerted way—from molecular to community levels—will remain a grand challenge in accomplishing DOE mission goals. All technologies that created this body of knowledge must be specialized in innovative ways to provide the same information at a microscopic (actually nanoscopic) scale. This facility will be capable of analytical measurements that are nondestructive and done in real time in living cells within a well-defined global and dynamic system. Understanding how these individual cells interact and function as a unit—a microbial tissue in some respects—to carry out complex processes is key to unlocking their vast potential for important applications and achieving our science goals.

5.4.2. Facility Description

The Cellular Systems Facility will combine advanced computational, analytical, and experimental capabilities for integrated analysis of spatial and temporal variations in biological systems—how, when, and why the various system components appear, disappear, function, and remain. The facility will determine the state of cellular systems, from the internal makeup, structure, and dynamics of individual microbial cells to complex communities and their environments. To achieve a systems-level understanding, simulation and modeling must be coupled tightly with experiments to define and analyze the complex regulatory and metabolic processes in microbial cells and communities. This facility will emphasize concurrent and dynamic measurements of proteins, molecular complexes, intracellular metabolites, regulatory molecules, and gene transcripts. The aim is to establish the state of cells within populations and communities as a function of changes in physicochemical and biological conditions, emphasizing measurements at spatial and temporal resolutions appropriate for the entities being measured.

5.4.2.1. Laboratories, Instrumentation, Quality Control, Computing, and Support

The facility will be 100,000 to 150,000 gross square feet, with laboratories containing necessary cultivation, isolation, and analysis instrumentation. Its frontier instruments will incorporate capabilities and techniques for extensive environmental control and monitoring, manipulating communities in real time in various ways, and temporal and spatial resolution. The exact configuration of these instruments awaits necessary technical developments as described below. The facility will have requisite offices, common space, and conference facilities for staff, administration, and users.

The Cellular Systems Facility will provide to the user community:

- Frontier instruments incorporating the capabilities to create, sustain, and monitor structured microbial communities and analyze them at the molecular through ecosystem levels.
- Models and the tools and computing and information infrastructure for developing and evaluating such models, along with the user-facility infrastructure needed to undertake such tasks.
- The GTL Knowledgebase as a source of data on all known aspects of microbial systems that have been studied as a foundation for further experimentation, model development, and the ensuing simulations.

The facility therefore will be highly data intensive, providing extensive linked data sets on the dynamic behavior of microbial cells and communities. It also will be compute intensive, providing unprecedented data analysis, modeling, and simulation. In addition, it will involve new computational approaches for data storage, analysis, and use in complex models. These systems-level data sets will be made available to the entire scientific community. This new knowledge will be invaluable for advancing the annotation of microbial and community genome sequence, identifying regulatory and metabolic networks in microbial systems, and understanding microbial contributions to ecosystems function.

5.4.2.2. Performance and Production Targets

The Cellular Systems Facility will have the experimental and computational capacities to measure and analyze hundreds of microbial systems per year at unprecedented levels of detail. A key distinguishing feature will be the sophistication and performance of its instrumentation and capabilities in computational modeling and simulation. Some key performance features as described in the following sections include:

- Performance of physical (i.e., structure) as well as functional (i.e., molecular profiling) mapping
- Spatial and temporal resolution to map all levels of the functional processes within a microbial community (i.e., nanometers to millimeters and microseconds to hours)
- Performance of nondestructive measurements
- Culturing capabilities for maintenance of realistic environmental conditions and microbial communities

- Integration of modeling capabilities and supporting computing infrastructure, including data-intensive bioinformatics, compute-intensive molecular modeling, and complexity-dominated systems modeling

This GTL facility, more than the others, must overcome major challenges associated with the lack of available technologies and instruments for measuring the dynamic state of living microbial cells. As stated above, it will benefit from technologies, instrumentation, and data developed by current and future GTL R&D and pilot projects by the time the facility comes online. Along with state-of-the-art capabilities when operations begin, the facility also will include extensive ongoing development of essential instruments and technologies to advance the measurement of activities and characteristics of cellular systems at the single-cell level.

DOE has an extensive and successful history of developing and applying new technologies to complex problems in the physical and chemical sciences. As in the genome projects, the agency can draw on multidisciplinary teams of biological, physical, computational, and other scientists and engineers from national laboratories, academia, and industry. GTL brings a tremendous opportunity for using these same talents in the Cellular Systems Facility to devise technological solutions to some of biology's most complex problems.

