

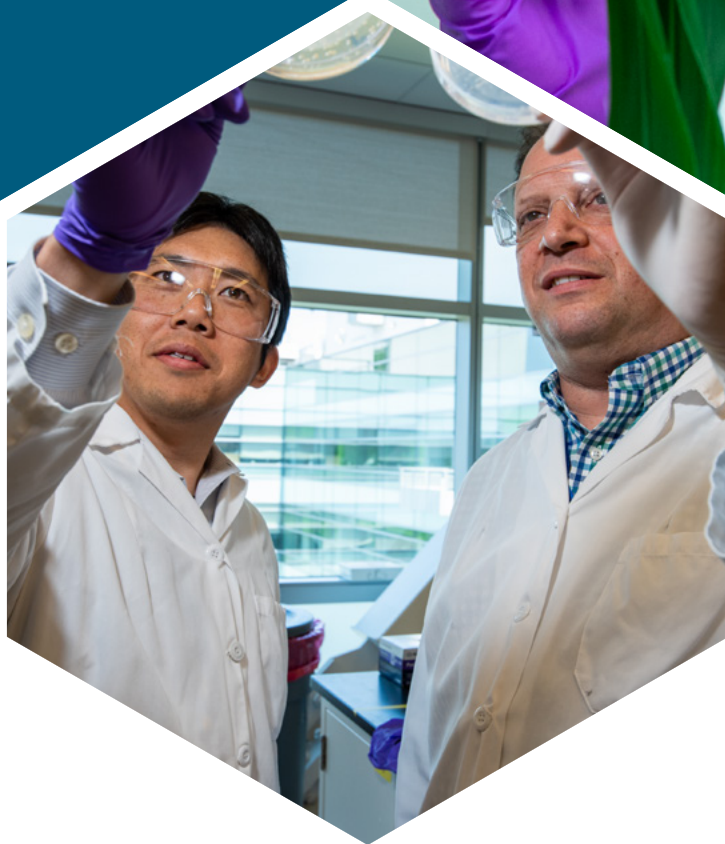


Lawrence Berkeley National Laboratory

BIOSCIENCES
10-YEAR SCIENTIFIC
STRATEGIC PLAN
2013-2023

2019







Lawrence Berkeley National Laboratory

BIOSCIENCES
10-YEAR SCIENTIFIC
STRATEGIC PLAN
2013-2023

REVISION 2019



INTRODUCTION



In 2017, I was appointed the Associate Laboratory Director for Biosciences at Berkeley Lab and am fortunate to lead such a vibrant research area. I believe deeply in the mission of the national laboratories to address national- and global-scale challenges using team science approaches across scientific disciplines. Berkeley Lab's Biosciences Area has long been a leader in such multi-disciplinary projects, from early discoveries in photosynthesis to today's long-standing programs in biofuels research. When we as an Area first embarked on strategic planning activities in 2012, our staff was given an opportunity to develop team science concepts that could advance as a consequence of our strengths in a multitude of biological research areas, including biophysics, biological engineering, environmental sciences, genomics, and computation for biology.

In 2019 as this new version of the Biosciences Strategic Plan (BSP) goes to press, powerful new technologies and methods such as machine learning and genome editing are enabling new advances in biological research, and the accelerating integration of physical sciences, mathematics, and engineering with biology is enabling new approaches to scientific inquiry that were previously not possible. It is through these advances that solutions to societal challenges such as access to sustainable and affordable sources of energy may finally be realized.

2019 marks the sixth year of active implementation of the BSP. In 2018, we assessed our progress towards completion of the five-year milestones described in the 2013 plan. We undertook a comprehensive assessment of our milestones, asking those responsible for leading their implementation if they had been achieved. An impressive 74% of the five-year milestones articulated in 2013 were considered completed, while another 21% were still in progress. Because most of our research aspirations were unfunded in 2013, Biosciences developed a program-development strategy to align our aspirations with potential funder interests and missions, and—more importantly—with great science befitting a national lab that our researchers were passionate about conducting. Our program-development strategy brings together a technical champion, professional program developer, and institutional support to advance national-scale research program concepts that engage scientific societies and research communities outside of Berkeley Lab. During the first five years of the plan, this formula resulted in five new programs that are important for achieving our 10-year Goals in Environment, Health, and Biomanufacturing.

The BSP was last refreshed following a complete reorganization of Biosciences' scientific and operations organizations. During that reorganization, new directors were hired for the scientific Divisions and program development was made a key element of their job descriptions. Our directors have taken different successful approaches to developing new research concepts and programs, and brought together scientists across disciplines at a scale that is only possible in national labs.

Under our Laboratory Director Michael Witherell, deliberate stewardship of our research, people, and resources has become a guiding principle across the lab. Within Biosciences, I am committed to upholding this principle and implementing it across our Area. Through formal and informal processes, we work with our sponsor partners to ensure our research programs fulfill the intended missions and meet their expectations. The BSP serves as an organizing tool that allows us to respond to funding opportunities and to suggest new avenues for future funding. In 2017, Biosciences staff, along with our colleagues at the Lab, developed a series of white papers that identified ways of addressing the Grand Challenges identified by the Biological and Environmental Research Advisory Committee that advises the Department of Energy's Office of Biological and Environmental Research (BER). Many of these white papers were informed by the BSP and discussions sparked by the plan. This is just one way that we respond to the national-scale challenges relevant to the BER.

As ALD, I am committed to stewardship of our talented Biosciences staff. When I think about the future of research and what it will take to achieve the five 10-year Goals outlined in the BSP, I believe it will require a workforce with diverse experiences, backgrounds, and expertise. Since becoming ALD, I have charged all hiring committees to consider how we recruit, on-board, and retain our workforce. For example, all hiring committees now take implicit bias trainings and are composed of members that span multiple organizational groupings. I've also made it a practice to encourage our early-career researchers to learn more about how the Department of Energy and other federal sponsors fund research.

Since the seeds of the Biosciences Strategic Plan were planted in 2012 planning sessions, transformational changes in science and technology have catapulted our approaches. This year's version of the BSP, more purposefully and strategically integrates computing throughout our scientific strategies to accelerate discovery and application of biological data. We have also responded in recent years to externalities such as the 2016 National Microbiome Initiative. Long an area of interest for Biosciences researchers, recent reports from the federal government and interest from stakeholders in agriculture and health research have spurred growth in this space. To assure an honest dialogue about responsible research conduct in engineering biology, since our last BSP refresh we have added a world-renowned bioethicist to our external advisory committee to ensure that our research is conducted with consideration of possible implications in mind.

Our efforts to develop, implement, and monitor our strategic plan are aligned with the planning for a BioCampus on Berkeley Lab's main site. Most of Biosciences' researchers are located in rented space in West Berkeley, Emeryville, and in Walnut Creek. The Integrative Genomics Building, the first building on the Biocampus, was dedicated in June of this year and is the new home of the Joint Genome Institute and the Systems Biology Knowledgebase. The second BioCampus building, the Biological and Environmental Program Integration Center (BioEPIC) is currently in an early planning phase and is envisioned to house flagship programs in environmental biology that bridge the gap from laboratory to field-scale experiments. These new buildings will allow Biosciences researchers to connect to their Berkeley Lab colleagues in computing sciences, earth and environmental sciences, and materials sciences to achieve the goals of the BSP. With a newly refreshed BSP and a new BioCampus building, 2019 is a very big year for the Biosciences Area.

Mary Maxon, Ph.D.

Associate Laboratory Director for Biosciences

TABLE OF CONTENTS

A decorative graphic on a dark blue background. It features two lines that meet at a central point, forming a V-shape. The line on the left is a light grey color, and the line on the right is a light green color. Both lines extend from the top right towards the bottom left, meeting at a point in the lower-middle section of the page.



Introduction	2
Biosciences 10-year Scientific Strategic Plan 2013-2023	6
Bioeconomy	7
A Strategic Framework	10
Four Key Scientific Challenges	11
Forward Looking, Secured by the Past	14
Biosciences at Berkeley Lab	16
Energy	21
Environment	35
Health	49
Biomanufacturing	63
Technologies	77
Five-year Progress Report Summary	88
Bioscience Capabilities	90



Lawrence Berkeley National Laboratory

BIOSCIENCES 10-YEAR SCIENTIFIC STRATEGIC PLAN **2013–2023**

REVISION 2019

Lawrence Berkeley National Laboratory's (Berkeley Lab) Biosciences' 10-Year Scientific Strategic Plan describes the vision for a national future strengthened by biological research achievements, and provides guidance for biosciences research activities at Berkeley Lab. It establishes a framework for maintaining that vision and achieving these goals from 2013-2023. Here we describe here the large-scale biological science challenges appropriate for a national laboratory and relevant to Berkeley Lab's own mission and values. Our plan lays out ambitious goals and relies on our capacity for multidisciplinary, collaborative research to bring bioscience solutions to the world. We have sharpened our focus on the uses of bioscience to address the energy needs of our nation, protect the environment, understand and improve health, and develop novel biomanufacturing technologies. Our plan is meant to be both a blueprint and a catalyst for achieving these goals. The 2019 version of the 2013 plan has been updated to reflect a deeper integration of computation with biological research—a vision that will accelerate discovery and application of biological processes through advanced computational methods. The 2019 version also further refines Biosciences' aspirations in environmental and health-related research to focus on the importance of the microbiomes and efforts to detect, understand and predict changes for applications in environmental remediation and biosecurity.

BIOECONOMY

In 2012, the United States was one of the first countries in the world to develop a bioeconomy strategy, defining a bioeconomy as "one based on the use of research and innovation in the biological sciences to create economic activity and public benefit."* Today, nearly 50 countries around the world have developed their own bioeconomy strategies that differ based on the prevalence of natural resources such as biomass and biodiversity, as well as scientific and technological advantages, such as the use of biotechnology and machine learning. Also in 2012, the Biosciences Area began strategic planning activities that aimed to maximize existing assets at Berkeley Lab such as capabilities in synthetic biology, genome mining, and supercomputing to achieve national-scale goals in energy and environment. Today, Berkeley Lab is home to a growing number of bioeconomy assets in biological and environmental science, and contributes to economic development and public benefit through demonstrated commitments to and achievements in research befitting a national lab, translating discoveries to industry, and stewardship of a diverse and multidisciplinary workforce.

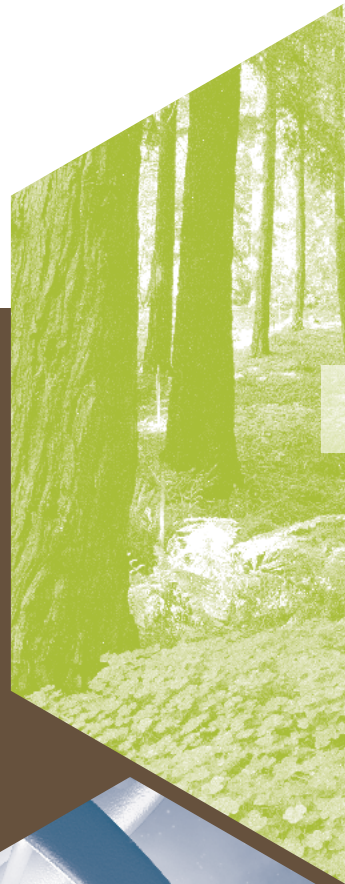
*White House, National Bioeconomy Blueprint; White House: Washington, DC, USA, 2012.

MISSION

Use integrated research teams to solve national challenges in energy, environment, health, and biomanufacturing

VISION

Berkeley Lab's Biosciences Area will lead the nation in using biology to solve energy and environmental challenges





ENVIRONMENT



ENERGY



BIOMANUFACTURING



TECHNOLOGIES

A STRATEGIC FRAMEWORK

Our strategy is grounded in the core mission of national laboratories to carry out basic research that addresses the nation's most pressing science and technological challenges. Our plan is informed and guided by the Berkeley Lab mission:

- Solve the most pressing and profound scientific problems facing humankind
- Conduct basic research for a secure energy future
- Understand living systems to improve the environment, health, and energy supply
- Understand matter and energy in the universe
- Build and safely operate leading scientific facilities for the nation
- Train the next generation of scientists and engineers

With that mission in mind, more than 650 Biosciences researchers and staff participated in an inclusive process to develop this plan, using surveys, focus groups, formal meetings, and informal discussions. Here we present a strategy to enhance and exploit key competencies associated with our discovery-sciences mission. Our overarching goal is to use basic research in bioscience to discover and illuminate paths toward practical solutions. We plan to accomplish our objectives with a strategy built on a framework of historical strengths and well-defined metrics of future success.

FOUR KEY SCIENTIFIC CHALLENGES

This 10-year Biosciences Strategic Plan focuses on four research areas: Energy, Environment, Health, and Biomanufacturing. In these areas, Berkeley Lab expertise is deep, the national need is great, and our commitment to solve problems is unshakeable. In each of these areas, our culture of team science and cross-disciplinary research can be brought to bear for maximum efficiency. Here are the challenges in each area:

Energy Research

How can we efficiently, cost-effectively, and sustainably transform the energy in lignocellulose and plentiful gas feedstocks to liquid transportation fuels using biological or bio-inspired approaches?

Environment Research

How can we develop principles to understand and model the dynamic reciprocity between organisms and their environments to predict how they will respond to changing environments? How do we manipulate environmental systems to remediate contaminants, build soil carbon, and support sustainable agricultural productivity?

Health Research

How do environmental challenges affect the health of diverse organisms in complex biological systems? What are the impacts of environmental changes and exposures, particularly due to anthropogenic impacts and perturbations, on biological systems and human health? Can we use fundamental knowledge of biological diversity in response to exposure to enable prediction, prevention, and treatments?