5.4.3. Technology Development for Cultivation of Microbial Communities

The classic definition of an unculturable microbe is that it cannot be grown in homogeneous suspension. Because of this, there is a dearth of information about the metabolic capacity of microorganisms that resist cultivation under laboratory conditions (Keller and Zengler 2004). This is due primarily to the difficulty or impossibility of simulating the chemistry and interspecies interactions of highly structured communities by suspended cell-culture techniques. Most microbes reside in these structured communities and display unique phenotypic states in response to microenvironments within those communities. Many current technologies require large populations of cells to measure gene expression, proteome, and metabolites, masking the true cell-to-cell heterogeneity. The inability to cultivate most structured communities formed by natural microbial populations limits our discovery of new genes, gene products, and resultant functionalities. New cultivation techniques must support the development of meaningful community structures, and the functions of community members must be measurable in that environment (see sidebar, Laboratory Cultivation Techniques to Simulate Natural Community Structure, p. 180).

The facility's cultivation systems will allow for the precise control and manipulation of environmental conditions and for the monitoring and culturing of microorganisms in meaningful community structures. Many microorganisms of scientific and biotechnological importance, including those sequenced by the Biological and Environmental Research Program, by GTL, and by the Joint Genome Institute's Community Sequencing Program, are relevant to various DOE missions. Some thrive in unusual or extreme environments or those in which gradients or temporal changes occur in physicochemical conditions.

Scientific investigations of these organisms and the communities in which they live thus require flexible, highly controlled, and instrumented systems that can provide a range of environmental conditions and sophisticated measurements. These conditions include monoculture or mixed cultures; nutrient status; extremes of pH, temperature, and salinity; exposure to contaminants and radiation; gas composition and pressure; light intensity; and the presence of solid phases. The ability to control and monitor the environment allows for rigorous investigations and interpretations of gene expression, regulation, and function at the level of individual cells, cell populations, and mixed communities. When required, cells of unusual or difficult-to-culture microorganisms will be cultured in sufficient quantities to provide protein for biophysical, structural, and functional analyses. In other cases, very small numbers of unusual or difficult-to-culture microorganisms might be studied using novel microscale approaches combined with specialized sensitive analyses of gene expression, proteomics, metabolism, and metabolite flux. Ultimately, for meaningful analysis of communities, measurements at the individual cell level will be essential. To select for study any cell in such a structured community, we need to be able to (1) remove it from its environment without inducing significant changes in the properties being measured or (2) conduct analyses of living cells in situ (i.e., without disrupting its environment or harming the cell).

Laboratory Cultivation Techniques to Simulate Natural Community Structure

Microbes associating within a biofilm surface offer the opportunity for members of a discrete population and individual organisms representing different species to establish fixed spatial relationships over extended periods of time. Surfaces enable microorganisms to establish high cell densities in localized areas. For example, products of cell metabolism in a colony of one type of microorganism diffuse to adjacent surface areas, forming strong concentration gradients within the intercellular volume of a biofilm (Beyenal, Davis, and Lewandowski 2004; Beyenal et al. 2004).

To identify the function of genes preferentially expressed by specific populations in the structured community, new cultivation techniques are being developed that incorporate surfaces for microbial colonization and RNA extraction (Finelli et al. 2003). During the past decade, researchers have developed reactors in which biofilms can be imaged using confocal scanning laser microscopy (CSLM) and other light-microscopic techniques (Wolfaardt et al. 1994). When combined with fluorescent in situ hybridization (FISH) to distinguish populations of cells in multipopulation biofilms and fluorescent reporters (green fluorescent protein) of functional gene expression, CSLM has been used to demonstrate how gene expression by one population affects gene expression in another proximally located population (Moller et al. 1998).

Pacific Northwest National Laboratory

The mobile pilot-plant fermentor shown here has a 90-L capacity and currently is used to generate large volumes of cells and cell products such as outer-membrane vesicles under highly controlled conditions. This fermentor allows the end user precise control of culture growth to produce high-quality samples. Future generations of fermentors will be more highly instrumented, possessing sophisticated imaging and other analytical devices developed to analyze interactions among cells in biofilms under an array of conditions.

5.4.3.1. Requirement Examples

- Development of techniques to isolate single cells from a natural community and analyze them for the expression of targeted genes, proteins, and other products (e.g., using fluorescent tags produced from metagenome sequences).
- Evaluation of metabolic processes carried out by cells in structured microbial communities, using molecular tags for RNA, proteins, or other reported moieties to map individual cell populations within the community (see imaging discussion below).
- For unsequenced microbes, application of techniques such as intact biofilm polymerase chain reaction (IB-PCR) to construct 16S rRNA clone libraries once a community is formed. The rRNA sequences of each population present could be used to establish phylogenetic links to other known populations (Reardon et al. in press). Techniques such as FISH then could employ 16S rRNA sequence data to construct oligonucleotide probes to locate different populations within the biofilm and identify putative associations between colocalized populations.
- For many systems, capabilities to assay the distribution of properties among a population of extracted cells from structured communities, sediments, or soils. Providing invaluable information, flow cytometry will be a high-throughput method of choice. Novel cultivation approaches also might be combined with single-cell sorting techniques such as emerging microfluidic lab-on-a-chip devices to grow and study currently uncultivable members of microbial communities.

5.4.4. Development of Genomic Capabilities

To test hypotheses about function, genetic manipulation to generate mutants and specific constructs containing tags or reporter molecules is an essential requirement for systems biology research. Highly roboticized capabilities will be essential for high-throughput construction and screening at the genome level. Examples of required basic capabilities include nucleic acid isolation and analysis, sequencing and annotation, expression analysis, gene cloning and expression, fusion tagging of genes (Gaietta et al. 2002), general tools to manipulate members genetically, and cell sorting. Many of these capabilities will be present in other GTL facilities, the JGI, and in researchers' laboratories and may be incorporated into this facility as needed. Molecular microbiology and, in fact, all microbiology support capabilities will require a highly developed system for information management and integration.

5.4.5. Technology Development for Imaging and Spectroscopy

The Cellular Systems Facility will employ a broad range of imaging modalities to monitor the structure of microbial communities and image (spatially and temporally resolving) the many molecular species critical to community function. The imaging capabilities of the Cellular Systems Facility and the Molecular Machines Facility are complementary—Molecular Machines images intracomplex structure and cellular location, and Cellular Systems focuses on spatially and temporally mapping multiple processes through the lifecycle of a community of cells in a complex environment. Imaging modalities available for both facilities are presented in Table 1. Characteristics of Available Imaging Modalities, p. 182; Table 2. Attributes of Available Techniques for Cellular Systems Characterization, p. 183; and Fig. 1, p. 176, all of which list the primary probe methods available, techniques based on them, analytical characteristics, prospects for further development, and computing requirements. To be applicable, these methods must be chemically specific, perform measurements nondestructively, and be capable of functioning in concert with other techniques (Toner et al. 2005).

5.4.5.1. Analytical Characterization of Cellular Systems

Critical to addressing key scientific questions will be the novel application of existing and emerging technologies that characterize systems in a continuous and spatially resolved way. Analyses now conducted on bulk samples must transition to nondestructive processes capable of characterizing systems ranging from a microbial community through multiple processes within a single living cell. Also, the power to view multiple systems with high spatial and temporal resolution must be augmented with the ability to identify, track, and manipulate living microbes in the presence of other strongly interacting species. To achieve this, many classic imaging techniques must be coupled with methods that can detect specific molecules or processes. Physical, chemical, and biological variables must be identified and tracked. Furthermore, the power to observe systems in action will need to be enhanced by the ability to interact with these systems.

Developing the capability to view biological systems in great detail will enable new high-throughput approaches to studying cellular systems. Each cell in a culture, consortium, or community presents a unique reflection of the biological response to the overall system's changing state. Each provides a set of multilevel outputs in response to the effects of changing parameters (e.g., environmental insults, nutrient gradients, and temperature). To the extent that this parallel data stream can be captured in real time, biological experiments can be conducted in a high-throughput manner rather than running as several series of experiments to evaluate each possible response (e.g., cell division, movement, and protein shedding) for each type of environmental change.

To enable these advances, new technologies must address the special requirements for observing biological systems. Ideally, techniques will be nondestructive, noninterfering, and compatible with the analysis of heterogeneous, living systems. They will need to document the state of each cell (or many cells or cell types) as time and environmental conditions change. Furthermore, physical and chemical information must be mapped onto community structure while detailing changes at the molecular scale. These analyses will necessitate the development of new software to provide intelligent processing of data. Ultimately, such tools will

Table 1. Characteristics of Available Imaging Modalities

	Technique	Unique Characteristics	Future Prospects	Bioinformatics
Visible Light (possible: 50 nm practical: 300 nm)	TIR Absorbance Scattering NLO Adaptive optics FRET Structured light illumination	Noninvasive In situ Wide range of time + length scales Functional analysis Coordinated release of caged molecules Microsurgery, microablation Characterization of individual cells and communities (biofilms)	Better probes, lanthanite dyes, quantum dots, nanoparticles, tetracysteine tags, genetically encoded nanoparticle sensors More versatile excitation sources Better detectors	3D visualization (online, offsite) Pattern recognition (spatial, spectral) Multiscale, multimethod data fusion
X Ray (20 nm)	Tomography Spectroscopy Microprobes	Thick, hydrated samples Whole cells Clean spectrum Organic functional group metal redox spectroscopy Molecular localization in ultrastructural context Characterization interactions	More versatile excitation sources Better detectors	
EM (0.3 nm)	Tomography Molecular microscopy: Single particle Cryo	Whole cells or sections High-resolution molecular localizations in ultrastructural context Correlation with fluorescence	More versatile excitation sources Better detectors	
Force Imaging	AFM tapping	Cell wall imaging Imaging of protein, nucleic-acid components	Better tips (higher-aspect ratio: Carbon nanotubes)	
Force (manipulation, perturbation)	Optical tweezers Magnetic tweezers	Mechanical characteristics (cell wall) Thermodynamics and kinetics of transient interactions Characterization of the molecular-machine mechanochemistry Correlated mechanical properties	Combined single-molecule fluorescence, optical tweezers	