Biomanufacturing Research

Can significantly reducing the cost and increasing the speed of engineering biological systems transform manufacturing in the United States? Can we solve challenges in energy, agriculture, health, and environment with new biomanufacturing approaches?

Foundational Technology Development

To achieve the ambitious scientific goals outlined in the above research areas, new technologies and the continued development of existing technologies are required. Biosciences' long history of technology development and application to challenging scientific problems has led to revolutions in bioimaging and functional genomics. With strategic technology development, new advances will enable future breakthroughs.

Progress That Is Measurable

In each of these four research areas and foundational technology development, our strategic plan specifies 10-year metrics and five-year milestones to assess progress toward and achievement of the four 10-year Goals, outlined below. Metrics along the full spectrum of research—from basic to applied—were selected to describe our vision for success and underscore our commitment to the breadth of scientific achievement from early discovery to applied solutions for each of our four primary research endeavors.

New learning and discoveries underpin successful approaches to tackle our increasingly complex scientific and societal challenges. Basic research is not intended to lead to immediate commercial benefit, but to new knowledge and theory. Basic research and discovery have historically played a foundational role in technological innovation.

A comprehensive understanding of a biological system enables predictions of how it will respond under certain conditions. It makes possible the reconstruction and redesign of components of the system—capabilities needed to move discoveries closer to solutions that address societal challenges.

Berkeley Lab is dedicated to "bringing science solutions to the world." Technologies developed at the Laboratory have generated billions of dollars in revenue and thousands of jobs. Berkeley Lab breakthroughs in energy-sparing technologies, such as more efficient lighting and windows, have also saved billions of dollars for industry and consumers.

With our metrics in place at the outset, we'll track our progress toward success in each of the four research areas along this "discovery-to-solution" paradigm.

Implementation

The BSP lays the foundation for strategic program development, implemented by a team of volunteer Mentors and Strategy Leads. Each section of the plan (Energy, Environment, Health, Biomanufacturing, and Technologies) is headed by a Mentor, who ensures that the strategies contained within are harmonized and offer opportunities to successfully meet the 10-year Goal. Each strategy is led by a Strategy Lead, who is a technical expert in that research area. While they may or may not actively be doing related research related to their strategy, Strategy Leads are the on-the-ground experts for implementation, identifying new opportunities or areas where Biosciences' could expand its strengths. The Mentor and Strategy Lead roles are independent of Biosciences' operational structure and implementation team members work together across the Divisions and User Facility. These roles offer staff an opportunity to gain additional experience in national-scale strategic planning and program development outside of their research duties. More information on the Mentors and Strategy Leads can be found on the Biosciences website.



FORWARD LOOKING, SECURED BY THE PAST

The long-term goals developed in this planning process tap into the core scientific competencies established at the founding of Berkeley Lab over 80 years ago and strengthened over decades. These competencies, which sprang from the Laboratory's focus on physical sciences and the synergistic academic environment offered by the University of California (UC) at Berkeley, evolved early in its history to include state-of-the-art biological science. Berkeley Lab investigators played pioneering roles in recent revolutions in genomics, computation, biophysics, synthetic biology, and imaging that are continuing to change the way biological research is conducted.

Our 10-year strategy will enhance Berkeley Lab's role as a leading center for the use of biosciences to meet national objectives for energy, environment, health, and biomanufacturing. As we execute our 10-year plan, we carry forward a legacy of transformative research. From the diversity of microbes, microbial communities, and plants, we will continue to uncover nature's secrets to gain a deeper insight about how biological systems work, how they interact with each other and with their environment, and how they can be manipulated to harness their processes and products. From the potential encoded in an organism's genome, we will work to define the principles that guide the translation of the genetic code into functional proteins and pathways. We will continue to advance our understanding of the metabolic and regulatory networks that underlie the systems biology of plants, animals, and microbes as they respond to their environments. Inspired by these processes, we will explore biological means to manufacture new and alternate sources of energy and materials that require less energy to produce and can restore balance to natural carbon cycles.

Biosciences at Berkeley Lab has been successful by many measures: high-impact scientific journal articles, promising technologies transferred to industry, multiple spinoff companies, mentoring and promoting stellar early career scientists and staff—and these successes bring a new opportunity to expand our impact. By extending existing capabilities and combining them in new ways, we hope to make rapid progress toward solving essential challenges in energy, environment, and health. A significant impediment to progress has been the distribution of Biosciences' effort between five sites across the San Francisco Bay Area. Accordingly, an important element of this 10-year Strategic Plan is consolidation of all biosciences-related activities at a single site to enable increased interactions amongst Biosciences and the wider Berkeley Lab research community, communication, and the efficient development of new technological infrastructure.

While basic biological research is the foundation for future technologies, it also provides the evidence for informed policymaking on topics of critical importance. Results of biosciences research efforts can provide guidance for decision-makers who, for example, must assess the present and future impacts of climate change and other environmental challenges, evaluate new avenues to energy independence, and develop new medical technologies.

Our vision for biosciences at Berkeley Lab depends on more than merely setting scientific goals for the future. To achieve the 10-year Goals outlined in this plan, and to meet the needs of the nation, we must strengthen several supporting activities: the transfer of promising technologies to the private sector to create public benefit; our outreach efforts to develop an enhanced public understanding of the science that will provide solutions for our future; and the inspiration and education of a new generation of diverse scientists to recognize future national needs and achieve the objectives required to meet them. We must facilitate research at the intersections of diverse scientific disciplines to create environments that promote creative thinking, attract the brightest and most inquisitive scientists, and accelerate transformative discoveries. Furthermore, underpinning our efforts at Berkeley Lab is the leadership we demonstrate in developing new best-practices for ensuring that environmental/ecological, biosecurity, social justice, and ethical concerns are considered in advance of research and development. Our work spans security from the implications of human genomic sequence for privacy and prediction to the impact of that innovations in biomanufacturing may have on who controls production, how with downstream effects on volatility of production may be lessened, and the disruption of stability of manufacturing professions and the inception of others. We pride ourselves on building trust with the public through transparency and direct conversation.



BIOSCIENCES AT BERKELEY LAB

While Lawrence Berkeley National Laboratory may be best known for its physics, the biological sciences have been a part of its mission almost from the beginning, when founder and namesake Ernest O. Lawrence recruited top-flight scientists to UC Berkeley in the 1930's.

Lawrence's younger brother John, a physicist and physician, is considered the father of nuclear medicine. At Berkeley Lab, John studied the biological effects of the byproducts of the atom smashers Ernest built, and carried out the first successful treatment of human disease with radioisotopes. Today nuclear medicine still plays a central role in the diagnosis and treatment of cancer and other human diseases, and today's health-related scientists at Berkeley Lab are building on these foundations in their research efforts to better understand cancer, DNA repair, genome structure and function, and neurodegenerative diseases.

Biochemists Melvin Calvin and Andrew Benson used radioactive carbon-14 from a Berkeley Lab cyclotron to map the route that carbon travels through a plant during photosynthesis—research that led to discovery of the “Calvin cycle” and the Nobel Prize in Chemistry in 1961. Today's physical bioscientists and engineers at Berkeley Lab are building on advances in the physical sciences and modern biology, including those of Calvin and Benson, to examine, characterize, and mimic biological molecules and molecular functions to create unique biological structures that can then be used to solve some of the 21st century's most difficult fundamental research problems.

Berkeley Lab conducted path-breaking research on medical imaging, including early development of computed tomography (CT) scans and positron-emission tomography, (PET) scans. Cancer studies broadened to include tracking the behavior of healthy and malignant cells in culture and animals, pioneering the development of 3-D human tissue models, defining cancers as diseases of tissue microenvironments, and identifying many of the impacts of radiation on cells and organisms. Studies of heart disease and Alzheimer's disease helped to characterize the role of oxygen radicals in aging and disease. Bioscience research at Berkeley Lab deepened our understanding of what was becoming known as “systems biology.”

The extensive work in biological sciences and pioneering studies on mapping and sequencing the model organism *Drosophila melanogaster* genome led to selection of Berkeley Lab as one of five centers for the Human Genome Project, the massive national effort to map and sequence the entire complement of human DNA. Berkeley Lab's Human Genome Center, which was consolidated into the Department of Energy's (DOE) Joint Genome Institute (JGI) in Walnut Creek, was responsible for sequencing a significant portion of the human genome. Since that time, the JGI has undertaken a considerable effort to determine the genome sequences of thousands of plants and microorganisms with the aim of using this genomic information to develop solutions to national-scale energy and environment challenges.

Aided by new sequencing methods, faster computers and more advanced algorithms, studies of gene regulation intensified. Berkeley Lab played a major role in the Model Organism Encyclopedia of DNA Elements Project, which resulted in greatly improved genome annotations and an expanded understanding of non-protein coding RNAs, chromatin ‘landscapes’, and genome functions. Rapid sequencing renewed interest in proteins, including how they are structured and how they work. X-ray crystallography at the Advanced Light Source, plus a range of powerful microscopic techniques, revealed structures of important proteins at the highest resolutions ever.

The focus on genetics and molecular biology developed naturally toward the discipline now called synthetic biology, which holds the promise of reducing dramatically the costs and time required to design, build, and characterize biological systems. These innovations have led to focused applications and the creation of a number of spin-off companies.

Under the leadership of Laboratory Director Steven Chu, the Nobel laureate who would become President Barack Obama's Secretary of Energy in 2009, Berkeley Lab embarked on an intensive effort to use the tools of genetics, supercomputers, and microbiology to develop biofuels and new sources of sustainable energy. The Joint BioEnergy Institute (JBEI) is one of the national centers created by the DOE in 2007 and refunded in 2017 to advance the development of biofuels. Building on a legacy of advanced research in biosciences, Berkeley Lab has the infrastructure and expertise to bring biological solutions to the energy, health, and environmental challenges of our time as well as provide the foundational underpinnings for a strong biological manufacturing industry.

Since the publication of the Biosciences Strategic Plan in 2013, changes at Berkeley Lab have positioned Biosciences to build upon its successes. With the support of Berkeley Lab Director Michael Witherell, Biosciences research programs, leveraging insights in fundamental research, expanded further into applied research. Berkeley Lab will soon be the home of an Advanced Quantum Testbed for quantum computing. Berkeley Lab has also invested significantly, along with its DOE and private industry partners, to develop Cyclotron Road, an incubator for new energy technologies and the companies that are building them. Researchers from the Biosciences Area have successfully created companies that are now being nurtured in Cyclotron Road on their path to market.

In addition to our focus on using science to bring solutions to the world, our strategy also embraces a Berkeley Lab commitment to transferring our knowledge to our surrounding communities. We will continue to combine our research efforts with efforts to reach out to our neighbors. Through workshops, internships, and educational programs at local schools, colleges, and universities, we will promote understanding of science and encourage young people of diverse backgrounds to make a career in biosciences part of their own strategic plan.



STRATEGIC GOALS

Biosciences for Energy

Develop and enable cost-competitive, economically and environmentally sustainable biological or bio-inspired energy solutions.

Biosciences for the Environment

Understand the genetic and molecular mechanisms governing the activities and ecology of organisms and multispecies communities, predict functions and interactions across scales, and harness microbes and plants for energy and environmental solutions.

Biosciences for Health

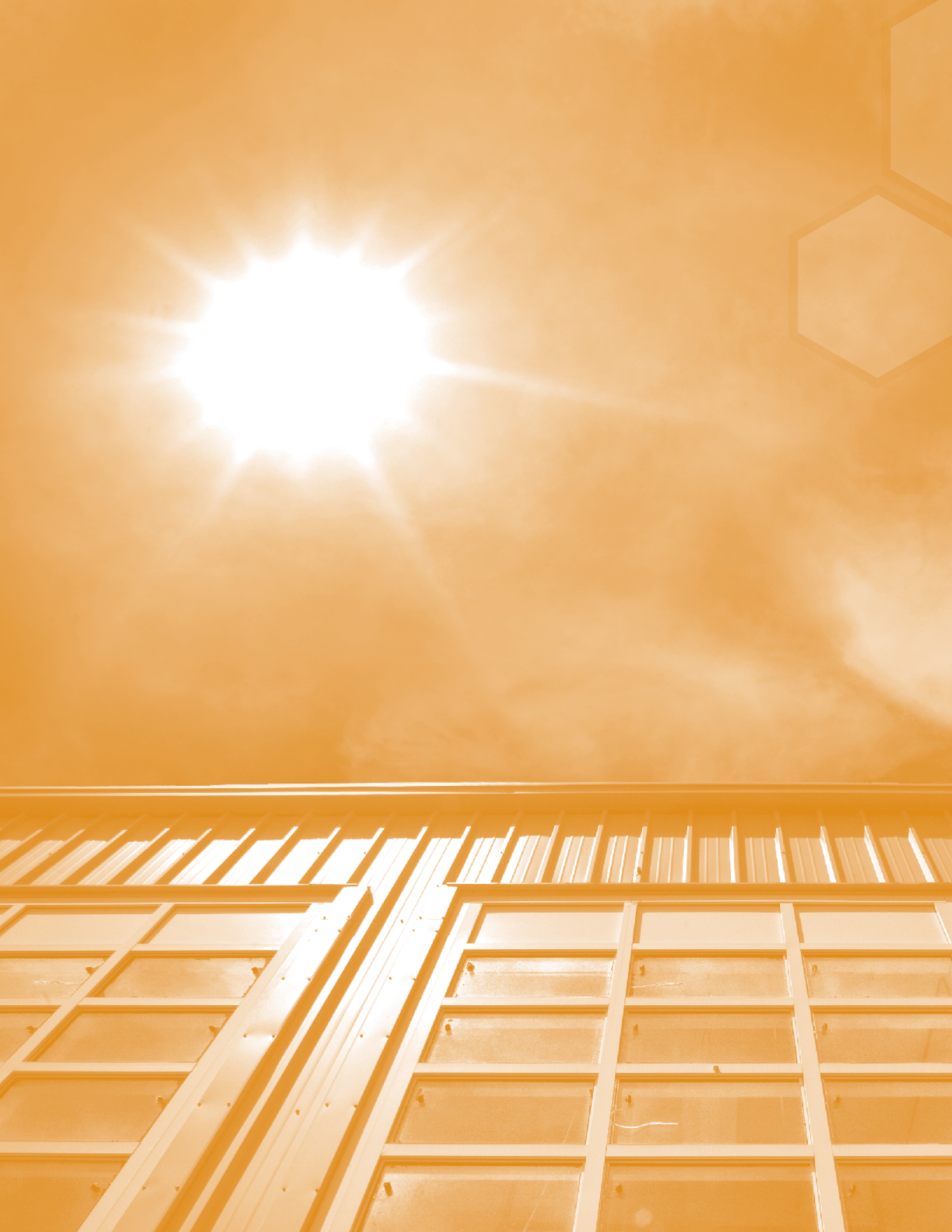
Develop and apply a predictive, multiscale, integrative understanding of biological systems to improve the bioresilience of humans and our ecosystems, including how biological variation affects individual responses to environmental challenges.