help elucidate the large-scale biochemical organization that characterizes community structure. Such new analytical approaches will be essential for assessing the community's physiological and phylogenetic makeup and for testing predictions derived from theoretical models.

A number of these scientific needs will require fundamental new developments in imaging technology—a transformational goal for GTL biology. Revolutionary advances will be essential for determining the dynamics of communities and their functions under various environmental conditions, defining the physical structure of cells and communities, detecting and tracking extracellular and intercellular molecules to define cell states, and, ultimately, understanding how molecular events are communicated in space and time.

Table 2. Attributes of Available Techniques for Cellular Systems Characterization

Scale of Analysis	Information Needed	Technoques for Structure and Imaging	Static Characterization Techniques	Dynamic Characterization Techniques
Proteins	Components Abundance Structure	X-ray crystallography (angstrom) Raman spectroscopy (angstrom) Neutron crystallography (angstrom) X-ray spectroscopy (sub-angstrom) Electron microscopy (SEM, TEM, STEM, tomography) Electron crystallography	Infrared spectroscopy Raman spectroscopy NMR (nuclear magnetic resonance) spectroscopy Microsampling Microfluidics Fluorescence Scattering	Infrared spectroscopy Raman spectroscopy NMR spectroscopy Microsampling Microfluidics Fluorescence Scattering
Molecular Machines	Components, active sites Function, role, interchangeability, stressed behavior	X-ray crystallography, Raman spectroscopy, neutron scattering, X-ray scattering, EM, multi- and hyperspectral fluorescence	Infrared spectroscopy Raman spectroscopy NMR spectroscopy Microsampling Sensors	Pump-probe spectros. Microsampling Sensors Labels (quantum dots, organic fluorescence) Laue X-ray crystallography
Cellular	Components Active sites Function role Interchangeability Communication Stressed behavior	X-ray microscopy Scanning probes Scanning probe microscopy (SPM) Atomic force microscopy (1.0 nm) Scanning near-field optical microscopy (NSOM or SNOM) Scanning tunneling microscopy Chemical force microscopy Electrostatic force microscopy Magnetic force microscopy Electron microscopy (SEM, TEM, STEM, tomography) Far-field vibrational imaging (>10 microns) Optical microscopy	Mass spectrometry NMR spectroscopy Probes Raman spectroscopy Neutron spectroscopy Infrared spectroscopy SPM PH meter Microsampling Sensors Multi- and hyperspectral fluorescence Optical microscopy (one or multiphoton, scanning optical tomography; 200 nm in conventional mode; 5 nm in FRET/FLIM modes, FISH, CARS, SHM)	Raman spectroscopy X-ray microscopy Scanning probes Mass spectroscopy NMR spectroscopy Probe spectroscopy Infrared spectroscopy SPM Microsampling Sensors Fluorescence Labels
Communities In lab In field	Components, active site, function, role, activators, interchangeability, stressed behavior How to communicate?	Far-field vibrational imaging (>10 microns) Optical microscopy (one or multiphoton, scanning optical tomography) NMR imaging Light-scattering spectroscopy Ultrasound	Infrared spectroscopy Raman spectoscopy NMR spectroscopy Microsampling Sensors	Pump-probe spectroscopy Microsampling Stop-flow chromotography Sensors Labels

5.4.5.1.1. Examples of Analytical Requirements

Intracellular Structure. Intracellular protein, RNA, and metabolite localization and kinetics of localization.

- Proteomics on replicate communities.
- Fine-scale cell ultrastructure.
- New multimodal capabilities for dynamic imaging of targeted intracellular molecules and their interactions (including machines) in individual cells and cell assemblies [e.g., with antibody labeling and electron microscopy (EM)].

Community Structure. Analytical instrumentation and techniques for determining overall community structure and identifying and characterizing spatial and temporal variations in metabolites, signaling and regulatory molecules, and the physicochemical environment within communities (see discussions under the Molecular Machines Facility, beginning on p. 143).

- Probes for the in situ measurement of extracellular metabolites in real time.
- Imaging and spectroscopy of population structure, gene expression, and metabolites in cell aggregates and subpopulations within communities.
- Characterization of cells in mixed communities by multispectral imaging of key cellular chromophores, possibly moving to on-the-fly cell-sorting platforms.
- Measurement of elemental distribution, oxidation state of elements, and biomolecules within and among communities.
- Quantitative imaging of metabolite (and signaling molecule) flux between cells in close proximity to or in contact with each other—one of the most critical needs for understanding how microbial communities function.
  - Analysis and assessment of the makeup and role of extracellular polymers in community structure, function, and stability.
  - Detection and frequency of genetic exchange, recombination, and evolution within communities.
  - Determination of macroscopic transport of water, solutes, and macromolecules and their relationship to microbial function.
- Characterization of interface physical and chemical properties.

Identified Development Needs

- Advanced chemical and biological probes, including engineered microorganisms, tagged biomolecules, and chemical sentinels that will help characterize microbial communities.
- Advanced tools for imaging characterization for use in the laboratory as well as in the field.

A variety of imaging and microspectroscopic techniques are emerging to meet these challenges. In general, imaging relates spatially dependent information. Characterization of additional dimensions, however, will be essential for relating system activity. Some commonly used imaging techniques include:

- **Short-Wavelength Techniques.** Analyses with electrons and X rays typically provide the highest spatial resolution. Although commonly associated with ultrastructural analyses, short-wavelength techniques are being extended for analyses of whole cells at atmospheric pressures. Additionally, X rays are useful for mapping trace metals, while spectroscopic measurements can provide chemical identification.
- **Optical Microscopies.** The current standard for live-cell imaging, these tools are ideal for studying dynamics across a broad range of time scales and are sensitive down to the single-molecule level. A number of physical scales can be assessed, and emerging techniques and new labels are improving the sensitivity and resolution of optical microscopy.

- **Long-Wavelength Techniques.** A variety of these procedures including vibration, magnetic resonance, and terahertz-based imaging can provide essential information on chemical structure, identity, and spatial arrangement. For example, vibrational signatures are molecularly specific and can produce direct chemical information without additional labels.
- **Other Techniques.** A broad range of unconventional imaging approaches are making an impact on biological studies. Most notable, the family of instruments comprising scanning probe microscopy enables molecular-scale resolution; and chemical, electrical, and physical properties can be measured simultaneously. Emerging tools based on optical and magnetic trapping are allowing measurement of mechanical properties while micro- and nanoscale structures permit sensing of chemical, physical, and biological attributes.

Clearly, many current imaging and microspectroscopic techniques possess significant attributes and provide information relevant to the study of biological systems. Significant advances still are needed to adapt many of these tools to the characterization of microbial cellular systems much smaller than eukaryotic cells. Advanced instrumentation, improved biocompatibility, new approaches for targeting and delivery of tags, and improved labels are but a few of the significant challenges that face imaging technologies. More significant, no technique alone can provide the broad range of information needed to understand community structure and system function. A combination of methods will be essential to extend the depth of information required.

5.4.5.1.2. Monitoring and Interacting with Cellular Systems

To enable effective systems-level studies, the ability to monitor systems in action must be enhanced with selective construction, manipulation, and interaction with the system. Only then can efficient experimental evaluations and effective iterations be achieved with pursuits in theory, modeling, and simulation. This integration will be a culminating product of the facility and an essential tool for studying microbes, consortia, and microbial communities.

Advanced cultivation systems that allow for precise control, manipulation, and monitoring of environmental conditions must be compatible with advanced imaging technologies. Chemical gradients will need to be controlled and monitored precisely while temporally measuring molecular-scale properties. Genetically defined organisms must be carefully arranged into ordered microbial communities, perhaps through molecular-scale patterning techniques resulting from nanotechnologies. Such highly defined systems will require integration with sensing capabilities and the ability to activate biomolecular networks remotely. The capacity for simultaneously imaging and specifically targeting reagent release or activation, as currently used in biomedical applications, is within reach for GTL systems biology studies. The creation of such compound, multifunctional instruments will enable the collection of information needed to understand and exploit complex biological systems.

5.4.5.1.3. Technology Development Progress and Benefits

5.4.5.1.3.1. Advanced Optical Methods – Laser or Synchrotron Based

- Optical spectroscopic methods can be used as tools for noninvasive characterization and monitoring of dynamic behavior.
- Measurements of absorption and in vivo fluorescence can be used to monitor the presence and relative concentration of optically active biochemical species.
- Light-scattering spectroscopy can probe the size distribution of community structures.
- Vibrational (infrared and Raman) spectroscopy is a technique for studying the composition of biological materials without perturbing or labeling the sample. Biological components (e.g., lipids, proteins, nucleic acids, and carbohydrates) and biofilm and microbial surfaces (e.g., minerals and polymers) have unique vibrational spectra based on their chemical structures.