Biosciences for Biomanufacturing

Develop and demonstrate accessible, scalable, flexible, cost-effective, and sustainable biology-based manufacturing infrastructure and expertise driven by applications in energy, health, materials, environment, and agriculture.

Technologies for Biosciences

Develop technologies infrastructure to measure, understand, predict, design, engineer, and control biological systems for solving energy, environmental, health, and biomanufacturing challenges.





ENERGY

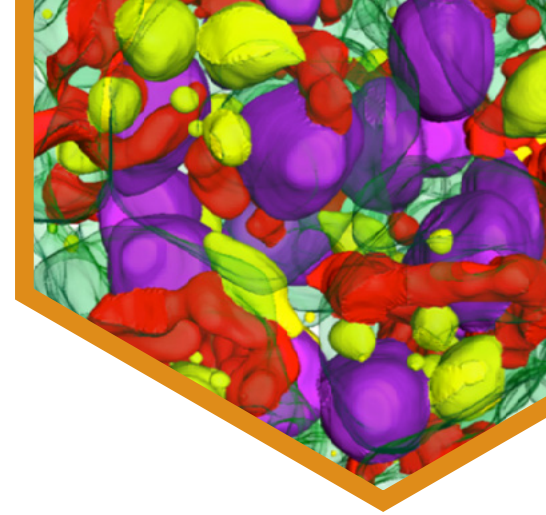
10-year Goal: Develop and enable cost-competitive, economically and environmentally sustainable biological or bio-inspired energy solutions.

Background and Motivation

Energy production is the world's largest industry and is based almost exclusively on fossil fuels, whose extraction and subsequent burning pollute land, water, and the atmosphere. As "clean" sources of fossil fuels are depleted, less-desirable sources are being tapped that require significantly more energy to extract and refine and may pollute the environment in other ways.

The development of alternative energy sources is a pressing national need, a major mission of DOE's Office of Science, and a central thrust of Berkeley Lab's strategic plan since 2006. This effort now involves all of Berkeley Lab's Biosciences Divisions and receives more than \$60 million in annual funding. Successful production of scalable bio-based sources of transportation fuel will be most rapidly achieved through an integrated, team-science approach that has been the hallmark of the Laboratory since its inception.

Biology has the potential to produce energy renewably, particularly liquid hydrocarbon fuels with the high energy density needed by the U.S. transportation infrastructure. However, biological mechanisms are relatively inefficient at harnessing sunlight and transforming it to hydrocarbons. Converting the full gamut of the lignocellulosic and non-lignocellulosic material to high-density fuels and replacements for petrochemicals is also necessary to make a meaningful difference in the current paradigm of bioenergy and biomanufacturing. A better understanding of photosynthesis, cell-wall synthesis in plants, and the hydrocarbon-forming pathways in all organisms would make it possible to build predictive models of these processes. These models could be used to engineer plants



10-YEAR Energy Goal

Develop and enable cost-competitive, economically and environmentally sustainable biological or bio-inspired energy solutions.

Energy Research Strategies to Achieve Goal

Lignocellulosic biofuels. Derive fuels and coproducts from biomass with new technologies.

Alternative biofuels. Directly convert one carbon (C1) feedstocks to fuels using microorganisms.

Artificial and engineered photosynthesis. Use bio-inspired reactions to create fuels directly from atmospheric CO₂ and sunlight.



to capture sunlight more efficiently and use *in situ* nutrient resources more efficiently while strengthening their resilience to otherwise harsh environments, as well as to alter sugars accumulation in plant cell walls and enhance their breakdown and usage. It will also empower the engineering of microorganisms to convert sugars, aromatics, and other carbon sources more efficiently into drop-in biofuels and fuel additives compatible with the transportation fleets of today and tomorrow. These advances are also critical for broad biomanufacturing by expanding the opportunities for designed biological systems. In addition, these efforts could facilitate design of bio-inspired catalysts that mimic photosynthesis to produce transportation fuels directly from sunlight and CO₂.

The Bioscience Area's approach to this problem is to develop a molecular description of the biological processes of photosynthesis, cell-wall synthesis, biomass breakdown, microbial metabolism and hydrocarbon biosynthesis. This new knowledge can subsequently be used forward and as a feedback loop to meet key strategic objectives in the areas of lignocellulosic biofuels production, alternative biofuels development, and both natural and artificial photosynthesis.

Energy Research Strategies

- **Lignocellulosic biofuels.** Derive fuels and coproducts from biomass with new technologies.
- **Alternative biofuels.** Directly convert one carbon (C1) feedstocks to fuels using microorganisms.
- **Artificial and engineered photosynthesis.** Use bio-inspired reactions to create fuels directly from atmospheric CO₂ and sunlight.

To achieve the Energy Goal by 2023, Berkeley Lab's approach employs three areas of strategic focus: production of fuels from plant biomass (lignocellulosic biofuels), production of fuels from gaseous C1 feedstocks (alternative biofuels), and non-biological production of liquid fuels directly from sunlight and CO₂ (artificial photosynthesis). These areas were chosen because (1) they collectively have potential to meet the long-term national need for sustainable cost-competitive alternatives to fossil fuels; (2) they are scientifically tractable within a 10-year span; and (3) they leverage specific facilities, organized research groups and core competencies that exist, or can be readily assembled, within the Biosciences Area at Berkeley Lab. These strategies will be executed in parallel.

Lignocellulosic biofuels: Derive fuels and coproducts from biomass with new technologies

The cellulosic biofuels strategy outlined by DOE in the 2005 *Billion Ton Study*, the 2011 *U.S. Billion-Ton Update*, and the 2016 *Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy* highlighted the amount of feedstocks required for the development of cellulosic biofuels' capacity to replace 30% of U.S. needs for transportation fuels, without significant negative impacts on human food and livestock feed production. The success of this strategy depends on:

- Improved biomass feedstocks with greater yields of convertible sugars and valuable lignin intermediates, greater tolerance to stress, and improved nutrient acquisition

aimed at reducing energy input and pollution and to allow for wider land use and less volatile production due to environmental variation

- Greatly improved biomass extraction and breakdown strategies
- Engineering of microorganisms capable of converting biomass sugars and lignin aromatics to high-energy-density hydrocarbon fuels and additives compatible with the current and future transportation fleet, including jet aircrafts and
- The development and demonstration of economically viable and scalable production processes for an advanced biofuel and co-product, including feedstock deconstruction, fermentation, chemical catalysis, and product recovery and purification.

Efforts on these fronts are underway at Berkeley Lab, primarily funded through JBEI and individual research programs funded by Biological and Environmental Research in the Office of Science at DOE, but supported also through collaborative interactions among JBEI, JGI, and the DOE Systems Biology Knowledgebase (KBase), and particularly with the Advanced Biofuels and Bioproducts Process Development Unit (ABPDU) funded by the DOE Office of Energy Efficiency and Renewable Energy.

JBEI focuses on the development of advanced bioenergy plants, efficient chemical and biological processes to extract sugars and lignin aromatics from these plants, and efficient biological conversion of the resulting sugars to advanced biofuels and co-products. This team of nearly 200 scientists has already engineered bioenergy crop plants with more cellulose and reduced or altered lignin; developed deconstruction processes based on ionic liquids to deliver clean cellulose and hemicellulose that can be more readily depolymerized into sugars using fewer enzymes than other deconstruction processes; and engineered microorganisms to produce drop-in biofuels for gasoline, diesel, and jet engines. In this effort, the broad retrosynthesis infrastructure, a key element in the



tools development for Biomanufacturing, would allow the improved prediction of potential pathways to a hydrocarbon product, while designed biological systems allow development of the corresponding plants and microbes. Current efforts are also directed at utilizing lignin aromatics and producing co-products in addition to fuels, thereby rendering the downstream biorefinery process more cost-competitive. Though co-products research is highly integrated with biofuels research, these aspects are further described in the Biomanufacturing section and are therefore not covered in this section even though co-products research is highly integrated with biofuels research.

The JGI's efforts have effectively complemented those of JBEI. The JGI has produced more de novo plant, fungal, and bacterial genomes than any other individual genome-sequencing center. Many of these genomes have direct relevance to Berkeley Lab's lignocellulosic biofuels strategy, providing the bases for molecular understandings of relevant pathways, and sources of molecular tools for engineering strategies. In addition, JGI is a world leader in the field of metagenomics, pioneering microbial genome assembly and annotation from complex microbial environmental communities. It has leveraged skills in this arena to mine biomass-degrading activities from unique environments ranging from termite gut to hoatzin crop, and from Yellowstone hot springs to tropical rainforests, which offer new avenues for biofuels investigations. Further, JGI has pioneered large-scale DNA synthesis and construct design in the service of characterizing these enzymatic activities and enabling manufacture of fuel production pathways under their new synthetic biology efforts.

Much remains to be done. Proof-of-concept studies in cell-wall modification, advanced biomass pretreatment (e.g., ionic liquids) and fractionation must be developed and demonstrated, cell-wall-degrading enzyme discovery must be extended, and processes developed and scaled for licensing to the cellulosic-biofuels industry. Finally, there remain several inefficiencies in converting sugars and lignin-derived aromatics to drop-in biofuels. Process efficiency research and scaling are being addressed in the bench- to pre- pilot scale development and demonstration environment at the ABPDU in collaboration with JBEI, industry partners, and other national labs.

Alternative biofuels: Directly convert one carbon (C1) feedstocks to fuels using microorganisms

The goal of the lignocellulosic biofuels strategy is to produce fuels from depolymerized lignin and sugars derived from plant cell walls. However, sugars and lignin intermediates derived from lignocellulosic biomass are complex and often contain contaminants that in the past have prevented them from being efficiently converted by microorganisms. Processes that use C1 sources greatly expand the repertoire of starting materials for biofuels and bio-based products, especially in areas of the country where these sources are more abundant than lignocellulosic biomass. Therefore, in addition to feeding microorganisms sugars and aromatics derived from plant biomass to create fuels, Berkeley Lab is also focused on another set of pathways: the direct microbial conversion of C1 feedstocks such as abundantly available CO_2 and its reduced derivatives (HCO_2^- , CH_3OH), synthesis gas ($\text{CO}/\text{CO}_2/\text{H}_2$), and waste gases such as H_2S and biogas methane, to fuels by engineered photosynthetic, methanotrophic, or chemoautotrophic microorganisms. The potential for fuel production by direct conversion has long been conceptually attractive, but major unsolved challenges continue to impede this effort. Microbial conversion of these substrates is ubiquitous in nature, but efforts to exploit species that might be useful for this purpose have found limited success. Efforts to engineer these organisms are hampered by a dearth of required genetic tools with which

to do the engineering and a lack of basic understanding of the metabolic pathways involved and how they are regulated. Tools for a range of new host microorganisms are featured in Berkeley Lab's Biomanufacturing strategies, which along with a systems-level understanding are needed to carry out the large-scale pathway engineering for the development of new production strains. In pursuing this strategy, Berkeley Lab leverages capabilities in the Biosciences Area, both existing and those being advanced in the Technology objectives, to obtain a systems-level understanding of these microbes. Berkeley Lab also leverages JBEI's significant experience with microbial fuel production from sugar substrates to transfer these pathways to microorganisms that generate fuels from C1 feedstocks. The ABPDU is developing bioreactors and upstream and downstream processes internally and in partnership with JBEI, industry, and other DOE laboratories to scale fuel production from C1 feedstocks. Synergistic JGI missions such as genomics of biofuel feedstocks and discovery strategies for fungal/microbial enzyme are valuable to this effort, as are JGI's major data resources (e.g. Integrated Microbial Genomes & Microbiomes system) that can be mined to achieve these objectives. The Biomanufacturing effort provides genome editing and strain engineering strategies as well as retrosynthesis infrastructure, that are highly synergistic with these objectives and are critically important to the success of this strategy.