- Use of these methods will provide new information on the following:
  - Large-scale (1- to 10-micron) biochemical organization.
  - Composition and distribution of extracellular polymer matrices.
  - Concentration and distribution of nutrients, metabolites, signaling molecules, and other macromolecules.
  - Interactions of biofilms and microbial communities with supporting surfaces.

Because vibrational spectromicroscopy is noninvasive, it can be performed on dynamic living systems in combination with other techniques. If synchrotron radiation is used as the photon source, a dynamic system can be studied directly on surfaces of geological materials (see Fig. 4.1, p. 27, Report on Imaging Workshop 2002).

Significant progress already has been made using confocal and two-photon fluorescence microscopy. The specificity of these techniques is provided by the exogenous chromophore targeted through an affinity reagent or fusion tag to a particular protein. The resolution is on the order of a micron and slightly higher for two-photon than for confocal microscopy. Delivering chromophores to remote regions within a community or cell is a particular challenge. Additionally, the identification of probes that maintain activity in diverse environments is required (see Fig. 4.2, p. 28, Report on Imaging Workshop 2002).

All these techniques can be used in an imaging arrangement to monitor changes in community behavior in real time. Improvements are needed in such areas as spatial resolution, the ability to provide quantitative information, and data-acquisition speed. Additionally, advanced light-microscopy techniques can be developed for high-resolution 2D and 3D mapping. Often with specificity to particular components associated with imaging, these techniques include surface-plasmon resonance, surface-enhanced Raman spectroscopy, imaging of second-harmonic generation, optical-coherence tomography, and coherent anti-Stokes Raman scattering.

5.4.5.2. Imaging Macromolecular Complexes

Many types of imaging technologies can be employed to identify and spatially and temporally localize macromolecular complexes and their interactions within a dynamic community environment. Some specialized techniques have specific applications to the analysis of macromolecular complexes in situ in live, fixed, or frozen cells or ex situ. The strengths of imaging techniques typically include detection sensitivity and the ability to identify complexes in cells. Imaging techniques are applicable to all classes of complexes. In many cases, however, the identities of one or more components of the complex must be known to prepare tagged probes for imaging analysis. This requirement limits the application of imaging to full identification of protein complexes. Currently, most imaging techniques are relatively slow; automation, however, is providing faster sample throughput, and improved computational tools are enhancing data acquisition and analysis. Imaging techniques relevant to identification and characterization of protein complexes are summarized below, with additional information on other imaging tools in Table 2, p. 183.

Tagged Localization. Used with visible, X-ray, or electron microscopies to identify sets of biomolecules labeled with tags. An in situ method applicable to live (visible light), fixed, or frozen cells, it also is applicable to tagged transient complexes and membrane-associated complexes. A limitation is that the complex must be labeled with a tag, requiring tag synthesis and introduction into cells. Spatial resolution in these modalities comes from the instrument response function of the exciting source (i.e., the exciting beam provides the resolution). More developed X-ray optics, more versatile excitation sources, and improved probes are needed. Lanthanide dyes, quantum dots, nanoparticles, tetracysteine-based ligands, and other probes are examples of some recently reported probes used with various imaging modalities.

Fluorescence Resonance Energy Transfer (FRET). Used to identify pairs of biomolecules labeled with tags as well as to provide information on biomolecule relationships. This in situ method is applicable to live cells, tagged transient species, and membrane-associated complexes. FRET is particularly good for structure and binding of extracellular ligands. Like other imaging techniques, it requires tag synthesis and introduction into cells.

Scanning Probe Microscopy. Identifies protein associations by scanning with a specific molecule attached to the tip, including transient molecules. The technique is capable of very high spatial resolution, depending on the length of probe time, and of single-molecule detection. It is most suited for the study of membrane-associated complexes with whole cells or for the study of isolated complexes. Like other imaging techniques, it requires that the identity of one component of the complex be known so a molecule can be attached to the tip as the probe molecule. The probe, for example, then can be used to identify interaction sites on a cell surface. The technique is labor intensive and slow. Identification is a one-at-a-time process unless multiprobe devices with individual probe molecules are employed. These multiprobe devices are under development to allow technique application in a highly parallel fashion. Computer modeling of protein folds would enhance data interpretation, and improved computation is needed for data visualization and manipulation.

5.4.5.3. Development Options

As previously mentioned, many techniques required for the facility have yet to be developed sufficiently to analyze microbes of less than one micron in complex and changing communities. Many potential options must be explored over the next few years to determine probe and detection modalities capable of providing necessary information under these demanding conditions. Options that may be explored regarding available techniques, their range of applicability, and information they might provide are shown in Table 2, p. 183, and Table 3. Cellular Systems Facility Technology Development Roadmap, p. 188. The bulk of intracomplex characterization of molecular machines will be carried out in the Molecular Machines Facility. The sidebar, The Super Imager, this page, details creation of super imagers comprising compound, multifunctional instruments that individually would include many of the capabilities listed. Many of these development issues are summarized in 6.0. Development Summary: Global, Crosscutting, and Long-Lead Issues, p. 191.

The Super Imager

The potential is to create compound, multifunctional instruments that individually include many of the following capabilities:

- **Mapping of molecular species such as RNA, proteins, machines, and metabolites through the use of fluorescent tags of various kinds**
- **Multiple excitation and detection wavelengths including both fluorescent and infrared absorption methods**
- **High-speed 3D imaging**
- **Nonlinear contrast imaging including second- and third-harmonic generation and coherent Raman scattering**
- **Lifetime mapping as sensitive probes of local environments**
- **Rotational correlation mapping for in situ analysis of protein structure and function**
- **Magnetic resonance imaging with 10-micron–scale analyses of metabolite concentrations and providing data on diffusion properties and local temperatures**
- **Acoustical imaging of the system's physical parameters with micron-scale resolution**
- **Atomic force microscopy (AFM) mapping of structures with added information provided by the controlled-interaction light with sharp metallic AFM tips to obtain optical resolutions of ~20 nm, one-tenth the diffraction limit**
- **High spatial resolution (nanometer scale) using X-ray and electron microscopies, including the use of special DOE facilities or perhaps the development of laboratory-based X-ray sources for imaging**

[Source: *Report on the Imaging Workshop for the Genomes to Life Program April 16–18, 2002* (Office of Science, U.S. Department of Energy, Nov. 2002); www.doegenomestolife.org/technology/imaging/workshop2002/]

5.4.6. Development of Computing Capabilities

Computational tools and infrastructure are required for efficiently collecting, analyzing, visualizing, and integrating large data sets to elucidate gene function and to model and simulate regulatory and metabolic networks, cells, communities, and ecosystems). These tools will support the development and validation of

Table 3a. Cellular Systems Facility Technology Development Roadmap

Technology Objectives	Research, Design, and Development	Demonstration: Pilots and Modular Deployment	Integration and Production Deployment	Facility Outputs
Technologies for Cultivation of Microbial Communities Precise control, manipulation, and monitoring of environmental conditions; interrogation Functional individual microbial cells in the context of characterized physiochemical environment Support for formation of structured communities	Requirements defined for analyzing individual cells within structured communities: Mixed microbial cultures • Suspended and structured • Biofilms Methods and approach to identify and track microbes and molecular complexes • Tagged probes » Increased variety of signals » Signal interpretation » Incorporation in cell • Arrays	Multiple flexible experimental systems to control and manipulate growth and conditions with multiplex measurements of activity including: • Chemostats • Microtechnologies • Remote sensing • Imaging • Surfaces to nucleate biofilms and other structures Multiple probes to identify community members Temporal monitoring of community structure and function	Integration of culturing capabilities within multiprobe instrumentation for simultaneous control, manipulation, and multimodal analyses of structured communities: • Nondestructive • Real time • Linked databases • Environmental, community, cellular, and molecular levels	Integrated, highly characterized, and real-time manipulatable structured microbial communities that simulate natural communities and niches: • Protocols • Extracted samples • Characterizations • Analytical images
Environmental Communities Sampling In situ measurements	Lab techniques extended to field use	Planned extension after operations begin		
High-Throughput Cultivation for Single-Cell Analysis Sampling techniques Controlled viable growth of single cells	Analysis from within structured communities, in microculture extracts, or in place: • Cell sorters • Lab on a chip and microfluidics • Single-cell analysis of "unculturable" environmental samples	Assessment of compatibility with analytical instrumentation and simulation fidelity of natural environments	High-throughput operational mode combining culturing techniques interfaced with multimodal, analytical, and manipulation modalities	Single cells prepared in conditions that simulate microniche environments in highly structured microbial communities such as biofilms (formerly unculturable)

(continued next page)

theories and models of community growth, function, and environmental response. New theory, algorithms, and implementation on high-performance computer architectures also are needed for modeling and simulating cellular systems. Enabling a broad range of biologists to access the large data sets and computational resources for discovery-based biology will require the development of web- and grid-based technologies (see 4.0. Creating an Integrated Computational Environment for Biology, p. 81, and Table 4. Computing Roadmap, p. 190).