Artificial and engineered photosynthesis: Use bio-inspired reactions to create fuels directly from atmospheric CO₂ and sunlight

Berkeley Lab's Melvin Calvin and Andrew Benson mapped the route that carbon travels through a plant during photosynthesis—research that led to discovery of the “Calvin cycle” and the Nobel Prize in Chemistry in 1961. Once achieved and scaled up, artificial photosynthesis, a chemical process that replicates the natural process of photosynthesis, could be significantly more efficient than biological fuel production processes, and would not require arable land, agricultural feedstock, or substantial inputs of energy or water. Over the past 50 years, basic research has steadily increased our understanding of the subtle and complex mechanisms behind natural photosynthetic systems as well as in the use of photochemical methods that mimic key steps in the process: splitting water and reducing carbon dioxide. However, significant impediments, such as the inability to control chemical reactions on a nanometer scale and the lack of a deep understanding of photosynthesis on temporal and spatial scales, have prevented our ability to design solar-energy-to-fuel conversion systems with the required efficiency, scalability, and sustainability to be economically viable.

In the past 20 years, nanotechnology—the making and manipulation of matter on nanometer scales—has advanced dramatically, and with it, the ability to control and optimize the intricate processes of artificial photosynthesis. New nanotechnologies allow Berkeley Lab researchers to work at the scale of atoms and molecules, to synthesize new catalysts to accelerate relevant chemical reactions, and to assemble molecular and nanostructured components into scalable hierarchical assemblies for the efficient conversion of carbon dioxide and water to solar fuels. Nanotechnology of inorganic materials, together with bioengineering technology that enables the manipulation of biological systems at the genomic level, makes it possible to integrate living organisms with inorganic catalysis into nanoscale biohybrids. Additionally, we can take advantage of the complementary strengths of abiotic and biotic systems—a robustness of inorganic systems and the complexity and nuanced function of biological systems—in energy harvesting and chemical production.



Berkeley Lab has committed to this strategy and is uniquely positioned to address it with a wide variety of assets. In addition to its historical and current scientific expertise, Berkeley Lab has unequalled state-of-the-art facilities relevant for studies of artificial photosynthesis. With its Molecular Foundry, a nanoscience facility, Berkeley Lab has become an epicenter for the synthesis of novel nanomaterials and advanced characterization, drawing scientists from around the world to its world-class instruments, materials, technical expertise, and training. Here again, the high-throughput DNA sequencing, synthesis and analysis provided by the JGI strengthen the foundation of essential synthetic biology aspects of the biohybrid goals. In addition, Berkeley Lab's ALS provides state-of-the-art macromolecular crystallography facilities and a planned national cryo-EM facility which are used to determine structural features of key proteins and complexes in these systems. Other proximal light sources like SLAC National Accelerator Laboratory's Linac Coherent Light Source, and laboratory-based unique transient spectroscopies are being used to study natural photosynthesis with the goal of translating discoveries of natural photosynthetic processes into bio-inspired design principles for artificial photosynthesis. These facilities, and technologies emerging from them are critical for monitoring photocatalytic processes in action, providing insights that rapidly lead to improved artificial photosystem designs. Additional knowledge for artificial photosynthesis will be gained through research conducted at Berkeley Lab's National Center for Electron Microscopy which provides cutting-edge instrumentation, techniques, and expertise for advanced electron beam mesoscale characterization of materials at high spatial resolution. The new envisioned bioimaging center will allow new insight to be gained into photosynthetic processes through multimodal imaging. The computing power of the National Energy Research Scientific Computing (NERSC) Center at Berkeley Lab plays an important role in guiding materials selection for optimal systems performance. At the Laboratory's Joint Center for Artificial Photosynthesis (JCAP), the DOE Energy Innovation Hub for Fuels from Sunlight (a partnership with Caltech), a dedicated team of scientists and engineers is developing a solar-powered technology at the testbed level for converting carbon dioxide into liquid transportation fuel.

2023 10-year Goal achievement measured by:

Lignocellulosic biofuels

- Understand fundamental elements of plant biology that underlie biomass yields and adaptation to stress by focusing on key gaps in knowledge of primary plant physiology and in plant-microbe interaction relevant to plant resilience
 - Elucidate how the secondary cell wall of bioenergy crops is synthesized
 - Understand mechanisms of water conservation and drought tolerance in plants
 - Advance the understanding of how nitrogen fixation occurs in model plants
 - Understand the molecular interactions between plants and beneficial microbes and their role in nutrient acquisition and stress tolerance
 - Engineer bioenergy crops for reduced inputs (water and nitrogen/phosphate fertilizers), enhanced tolerance to stress, improved sugar yields, and facilitated production of useful compounds from lignin, including fuels
 - Engineer plants and microbes to stimulate nitrogen fixation
 - Engineer model plants and crops to grow with less water
 - Engineer lignin in biomass crops to be an economically useful polymer
 - Increase the C6 to C5 sugar ratio
- Develop predictive models that will facilitate engineering of optimized secondary cell wall synthesis and saccharification by specific genetic manipulations
- Develop deconstruction processes that enable efficient utilization of all plant biomass components (cellulose, hemicellulose, lignin)
 - Fractionate biomass so the individual components can be separated into different streams
 - Develop pretreatment technologies that allow for direct conversion of all the plant polymer components in one bioreactor
- Develop inexpensive methods to depolymerize plant polymers to intermediates for microbial conversion
 - Develop a suite of enzymatic cocktails that depolymerize cellulose, hemicellulose, and lignin without product inhibition and inhibition from contaminants resulting from the deconstruction process
 - Discover chemical catalysts that selectively depolymerize plant polymers to monomers that can be converted to fuels by fermentation
- Understand the mechanisms of lignocellulose deconstruction by microbial communities
 - Probe the variety of biomass degradation strategies employed by microbes in natural environments
 - Establish a synthetic microbial community that deconstructs one lignocellulosic substrate
 - Understand synergies between fungi and bacteria in a mixed eukaryote-prokaryote microbial community
 - Define the pathways for deconstruction and metabolism of lignin in one synthetic or adapted microbial community
- Engineer microorganisms tailored for consumption of deconstructed biomass (sugars and aromatic compounds) and production of drop-in biofuels at high yield

- Develop metabolic pathways for production of hydrocarbons with fuel properties equivalent to those found in petroleum-based gasoline, diesel, and jet fuels, and engineer these pathways into diverse hosts
 - Use metabolic engineering to generate microorganisms that rapidly convert multiple substrates to desired end products simultaneously
 - Develop machine learning approaches to optimize metabolic flux in engineered organisms
 - Develop genetic tools, including biosensors, and machine learning algorithms to enable multi-gene engineering in diverse hosts to rapidly engineer complex genetic traits such as fuel yield and stress tolerance
 - Engineer microbes to tolerate or detoxify biomass deconstruction inhibitors and fuel products, and other stresses introduced during bioreactor scale-up
 - Engineer microorganisms to produce fuels under anaerobic conditions
 - Engineer fuel-producing microorganisms whose communities are resistant to invasion by contaminating organisms and that are genetically-constrained to growth in industrially defined conditions
- Develop and demonstrate bench- and pilot-scale processes for fuel production
 - Understand gene expression and metabolism at different scales of production to enable efficient and predictive scale-up of manufacturing processes
 - Identify, implement and prototype new unit operations for more flexible and efficient production of precursor biomolecules and finished fuel derivatives
- Develop predictive models and computational methods to quantify the potential cost, net energy, greenhouse gas emission, and water impacts of research advancements in the production of feedstocks, biofuels and coproducts
 - Model the shifts in biomass yield, land requirements, and fuel yield resulting from new engineered feedstocks
 - Identify key drivers of feedstock and biorefinery costs, emissions, and resource requirements to guide research towards more sustainable pathways

Alternative biofuels

- Use a systems-level understanding to identify key bottlenecks that will limit fuel production in the metabolism of photosynthetic, methanotrophic and chemosynthetic microbes
 - Identify key and structurally characterize key proteins involved in the metabolic pathways essential for conversion of gas feedstocks
 - Identify key regulatory elements that sense and respond to the presence of gas feedstocks in potential fuel production hosts
 - Develop computational methods to predict the metabolism of C1-converting microbes
 - Combine computational predictions and high-throughput automation to develop and predict beneficial changes in metabolism using machine learning
- Engineer photosynthetic, methanotrophic and chemosynthetic microbes to produce fuels from gas feedstocks
 - Develop a synthetic biology toolbox leveraging broader methods being developed in the Biomanufacturing strategies (transformation, promoters, genome integration of heterologous pathways), for at least three C1 conversion hosts

- Produce energy-dense biofuels from gas feedstocks in a photoautotrophic, methanotrophic or chemoautotrophic host
- Improve the total carbon conversion in host microbe C1 feedstock use by 10% which will approach theoretical maximums for substrate conversion
- Scale fuel production from gas feedstocks through bioreactor and process development
 - Demonstrate production of an energy-dense fuel from a gas feedstock at pre-pilot scale
 - Design bioreactors in collaboration with Berkeley Lab researchers that maximize C1 conversion to fuel by improving mass transfer
 - Integrate product separation with C1 conversion to improve fuel yield

Artificial and engineered photosynthesis

- Develop scalable artificial photosynthetic systems based on bioinspired engineered and biohybrid approaches that produce energy dense liquid fuels from light energy and CO₂, well beyond CO or formate molecules currently attainable
- Couple predictive models with advanced nanoscale engineering to improve artificial photosynthesis
- Improve the biomass productivity of natural photosynthetic systems
- Use new levels of understanding of photosynthesis to predict ways it can be improved in plants and microbes
- Redeploy the photosynthetic apparatus in a previously non-photosynthetic host with superior coverage of the solar spectrum, improved conversion efficiency to fuel products, and durability
- Understand the fundamentals (e.g., excited state and charge transfer dynamics) that govern the multidimensionally (time and space) controlled chemistry in photosynthetic enzymes and artificial systems in conjunction with the development and application of advanced characterization methods. That includes using time-resolved X-ray crystallography and X-ray spectroscopy at X-ray free electron lasers, and multidimensional X-ray spectroscopy at the synchrotron and X-ray free electron laser facilities, in combination with time-resolved 2D electronic and vibrational spectroscopies

2018—Five-year milestones for energy research strategies

Lignocellulosic biofuels

- Cell-wall biosynthesis and assembly elucidated through identification of new genes, alleles, and metabolic pathways controlling cell-wall recalcitrance, sugar and lignin content, and fermentation inhibitors
 - Secondary cell walls engineered to have more C6 sugars and fewer C5 sugars
 - Cell walls engineered to have easily cleavable or altered lignin
 - Biomass traits have been engineered from model plant systems to potential bioenergy crops

- Tools developed to determine metabolite levels and metabolic bottlenecks in plants
- Knowledge of drought tolerance in model crops advanced through identification of new genes, alleles and metabolic pathways
- New pretreatment methods developed that reduce cost and efficiently fractionate lignocellulose into targeted lignin and sugar output streams
- Enzymes discovered and developed for optimal performance under pretreatment/ saccharification conditions (temperature, pressure, presence of ionic liquids)
- New enzymes and cofactors discovered and engineered for biomass deconstruction that are tolerant of pretreatment regimens
 - Efficient enzymes capable of depolymerizing lignin into aromatic hydrocarbons discovered to create a strong value stream for lignin
 - Models developed to predict how modifications to secondary cell-wall biosynthesis and degradation to improve biomass yields developed
 - Predictive models to describe release of sugars from the plant secondary cell wall developed
 - Hydrocarbon biosynthetic pathways and associated transporters engineered into microbes to convert sugars to transportation fuels
- Native hydrocarbon biosynthetic pathways in plants and microbes described
- Predictive models to describe metabolic fluxes developed and used to predict bottlenecks in biosynthetic pathways in microorganisms
- New bench- and pilot-scale unit operations and integrated processes developed and demonstrated for the production of a lignocellulosic biofuel

Alternative biofuels

- Developed and implemented gas-fed bioreactor processes and associated systems and upstream and downstream unit operations for fuel production to enable demonstration at bench and pilot scale
- Identified essential genes required for growth and CO₂ metabolism in a photoautotroph
- Developed synthetic biology toolbox related to hydrocarbon biofuels production in a chemoautotrophic host
- Converted synthesis gas to hydrocarbon biofuels under anaerobic conditions

Artificial and engineered photosynthesis

- Developed an advanced mechanistic understanding of photosynthesis in plants and microbes
 - Identified the determinants of efficient photosynthesis in plants, algae, and bacteria
 - Defined the repair mechanisms of the photosynthetic apparatus
- Explored options for reaching mA/cm² currents between artificial systems and cellular organisms

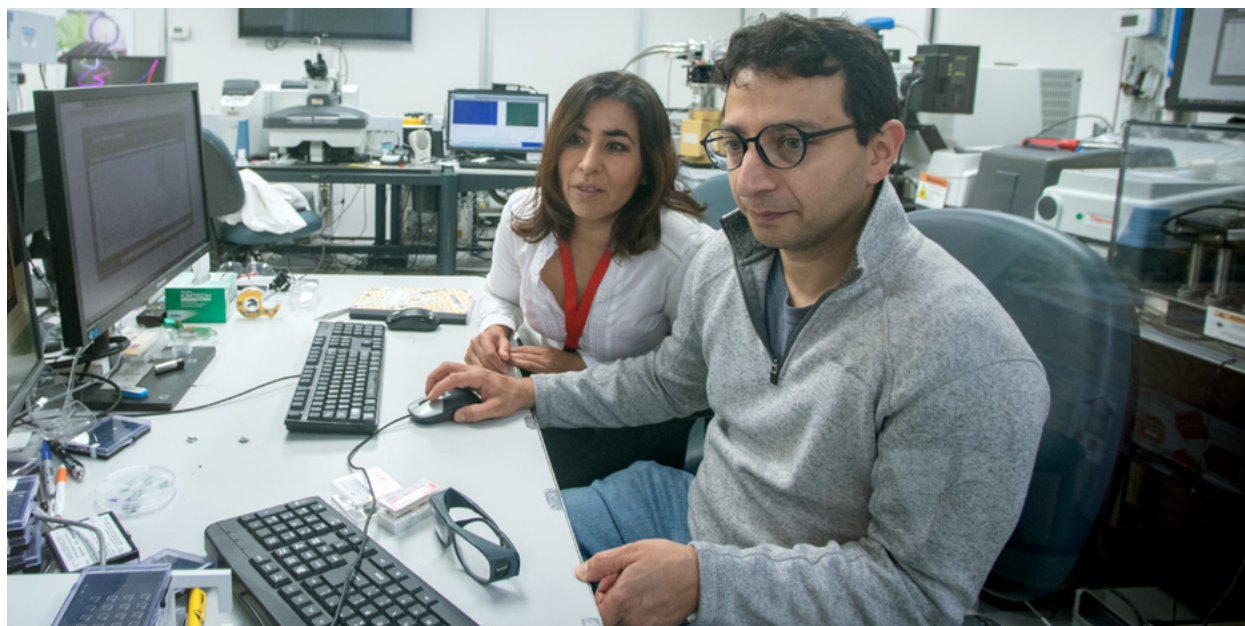
- Synthesized membranes capable of separating carbon-based fuels from oxygen
- Demonstrated capability for photocatalyzing conversion of CO₂ to a carbon-based fuel beyond CO and formic acid
- Designed the first prototype devices for testing components (catalysts, light harvesters, membranes, interfaces, etc.) as an integrated system
- Demonstrated photosystem for unassisted CO₂ reduction by H₂O under membrane separation on the nanoscale
- Developed a mechanistic understanding of light driven H₂O oxidation on robust Earth abundant catalyst
- Performed analysis of components, materials and chemical inputs, and hardware designs to provide information on manufacturability, life-cycle costs, and reusability to ensure the system's scalability

A look ahead to 2028

Lignocellulosic biofuels

How do we optimize the synergistic production of biofuels and bioproducts from lignocellulosic biomass?

- Develop new pathways for bioproduct synthesis in plants and industrial microbes
- Understand and improve bioenergy crop resilience to environmental stress
- Combine chemical and biological conversion techniques to expand the types of molecules that can be produced



- Develop efficient, economical and sustainable separations technologies for integrated biorefineries
- Integrate machine learning methods in development of predictable bioengineering and process optimization

Alternative biofuels

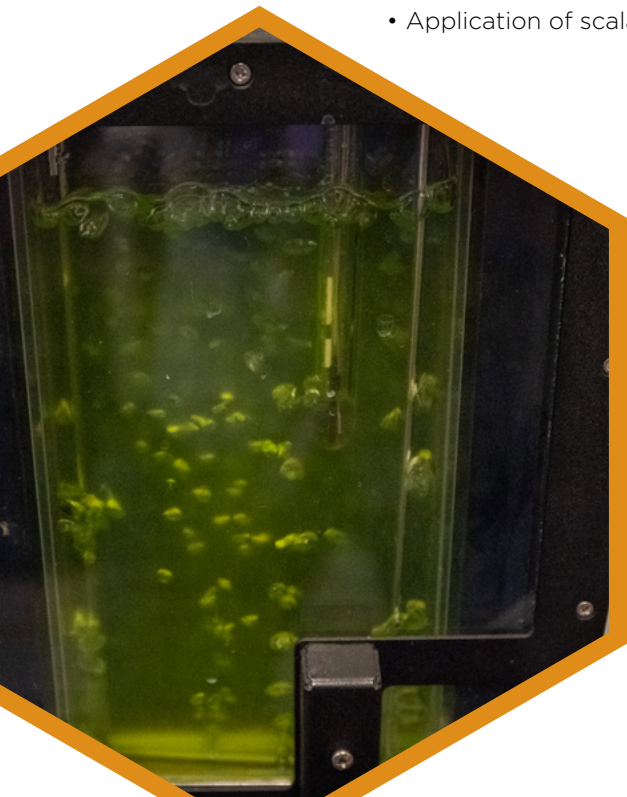
How can we deploy microbial C1 metabolism to use CO₂ at atmospheric concentrations to draw down CO₂ levels directly?

- Develop modular biological CO₂ concentrating mechanisms
- Engineer modified or fully synthetic CO₂ metabolic pathways that are more efficient than native pathways
- Integrate CO₂ conversion with other renewable sources of carbon
- Design bioreactors that work with air as a carbon and energy source

Artificial and engineered photosynthesis

How can we leverage nanotechnology and new knowledge about photosynthesis to deploy new natural and artificial photosynthetic systems with improved usage of light and CO₂?

- Design and utilize artificial photosynthetic devices based on the design concept learned from natural photosynthesis
- Production of liquid fuels from CO₂ using artificial systems
- Develop light-driven nitrogen fixation systems
- Application of scalable biohybrid-systems for producing fuels from sunlight

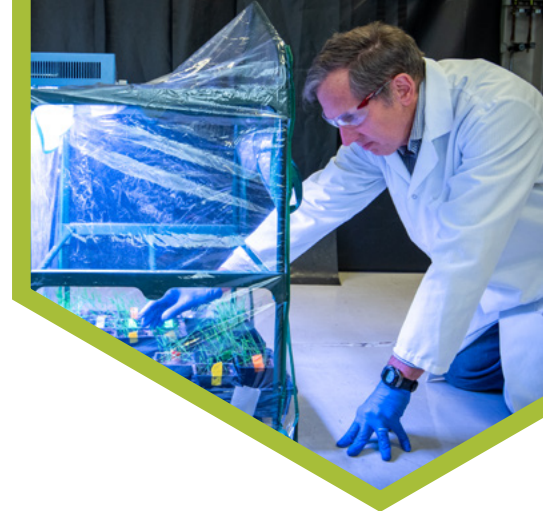








ENVIRONMENT



10-year Goal: Understand the genetic and molecular mechanisms governing the activities and ecology of organisms and multispecies communities, predict functions and interactions across scales, and harness microbes and plants for energy and environmental solutions.

Background and Motivation

Globally, we face a confluence of environmental and societal challenges that make it difficult to provide resources for a rapidly growing population, while safeguarding vital environmental ecosystem processes. To address such 21st century challenges, it is imperative that we drastically improve our mechanistic understanding of environmental organisms and their ecology across dynamic and changing environments. This approach will provide foundational insights into how microbes, plants and other organisms interact to process nutrients and contaminants that will enable us to predict how soil processes and water quality changes with anthropomorphic and natural inputs. These insights will provide the foundation for developing urgently needed mechanistic approaches for management of vital nutrient cycles and harnessing beneficial plant-microbe-metazoan interactions for improved bioenergy and food crops with increased productivity, enhanced drought resistance, and lower inputs (energy, fertilizer, pesticides).

The tremendous advances in DNA/genome sequencing now enables inexpensive and rapid sequencing of prokaryote and eukaryote genomes and communities from environmental samples. As a result, we have, and continue to accumulate, vast amounts of sequence information from all types of organisms and environments. Functional genomic technologies including gene expression, epigenetic profiling, large-scale protein and metabolite measurements and powerful imaging modalities are providing unprecedented views into the

10-YEAR Environment Goal

Understand the genetic and molecular mechanisms governing the activities and ecology of organisms and multispecies communities, predict functions and interactions across scales, and harness microbes and plants for energy and environmental solutions.

Environment Research Strategies to Achieve Goal

Predictive understanding of environmental organisms. Discover and deeply characterize the genetic and molecular mechanisms of environmental organisms that drive and respond to environmental changes.

Molecular ecosystems biology-based solutions. Using biological and environmental characterization of natural and laboratory ecosystems to understand native ecosystem processes, predict responses, and harness plants and microbes for energy and environmental solutions.



activities encoded in organismal and community genomes. Together with sequencing and environmental characterization, these advanced technologies have the potential to provide the scientific foundation to accurately predict the functions of genes, organisms and whole communities. This approach will transform ecosystem science and enable the development of new theories and models to predict microbial community nutrient cycling and environmental responses and will enable harnessing organisms to promote beneficial outcomes. However, we are far from a predictive level of understanding of these ecosystems. Gene function is unknown for about half of the genes in the genomes of most environmental microbes. Even less is understood about how to scale from the genes in a genome to the phenotypes and activities of an organism and then, in turn, to move from an individual organism and its genome to the highly interdependent communities of organisms.

Deploying advanced sequencing, functional genomic and imaging technologies to study microbe-plant-animal communities in realistic environments is the best way to learn about the function of their genomes. For example, a large fraction of the poorly annotated genes in environmental organisms are believed to function only under specific environmental conditions or in the context of other species with which they interact. Hence, discovering the function of such genes/genomes will require studying them in an ecological context. However, most studies to date have been performed either on isolated organisms or fully complex environmental communities, both of which have major challenges. The same is likely true for a large number of plant genes, which are presumably dedicated to selecting and maintaining beneficial microbiomes (i.e. interconnected communities of viruses, bacteria, and fungi living in close association with plant hosts or in other environments), protecting from pathogens, response to abiotic and biotic stress and enabling nutrient acquisition. Yet, because of the extreme complexity and undefined nature of natural ecosystems, it is very difficult to determine the functions of specific genes/genomes and ecological interactions under field conditions.

To address these challenges, the Berkeley Lab is pioneering the development of laboratory ecosystems and advanced technologies for characterization and genetic, environmental and biochemical manipulation of diverse environmental organisms within these environments. This approach provides unprecedented control of and insight into environmental biological processes. The results obtained will feed directly into developing new theory, data scientific and modeling tools to enable predictive approaches on biomolecular, organismal and community functions, activities and dynamics.

Central to these efforts are our rapidly expanding abilities to cultivate, characterize and manipulate diverse organisms and communities within controlled environments. The model fabricated ecosystems from single plant scale 'EcoFABs' to mesocosm scale 'EcoPODs' are being developed based on field studies of native ecosystems. These provide controlled environments needed for genetic, synthetic biology, functional genomic, and imaging approaches to determine the roles of specific organisms, genes, metabolites, and environmental factors mediating activities important for ecological attributes. This will enable both systems biology-based discoveries and reduction to causal mechanisms for development of predictive models and new ecological theories that can be tested and refined under laboratory and native ecosystem conditions. The ultimate goal is to enable the development of effective approaches for environmental clean-up, balancing nutrient cycles, and harnessing plants and beneficial microbiomes to build soil carbon and increase agricultural and ecosystem productivity.

Berkeley Lab is uniquely suited to address these grand environmental challenges as a result of our extensive expertise and one-of-a-kind capabilities. In particular, the DOE-funded ENIGMA Scientific Focus Area (SFA) has a long history of groundbreaking work in advancing fundamental scientific understanding of genomics of critical environmental organisms and microbial communities that impact contaminated field sites and nutrient cycling. This SFA has developed state-of-the-art functional genomics, high-throughput genetics, systems biology, mass spectrometry, bioimaging, bioinformatics, computational modeling, and molecular environmental microbiology approaches that greatly enhance our ability to achieve Biosciences' Environment strategies. The Environmental Genomics and Systems Biology Division (EGSB) brings together experts in microbe, plant, and metazoan biology; genomics; biotechnology; data science; and modeling to address these critical challenges in energy and the environment.

Biosciences' Environment strategies are highly integrated with Berkeley Lab assets, such as the JGI, an integrative collaborative Genome Science User Facility. The JGI provides advanced sequencing, DNA synthesis, and computational analysis capabilities to the energy and environmental genomics communities. Examples of its activities include the sequencing of DOE mission-relevant genomes, the development and application of advanced "sequence-to-function" capabilities, the sequencing and decoding of the unexplored "dark matter" of microbial genomes, and the synthesis and functional characterization of large DNA constructs to better understand microbial and plant secondary metabolism and plant-microbe interactions.

The Environment strategies will also leverage the Biosciences-led Agile BioFoundry (ABF), the recently established multi-national lab effort that is aimed at democratizing and accelerating the engineering of biology for desired purposes. Of specific relevance is the ABF's potential to provide synthetic biology tools that allow rapid characterization and manipulation of biological activities through multi-modal measurements, machine learning, and artificial intelligence. For example, ABF capabilities will enable the synthesis, assembly, and expression of putative secondary metabolite biosynthetic pathways to discover their effects on microbial and plant ecology.

High-resolution multi-scale imaging that is used to examine microbe-microbe and plant-microbe interactions effectively leverages the cutting-edge imaging capabilities of the ALS and Molecular Foundry. Sensor development across Berkeley Lab will go hand-in-hand with the advancement of these model ecosystems. The supercomputing resources required to analyze and model complex systems biology and imaging datasets are available at the National Energy Research Scientific Computing Center (NERSC). KBase provides critical data and computing resources to enable researchers to collaboratively generate, test, and share new hypotheses about gene and protein functions and to build and share predictive systems models of microbes and microbial communities, and their interactions with plants and other biotic and abiotic components of ecosystems. Additionally, the National Microbiome Data Collaborative (NMDC) is building sustainable computational infrastructure to facilitate transparent and reproducible environmental microbiome research.

Berkeley Lab's Earth and Environmental Sciences Area (EESA) is an important partner in this effort with complementary expertise in environmental characterization, remote sensing, hydrology, geophysics, isotope chemistry, environmental microbiology, ecology, biogeochemistry and multi-scale modeling. Hence, EESA, in collaboration with Biosciences scientists working to achieve the Environment Goal will play a key role in developing laboratory ecosystems that are deeply connected to and accurately

recapitulate key aspects of native ecosystems. The Bioscience's strategies of incorporating and integrating new knowledge of biological mechanisms obtained through world-leading systems biology approaches are synergistic with EESA programs focused on developing multi-scale mechanistic biotic-abiotic simulation capabilities. This cross-Area integration and innovations in multiscale models from biomolecules to miles will enable powerful macrosystem-scale predictions of ecosystem function relevant to climate and energy-related challenges.