Table 3b. Cellular Systems Facility Technology Development Roadmap

Technology Objectives	Research, Design, and Development	Demonstration: Pilots and Modular Deployment	Integration and Production Deployment	Facility Outputs
Temporal and Spatial Localization of RNAs, Machines, and Metabolites Analytical measurement contexts: • Environmental • Community • Intercellular • Intracellular	Requirements defined for multimodal measurements: • Environmental physicochemical variables • Intercellular biomolecules • Intracellular biomolecules • Metabolites • Community overall biochemical and biophysical functionality Examples of needed instrumentation with molecular specificity, sensitivity, and spatial and temporal resolution: • NMR for community-scale microscopy (e.g., metabolites, signaling molecules) • Small molecules in living cells • Gene expression in living cells • Proteins and machines in living cells, including dynamics and interactions • Biomolecular mapping microscopies [confocal, CryoEM, SPM (AFM, STM, others)] • Image-interpretation tools • Visualization • Computational systems • Databases	Modular analytical and imaging instrumentation and methods integrated with culturing, monitoring, control, and manipulation modalities to assess: • Viability of integrated approaches • Compatability with living systems • Intermodal interactions • Ability to meaningfully assess single cells • Data integration • Simulation and modeling integrated into experimental methods • Visualization of multimodal analyses and system monitoring and manipulation	Integrated culturing capabilities within multiprobe instrumentation for simultaneous control, manipulation, and multimodal analyses of structured communities: • Nondestructive • Real time • Linked databases	Characterizations of microbial communities in realistic environments at the environmental, community, cellular, and molecular levels: • Spatial • Temporal • Functional • Process • Molecular Databases and query tools Protocols QA/QC

To develop and incorporate the necessary technologies and methods into a high-throughput production environment, a phased process will be followed as described in this roadmap. The process includes research, design, and development; modular and pilot-scale deployment; and final integration and scaleup into operational procedures.

Table 4. Computing Roadmap: Facility for Analysis and Modeling of Cellular Systems

Topic	Research, Design, and Development	Demonstration: Pilots and Modular Deployment	Integration and Production Deployment
LIMS and Workflow Management Participate in GTL cross-facility LIMS working group	Available LIMS technologies Process description for LIMS system Crosscutting research into global workflow management systems Approaches to guiding experiment-based production protocols to inform how best to produce a protein as an AI system helps develop strategy for production	Prototype cellular systems LIMS system* Characterization design strategy system Workflow-management system for identification and characterization Process simulation for facility workflow	Cellular systems LIMS and workflow system Workflow integrated with other GTL facilities and experimental strategy systems
Data Capture and Archiving Participate in GTL cross-facility working group for data representation and standards	Data-type models* Technologies for large-scale storage and retrieval Preliminary designs for databases	Prototype storage archives Prototype user-access environments	Archives for key large-scale data types* Archives linked to community databases and other GTL data resources GTL Knowledgebase feedback
Data Analysis and Reduction Participate in GTL cross-facility working group for data analysis and reduction	Algorithmic methods for various modalities* Grid and high-performance algorithm codes Design for tools library Approaches for automated image interpretation in confocal light microscopy/FRET	Prototype visualization methods and characterization tools library* Prototype grid for data analysis, with partners Prototypes for automated image interpretation in confocal light microscopy and FRET Analysis tools linked to data archives	Production-analysis pipeline for various modalities* on grid and HP platforms Large-scale experimental data results linked to genome data Automated image interpretation in confocal light microscopy and FRET Repository for production-analysis codes Analysis tools pipeline linked to end-user problem-solving environments
Modeling and Simulation Participate in GTL cross-facility working group for modeling and simulation	Existing technologies explored for cell-system modeling and simulation Research methods for reconstruction of protein interaction, regulatory networks, metabolic pathways, and community interactions Mathematical methods for multiscale stochastic and differential-equation network models	Experimentally guided metabolic reconstruction Signaling and regulatory-network reconstruction and simulation Efficient modeling methods for community-interaction networks Mature methods for reconstructing protein-interaction and regulatory networks	Production pipeline and end-user interfaces for cellular and community-level combined network reconstruction and simulation Production codes for image time-series analysis
Community Data Resource Participate in GTL cross-facility working group for serving community data	Data-modeling representations and design for databases: In vivo protein expression and localization, cell models and simulations, community models and simulations, cellular and community methods and protocols	Prototype database End-user query and visualization environments Integration of databases with other GTL resources	Production databases and mature end-user environments Integration with other GTL resources and community protein-data resources
Computing Infrastructure Participate in GTL crosscutting working group for computing infrastructure	Analysis, storage, and networking requirements for cellular systems data Grid and high-performance approaches for large-scale data analysis for cellular and community networks and simulations and to establish requirements	Hardware solutions for large-scale archival storage Networking requirements for large-scale grid-based MS and image data analysis	Production-scale computational analysis systems Web server network for data archives and workflow systems Servers for community data archive databases

* Data types and modalities include MS, NMR, neutron scattering, X-ray, confocal microscopy, cryoEM, and process metadata. Large-scale experimental data results are linked with genome data, and feedback is provided to GTL Knowledgebase.