Together these strategies and effective partnerships at Berkeley Lab will drive critical advances in our understanding of the activities and dynamics of interactions of complex microbial communities and microbe-plant associations. Ultimately, this will lead to designed interventions that improve environments and to the discovery of novel environmental organisms and biomolecular mechanisms that can be harnessed for bioenergy and biomanufacturing.

Environment Research Strategies

- **Predictive understanding of environmental organisms.** Discover and deeply characterize the genetic and molecular mechanisms of environmental organisms that drive and respond to environmental changes.
- **Molecular ecosystems based solutions.** Using biological and environmental characterization of natural and laboratory ecosystems to understand native ecosystem processes, predict responses and harness plants and microbes for energy and environmental solutions.

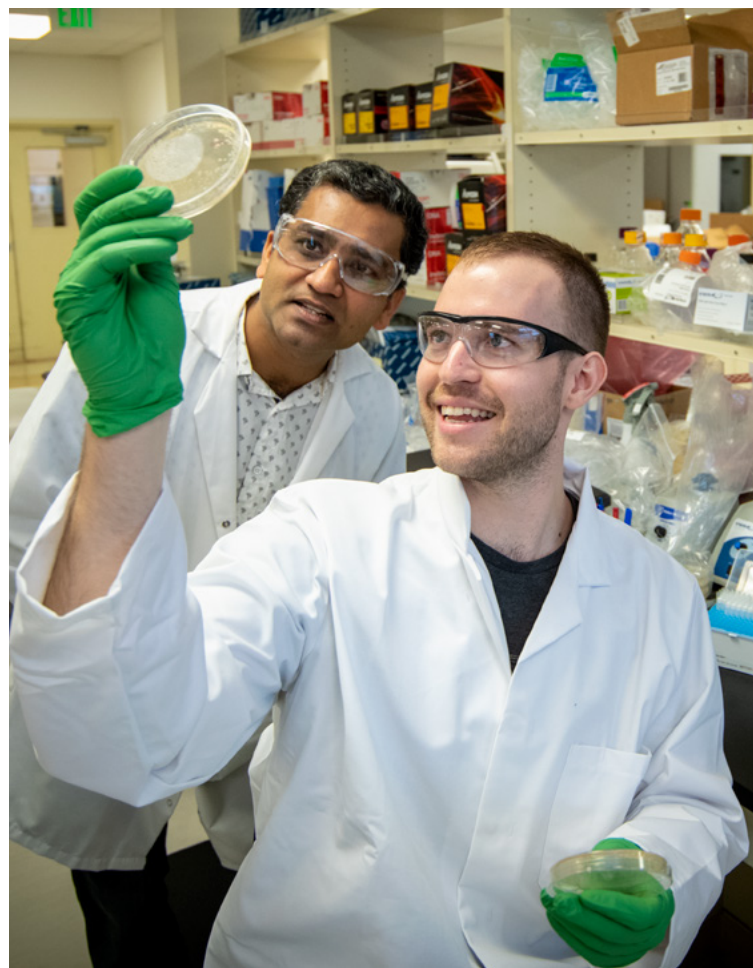
To achieve the Environment Goal by 2023, we focus on two strategies: (1) foundational science to develop a predictive understanding of environmental organisms, and (2) developing molecular ecosystems biologybased solutions. We believe that advancing and coupling these elements are necessary to develop new classes of environmental and energy solutions. Berkeley Lab has the potential to make a unique contribution to understanding integrated environmental system behavior by linking organized research efforts and expertise in molecular microbiology, microbial ecology, and subsurface and terrestrial ecosystem science with global climate expertise. Broadly across Biosciences, researchers are developing a deep understanding of microbial, plant, and metazoan genomics and biomolecular mechanisms using integrated systems biology and computational learning approaches to study the dynamics, stabilities, activities, and interactions of organisms with their communities. To facilitate this work, Biosciences is pioneering the development of advanced functional genomic technologies from EcoFABs to EcoPODs that together will enable discovery of causal mechanisms through manipulation of genetic, organismal and abiotic system components. These model ecosystems are continually referenced to native ecosystems through field studies and refined until the models accurately represent key aspects of the native environments, they provide a powerful framework for determining the necessary and sufficient components and features of an ecological process of interest. They also provide testbeds for examining important ecological questions such as functional redundancy, selective and neutral forces, population dynamics and adaptation to environmental changes.

Our efforts to understand natural ecosystems are multifaceted and interconnected. We use measurement of *in situ* dynamics of natural and manipulated systems as a function of environmental constraints to identify the primary environmental drivers of community

structure and activities. To discover the mechanisms that drive ecosystem dynamics, we integrate technologies for constructing model ecosystems with *in situ* monitoring and manipulation. This work includes spatially and temporally defined sampling coupled to systems biology based analysis (e.g., stable isotope tracing of metabolite, protein, RNA, and DNA molecules, as well as imaging approaches) of both microbes and their hosts. Controlled manipulations within these systems will enable the discovery of causal mechanisms that predict dynamic responses of organisms and their reciprocal impact on their environment, thus enabling the development of effective environmental solutions (e.g., microbiomes that clean contaminated water and growth-promoting microbiomes for bioenergy crops). Through the development and application of emerging third-wave machine learning and artificial intelligence strategies, predictive models will be able to capture mechanisms for resilient energy, agricultural, and environmental properties in microbial communities and plants that can be used for sound ecological solutions.

Predictive understanding of environmental organisms: Discover and deeply characterize the genetic and molecular mechanisms of environmental organisms that drive and respond to environmental changes

Microbes and plants in terrestrial environments carry out a wide range of essential biogeochemical processes. In natural ecosystems, microorganisms live in complex communities, interacting with plants, animals and other microbes. Advances in DNA sequencing now enable rapid and inexpensive interrogation of genomes, from both isolated organisms and mixed environmental communities (metagenomics). As a result, we have and will continue to accumulate vast amounts of sequence information from all types of organisms and environments. The DOE JGI, a user facility within the Biosciences Area, plays a global leadership role in generating and analyzing such data sets from organisms and environmental samples that are relevant to the DOE mission. These sequence data, in principle, contain the information necessary to make accurate predictions of organismal metabolism, community interactions and fitness under controlled conditions. However, we lack vital information on functional attributes of these sequences and computational models and resources necessary to make accurate predictions. For example, approximately half of the genes in any given bacterial genome are of unknown function and we know very little about the genetics and ecology of viruses, which are numerically the most dominant microbes on the planet. Higher organisms including plants, fungi and soil metazoans also play critical roles in global nutrient cycles and other environmental processes, yet their genomics remain poorly understood. This gap between sequence and function greatly diminishes the utility of sequence information and has led to correlative research (e.g., genes correlated with different states or environments). Bridging the sequence-to-function gap and generating a mechanistic, molecular understanding of organismal activities is one of the grand challenges in biology in the post-genome era.





Since function is intimately connected with environmental context, simple pure culture laboratory experiments are insufficient for discovering the functions of many genes. For example, a microbial pathway for making a secondary metabolite that is essential for mediating plant interactions only has a function in the context of a plant. Thus, our functional genomics approach includes characterization of gene functions across environmental conditions and model ecosystems. This approach, coupled with technological innovations in imaging, DNA sequencing, functional genomics, bioinformatics, and systems biology developed at Berkeley Lab and elsewhere, will allow us to study gene functions in the context of changing environmental conditions.

Similar to the "microbial dark matter" of uncultivated, uncharacterized organisms, there exists a vast "dark biochemistry" in nature encoded in the DNA of these uncultivated environmental organisms that must play vital ecological roles that have yet to be discovered. To address this challenge, we use environmental metagenomics integrated with bioinformatics, DNA synthesis, metabolomics and cheminformatics to allow expression of biosynthetic pathways from uncultured environmental microbes in host organisms to identify metabolites and examine their biochemical ecology in laboratory ecosystems. This thrust is closely aligned with the strategic direction of the JGI, which is actively expanding its portfolio of capabilities to include "sequence-to-function" genome-based technologies and approaches with the goal to make them available to their user communities.

This new, more complete pictures of microbial genomes within realistic environmental contexts will provide important new insights into how microbes adapt and evolve, how dynamic their populations are, and the nature of interactions between the microbes and their physical/chemical environment and with plant and metazoan hosts. This functional genomic and physiological understanding will greatly enhance genomics-based modeling approaches and effective manipulation of organisms and environmental systems for environmental or energy benefits.

Molecular ecosystems biology based solutions: Using biological and environmental characterization of natural and laboratory ecosystems to understand native ecosystem processes, predict responses and harness plants and microbes for energy and environmental solutions

Climate change and anthropogenic inputs into the environment are impacting carbon, nitrogen and other nutrient cycles in soils. Microbial communities play critical roles in mediating these processes, as well as many other environmental effects of energy production. Improved understanding of the molecular mechanisms mediating microbial, metazoan, and plant ecology is urgently needed to accurately predict native ecosystem

processes and harness plant and microbial ecosystems for environmental clean-up and sustainable production of biofuels feedstocks and other crops.

To address these critical challenges, Biosciences is developing and using model ecosystems connected to field studies and advanced bioinformatics resources to understand mechanistic aspects that affect nutrient cycling and plant-microbiome interactions that enhance plant growth and abiotic stress tolerance. This, in combination with advanced genetic and systems biology approaches will provide new insights into microbial, metazoan and plant genomics and enable elucidation of causal mechanisms that determine the resilience and change of natural ecosystems in response to climate and anthropogenic inputs. This mechanistic understanding of microbial, soil metazoan, and plant ecosystems will enable the development of new technologies, including defined microbial communities for environmental clean-up and enhanced soil carbon storage as well as new plant cultivars coupled with specific microbiota that increase productivity and efficiency (energy, nutrient, and water-use) to enable bioenergy production on marginal land.

This goal is important since nearly all arable land is already under cultivation. Farmland is also lost due to urbanization and soil degradation, making it critical to develop approaches for cultivation of bioenergy crops on damaged soils. Typically, these degraded soils have very low organic carbon content, which plays a critical role in supporting beneficial microbiomes, increasing water infiltration, preventing erosion and retaining water and nutrients. Building soil carbon content through bioenergy crop cultivation has tremendous implications for both increasing the productivity of damaged soils while at the same time decreasing atmospheric carbon. Soils are a major global carbon reservoir and understanding soil carbon cycling requires studying plants, which are the major carbon input, as well as the activities of microbes and soil fauna that transform these inputs, and the abiotic soil factors that control plant and microbial biochemical activities.

Unfortunately, processes governing soil carbon creation and retention are poorly understood, so that predictive power regarding how environmental conditions impact soil carbon is lacking, much less how we can harness crops and environmental organisms to build soil carbon. Thus, there is great urgency to develop the necessary foundational understanding of soil carbon and other nutrient cycles to enable accurate predictions, develop beneficial practices, and inform effective policies and land management practices. We will help address this challenge by extending systems biology approaches into ecosystems to identify causal principles governing carbon cycling and other ecosystem services.

These efforts will be synergistic with JGI and KBase. JGI has built user communities around topics such as the role of secondary metabolites in plant-microbe interactions and, along with this, substantial expertise in relevant systems and technologies that will cross-fertilize efforts in the Divisions. JGI has also played a global leadership role in producing reference genomes, as well as functional genomic data sets (transcriptomes, epigenomic data, population-scale data) for many of the plants that are of interest to the DOE. This substantial expertise and first-hand access to massive data resources will facilitate research in this area by the Biosciences Divisions. KBase accelerates users' ability to turn complex data into predictions of molecular, organismal, and community function. Therefore KBase will be a vital tool to effectively transfer knowledge about specific biomolecules, organisms, and ecosystems to related systems enabling exploration of new biology and development of predictive models. The resulting "Molecular EcoSystems

Biology"-based predictive understanding will enable Berkeley Lab to help develop more accurate models and biology-based solutions for environmental clean-up, soil carbon management, and much needed low-input agriculture.

2023 10-year Goal achievement measured by:

The 10-year Environment Goal strives to achieve a deep scientific understanding of the major drivers and consequences of environmental change arising from both natural variability and human activities and to develop new environmental solutions that consider integrative system behavior—some of which span both Environment research strategies. Many objectives will be met through investigations carried out using relevant model laboratory ecosystems integrated with field-study sites and synthesis of resulting data sets.

Predictive understanding of environmental organisms

- Use model ecosystems (e.g., EcoFABs and EcoPODs) to discover the functions of genes and metabolites that mediate microbial interactions
- Discover the molecular mechanisms by which microbial communities in groundwater and sediment are altered by anthropogenic contaminants
- Enable data-driven prediction of gene functions for thousands of microbes and metagenomes per year and tens to hundreds of plants
- Develop methods for transferring knowledge about specific biomolecules, organisms and ecosystems to related systems so that exploration of new biology is accelerated, and to better understand the behavior of complex ecosystems
- Characterize the metabolism, stress responses, and interactions of diverse environmental microbial isolates (such as fungi, algae, bacteria, archaea, and viruses) from relevant field sites using a multi-omics approach including sequencing, mutant phenotyping, and genome engineering
- Develop a multi-modal and multi-scale computational infrastructure to accurately and rapidly predict microbial metabolism, gene regulation, and stress response for microorganisms in key environments, such as: contaminated sediments, soils, wetlands, deserts, agroecosystems, and grasslands
- Integrate diverse high-throughput data types, including genomics, transcriptomics, imaging, epigenomic, and metabolomics approaches and computational learning technologies to discover the roles of novel genes, proteins, regulatory sequences, and metabolites in plant and metazoan responses to environmental change
- Use multi-omics measurements in controlled ecosystems of model plants and bioenergy crops to understand the molecular processes through which plants interact with, select, and maintain beneficial bacteria and fungi to improve plant fitness
- Pioneer the development and application of *in situ* and *in vivo* sensors that provide spatially and temporally defined measurements of cellular responses to environmental changes and, secondly, link it to synthetic processors that amplify the adaptive molecular signals enhancing organismal fitness

Molecular ecosystems biology-based solutions

- Create and validate a laboratory-based model plant-soil-microorganism ecosystem (an “EcoFAB”) to investigate the molecular basis of microbiome-driven plant growth promotion
- Implement and further optimize *in situ* sensor technology to explore temporal and spatial heterogeneity and dynamics of the plant-microbe-soil-atmosphere system from lab to field scales
- Develop, test and refine model predictions of native community dynamics to mechanistically account for the material and energy flow at contaminated field sites
- Use model ecosystems, along with genome and community editing, to identify the key microbial functions and cultivated and uncultivated organisms that stabilize microbial communities, and design interventions that decrease energy and nutrient inputs necessary to achieve key ecosystem service goals
- Develop and apply computational learning strategies that integrate detailed imaging, genomic, and metabolic data from model ecological models to accurately predict nutrient cycling and/or biotic interactions in complex ecosystems
- Understand how to harness plants and their microbiomes to achieve a quantitative increase in soil organic carbon storage based on computational predictions from simulated ecosystem experiments
- Demonstrate the ability to design rhizosphere communities within fabricated ecosystems that improve the low-input growth of an important bioenergy crop
- Identify new metabolites exuded by plants that help select and maintain beneficial microbes, which, in turn, improve the low-input productivity of plant biomass production
- Develop computational learning tools to enable mechanistic predictions and discoveries from the integrative analysis of complex, multimodal, multi-scale data collected on artificial and natural ecosystems—including deep learning and causal inference modalities suitable for low-sample regimes

2018—Five-year milestones for Environmental Research Strategies

Near-term efforts largely consist of advancing the functional genomics of environmental and model organisms and studies on controlled and reproducible model ecosystems that recapitulate important aspects of native ecosystems to link genomics to ecosystem processes. Emphasis is placed on studies that improve our understanding of nutrient cycles and biochemical ecology under dynamic environments to harness microbiomes to improve plant productivity and efficiencies (e.g. water, nutrient). We will also obtain new insights into the flow of matter and energy through ecosystems to understand how these processes are mediated by specific organisms to predict critical ecosystem features such as resiliency or productivity.

Predictive understanding of environmental organisms

- Used model ecosystems and bacterial mutant fitness profiling to discover new genes mediating microbial interactions under environmental constraints
- Developed and applied new functional genomics technologies for discovery and validation of gene and noncoding regulatory DNA functions that impact the fitness of plants and microorganisms under multiple environmental conditions
- Used exometabolomic profiling of groundwater and soil bacteria to model how soil metabolites shapes microbial community structure
- Discovered the molecular basis of plant-growth promotion for three bacteria, each from a different phylum, to determine if these mechanisms are conserved
- Used plant mutants with defined microbiomes to discover plant genes that select for beneficial microbiomes
- Identified over 100 metabolites exuded by plant roots and characterized how these metabolites are used by isolate soil microbes and how they impact microbial community structure
- Developed technologies for genome editing, mating, and selection to gain genetic mastery of model microorganisms, plants, and metazoans and applied these strategies to understanding key biological processes in the environment

Molecular ecosystems biology based solutions

- Constructed model desert ecosystems that recapitulate key aspects of native ecosystems for controlled analysis of microbial metabolic process
- Used synthetic biology tools to construct reporter microbes and used these in conjunction with advanced imaging technologies to determine where and when metabolites are being produced within a microbial community
- Elucidated the metabolic strategies of key environmental microbial community members and the functions of specific genes that are important for community metabolism through a combination of systems biology, physiological analysis, functional genomics, and high-throughput microbial isolation
- Used manipulation of environmental variables and spatially defined sequencing to determine the environmental controls on spatial distribution of microbial communities in soil environments
- Applied our understanding of resource competition in a model ecosystem to accurately predict soil metabolite composition and community structure
- Constructed model agroecosystems that recapitulate key aspects for controlled analysis of plant-soil-microbiome process

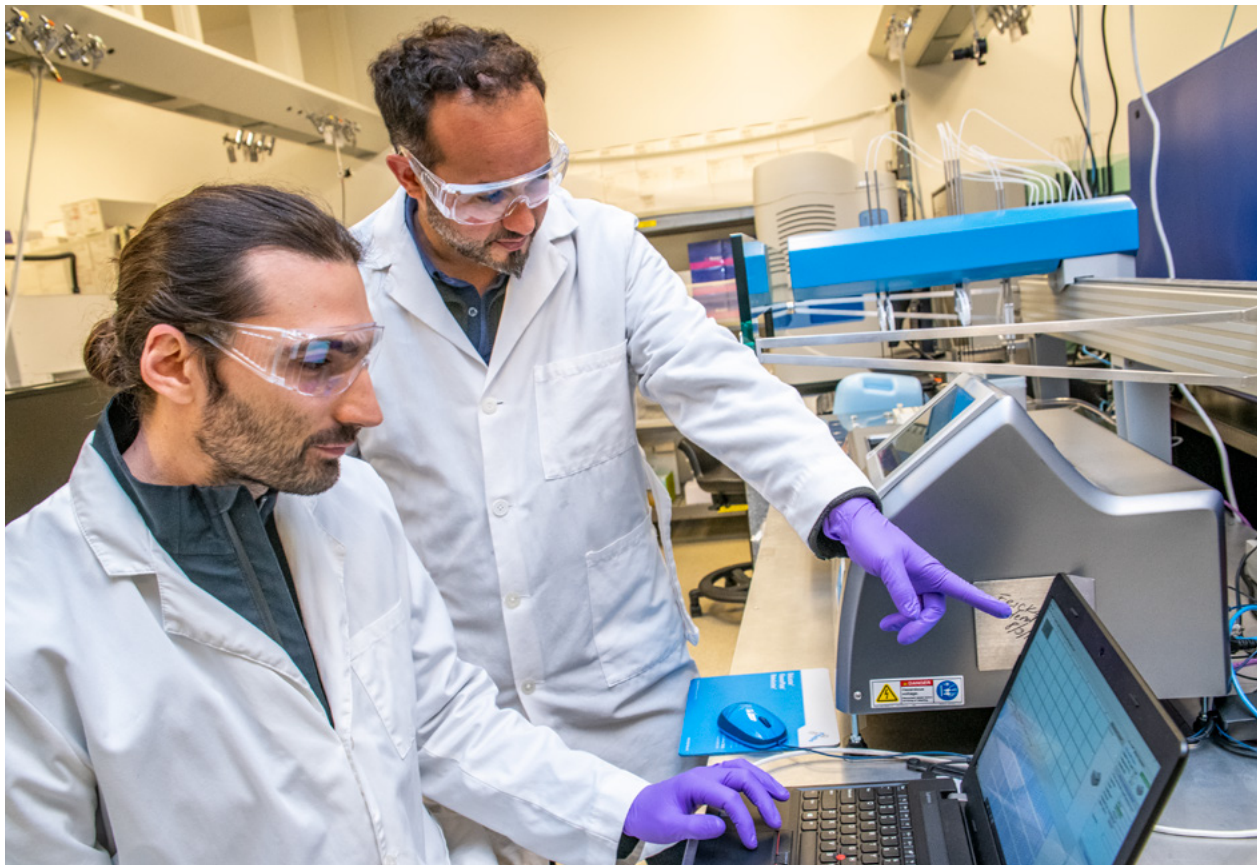
- Demonstrated ability to alter microbiome structure and soil carbon cycling through targeted modification of plant genes
- Discovered at least one novel microbial metabolite that is beneficial to the plant host and confirmed through introduction of the biosynthetic pathway into a model plant-associated microbe
- Developed a “data ecosystem”, computational infrastructure that enabled integration of microbial and multi-cellular systems biology modeling for interactions in microbiomes, especially in native and model ecosystems

A look ahead to 2028

Predictive understanding of environmental organisms

How can we develop molecular profiling methods with substantially higher levels of resolution and computational analysis tools capable of modeling highly complex systems?

- Develop and apply single-cell profiling and imaging technologies that enable the study of plants and fungi as complex, three-dimensional multicellular organisms or community assemblies, with heterogenous gene expression and distinct functions across tissues and cells

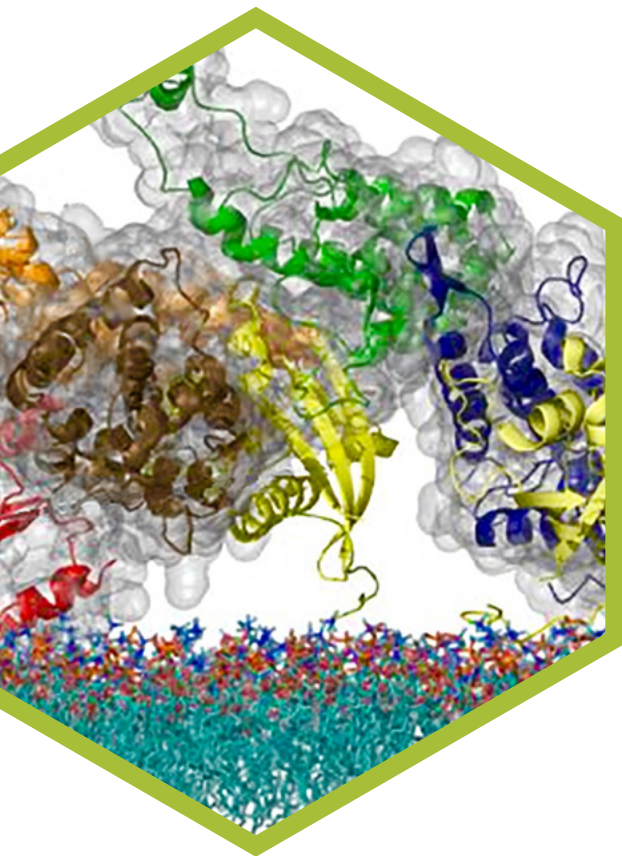


- Develop experimental tools and learning strategies that will facilitate extrapolation of insights from model ecosystems to landscape scales and across spatiotemporal scales
- Create interpretable and explainable machine learning and artificial intelligence methods to build causal models from which insights can be extracted and harnessed for developing intervention or perturbation strategies with predictable outcomes

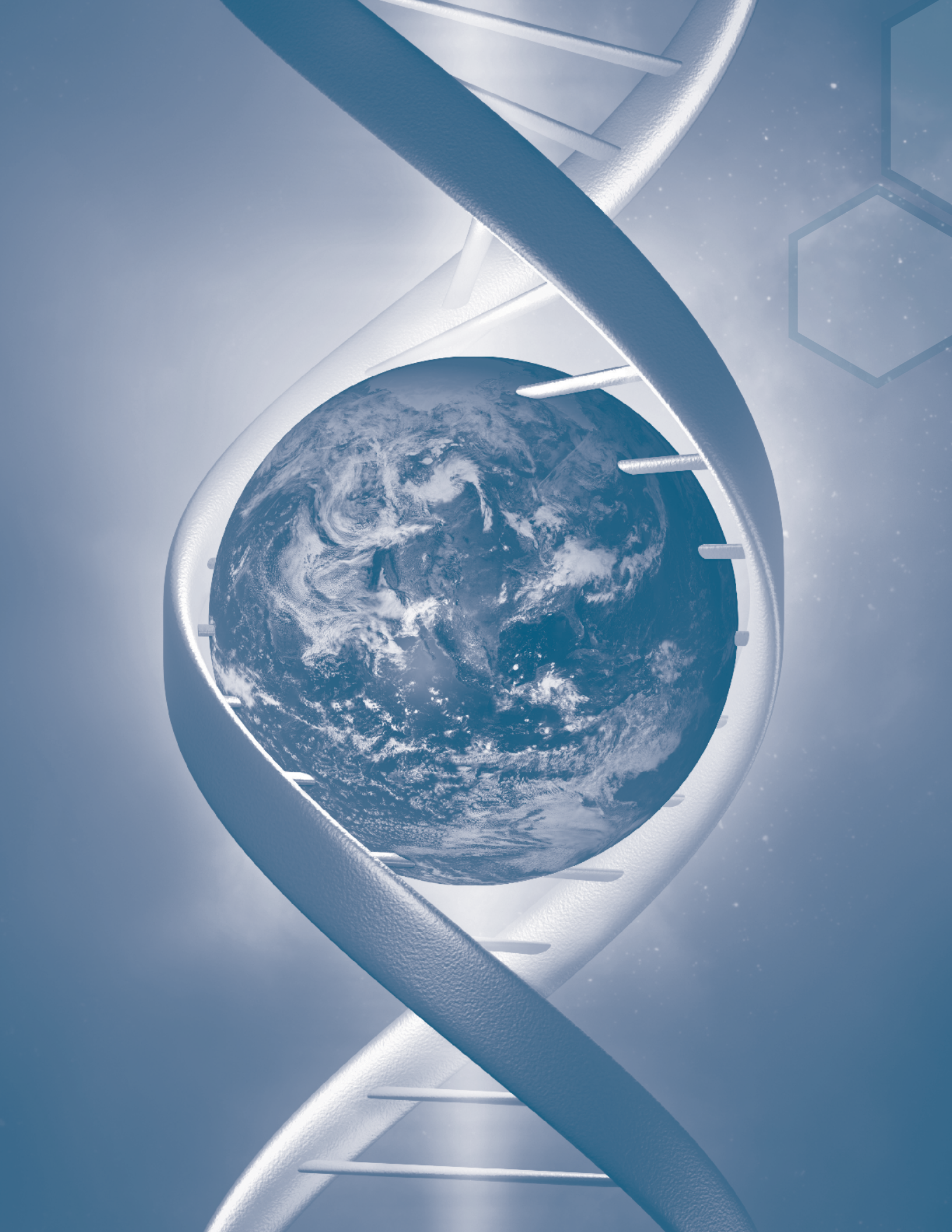
Molecular ecosystems biology-based solutions

How can we form the scientific foundation for the establishment of sustainable environmental management strategies that reduce external inputs and drive co-benefits?

- Develop bioenergy crops coupled with specific microbiomes that enhance soil carbon, are resilient to a range of climatic changes, and provide transformative insights into effective management of contaminated field sites
- Decipher dynamic reciprocity between environmental factors and biological activities across our integrated portfolio of laboratory ecosystems to drive the development and application of break-through technologies
- Advance community editing technologies for targeted engineering of microbial community functions and composition
- Establish next-generation sensors based on quantum sensing for non-destructive *in situ* and *in vivo* real-time molecular measurements that enable tight integration between experimentation and modeling enabling us to form digital twins of fabricated ecosystems









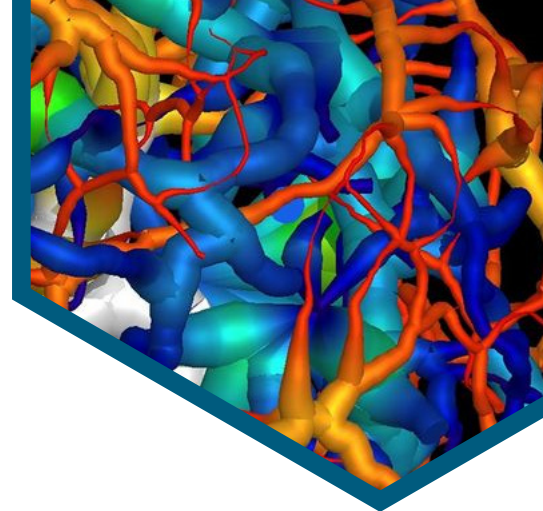
HEALTH

10-year Goal: Develop and apply a predictive, multiscale, integrative understanding of biological systems to improve the bioresilience of humans and our ecosystems, including how biological variation affects individual responses to environmental challenges.

Background and Motivation

Organisms at all evolutionary levels have complex responses to natural and anthropogenic changes or challenges in their environments. External factors that can influence biological health include diet, temperature, climate, water and air quality, chemicals, radiation, nanomaterials, and byproducts from biomanufacturing and energy production. Most chronic diseases in higher organisms—including cancer, cardiovascular and neurodegenerative disease—are caused by adverse gene-environment interactions, with the environmental component playing a major role. Environmental factors induce effects by altering molecules, cells, and physiological processes inside organisms. For example, in humans, environmental exposures such as radiation, physical trauma, and certain drugs damage brain tissue, leading to behavioral changes and memory loss, anxiety, and neurological diseases. In animals, environmentally induced defects in neural function impact basic behaviors (finding food, mating behavior, aggression, etc.) that transcend the individual organism and affect the long-term survival of individual species and entire ecological systems.

A lack of understanding of the biological and health responses to environmental challenges also has a large impact on the US economy. For example, energy production impacts air and water quality by generating airborne contaminants that increase rates of asthma and cancer, and broadly reduce human resiliency. In the United States, an additional \$886 billion would be spent per year



10-YEAR Health Goal

Develop and apply a predictive, multiscale, integrative understanding of biological systems to improve the bioresilience of humans and our ecosystems, including how biological variation affects individual responses to environmental challenges.

Health Strategies to Achieve Goal

Multiscale understanding of human biology. Develop and deploy functional genomics, physiological monitoring, imaging, and computational modeling to enable foundational insights into human biology for applications in health and biodefense.

Biological responses to environmental challenges. Design and integrate experimental and computational approaches to understand how individual genetics, epigenetics, and microbiomes impact molecular, cellular and organismal responses to environmental challenges, and to identify biomarkers for disease risk assessment, as well as prevention and mitigation strategies.



if the price of energy reflected the hidden costs associated with lost work time and early mortality due to energy-induced environmental challenges. Improvements in understanding the health consequences of energy production lead to cost-effective regulation: the Clean Air Act generates an estimated net \$83 billion per year for the US economy. Other environmental challenges, such as intentional or unintentional deployment of chemical toxicants, have similar "hidden" economic costs.

Berkeley Lab envisions the emergence of a concerted national effort focused on the protection of human health through advances in environmental stewardship. It is therefore of central importance to understand the health consequences of energy production and other environmental challenges, and to develop new remediation strategies and guide technologies toward safe and sustainable solutions. To realize this vision, advances are required in exposure biology, neurology, and the science of individual susceptibility, precision ecotoxicology, and many other research areas. One of the greatest scientific challenges is balancing the potential risks associated with innovation and sustainable economic growth with maintaining the fitness of humans and the Earth's biomes.

It is well known that genetic background and life history (e.g. gender, age, health, and diet) can modulate tissue and organismal responses to environmental exposures. We have recently come to appreciate how the trillions of microbes present in humans and other species can greatly impact organismal and ecological health. The host plus of all its microbial symbionts defines the holobiont. These associated microbes produce enzymes, vitamins and anti-microbial substances that contribute to individual susceptibility to environmental toxicants. For example, microbial metabolites produced in the human gut or soil metazoans can be much more or less toxic than their parent compounds. The internal chemical context of organisms continually fluctuates during life, due to changes in external and internal sources, so social and environmental factors also must be considered when evaluating impacts on fitness and disease.

Berkeley Lab aims to address scientific and societal challenges in a comprehensive manner. A major challenge for 21st-century biology is to develop a deeper understanding of the types of responses to environmental challenges at many interrelated biological levels, including molecules, cells, cellular communities, tissues, and organisms, and the multitude of timescales associated with dynamics across these different scales. The ability to assess the genetic contributions to health and disease, the development of quantifying technologies for analyzing the genome, transcriptome, metabolome, fluxome, and proteome ("omics"), and computational methods for integrating diverse and large data sets, have blossomed in recent years. However, quantitative assessment of the extent of organismal exposure to environmental challenges, and its relation to fitness and disease, have lagged due to technical limitations in assessment and monitoring, and a lack of comprehensive and accurate data. Thus, it is essential to develop an integrative, quantitative, and predictive understanding of the biological responses to environmental challenges, and how they impact the fitness of humans and ecological biomes. Through the development of biologically informed solutions to pressing societal problems, this knowledge will lead to reduction in risks associated with deploying new cutting-edge products. It will also ensure that economic growth is balanced with the policy reforms and targeted interventions required to maintain the fitness of humans and the ecosystems we depend on.

Human health is tightly coupled to the health of the biome, especially the plant and microbial communities that perform a variety of essential ecosystem services, including nutrient cycling, clean air and water, abundant food supply, and chemical transformations. Disruption of these services by environmental challenges (e.g. climate change) will have profound impacts on human health and the global economy. Further, an improved understanding of the potential risks and benefits of chemical influences on individuals and ecosystems will enable new approaches to streamline product design in areas including biomanufacturing, agriculture, and pharmacology. Screening procedures that broadly characterize the biological activities and potential risks of compounds at low cost and high-throughput have the potential to identify new bioactive compounds, including herbicides, antimicrobials, and prebiotics.

Unraveling the complexities of biological responses to the many internal and external environmental challenges ahead—and their impact on the health of humans and the biome—requires technical and multi-disciplinary advances allowing a deep, multiscale, integrated knowledge of mechanistic and phenotypic responses to these factors. For example, determining the roles of environmental factors to the initiation and progression of cancer and neurological diseases requires understanding, predicting, and mediating complex reciprocal interactions among multiple levels of genetic and physiological functions, in both host organisms and their microbiomes. These same skills and disciplines are required to monitor and analyze the health of humans and the biome. Whether the subject is the human body, microbial communities or a critical insect population, what is needed is a quantitative understanding of the short- and long-term responses to environmental challenges encountered by hosts and their resident microbiomes, as well as their combined effects on organismal fitness, behavior, and long-term health.

Berkeley Lab has a long history of successfully integrating multiscale, cross-disciplinary approaches to address challenges of similar size and complexity. The Biosciences Area aims to address bioeconomy-related challenges in a comprehensive manner, by leveraging advanced facilities and equipment, a culture of interdisciplinary team science, and an historic and deep level of biological and technical expertise. Relevant disciplines include structural biology; biochemistry; cell and organismal biology; microbial communities; genetics, genomics, and epigenetics; metabolomics and proteomics; multiscale imaging and physiological measurements; and data integration through advanced computational analyses. Berkeley Lab has extensive experience with integrating data from multiple levels of biological function to understand the impact of environmental challenges—including radiation, climate change, toxicants and third-hand smoke—in model systems, and the effects of exposures on the etiology of diseases such as cancer and neurological disorders. Importantly, the Biosciences Health efforts leverage Berkeley Lab's extensive technical capabilities and National User Facilities, including the ALS, JGI and NERSC. In addition, partner faculty at University of California (UC) campuses at Berkeley, San Francisco, and Davis and other Bay Area institutions bring additional expertise and experience in molecular epidemiological studies of human health and disease, which complement the Berkeley Lab team science mission and capabilities in elucidating basic biological mechanisms.

Together, Berkeley Lab and partner institutions are poised to integrate basic, mechanistic information about the responses of biological systems to environmental challenges with accurate measurement and understanding of the extent and magnitude of exposures and their impact on health and disease. Achieving this mechanistic understanding will address a pressing national need to accurately predict the impact of environmental challenges on biological systems. This knowledge will then be leveraged for risk management, by eliminating or modifying detrimental compounds from the environment, identifying individual humans or organisms that are sensitive to exposure, and developing innovative bio-based solutions. Berkeley Lab's efforts will result in the development of safe, sustainable energy and materials and reduced exposures to harmful environmental factors, improved public health and personalized medical interventions, and increased understanding of the impact on organisms that are essential to biome fitness. Given sustained effort, Berkeley Lab will be a world leader in generating scientific discoveries that have long-term, high-value impacts on improving the fitness of humans and the biome, and quality of life, resulting in significant positive impacts on the overall economy.

Health Research Strategies

- **Multiscale understanding of human biology.** Develop and deploy functional genomics, physiological monitoring, imaging, and computational modeling to enable foundational insights into human biology for applications in health and biodefense.
- **Biological responses to environmental challenges.** Design and integrate experimental and computational approaches to understand how individual genetics, epigenetics and microbiomes impact molecular, cellular and organismal responses to environmental challenges, and to identify biomarkers for disease risk assessment, as well as prevention and mitigation strategies.

National laboratories excel at integrating the multiscale, complex data required to address these important challenges. To achieve the Health Goal by 2023, Berkeley Lab will: 1) build a cross-disciplinary platform that provides a comprehensive, integrated understanding of positive and negative responses to environmental challenges, using complementary metazoan and microbial model systems that cover levels of function from molecules to organisms, and 2) assess the impact of environmental challenges on human health and disease. These focus areas will integrate multiscale data using bioimaging, genomics, proteomics, metabolomics and computational technologies to rapidly phenotype and quantitatively interrogate complex, dynamic biological systems. Technological and intellectual advances at Berkeley Lab and elsewhere will accelerate the success of the Health research strategies.

Berkeley Lab's work on understanding human and biome health responses to environmental perturbations will be performed in close integration with the other Biosciences focus areas (Energy, Environment, Biomanufacturing, and Technologies for Biosciences). For example, development of the bioimaging, computational, and functional genomics tools described in "Technologies for Biosciences" will be crucial to the success of this program. Through these integrated efforts mechanistic understanding will be

developed that will ultimately inform prevention, diagnosis and treatment. The Health component research strategies will also play key roles in development of cross-cutting technologies for rapid assessment and imaging of phenotypes that will be utilized across the Biosciences Strategic Plan.

Together, these strategies address short- and long-term national needs by generating a comprehensive understanding of the impact of environmental challenges on the fitness of biological systems, and developing innovative bio-based solutions to assess hazards and mitigate health problems related to environmental exposures. These strategies are designed to be achievable in a ten-year time-span. The technical and intellectual strengths of the Biosciences Area and of other Berkeley Lab scientific Areas will be brought together to achieve the advances needed to enable productive application of renewable biological resources that protect and improve the economy. As a result, novel methods will be developed that connect nanoscale to organismal behaviors by collecting multiscale data simultaneously from the same animals or tissues. This will enable a predictive, multiscale understanding of responses to environmental challenges, which will promote responsible and sustainable economic growth, as well as ensure the fitness of humans and the biome.

Multiscale understanding of human biology: Develop and deploy functional genomics, physiological monitoring, imaging, and computational modeling to enable foundational insights into human biology for applications in health and biodefense

Berkeley Lab has a long history in developing advanced technology platforms, large-scale facilities, high-throughput workflows and multi-investigator teams to explore fundamental processes of human biology and environmental impacts on human health. In particular, Berkeley Lab has been a major contributor to the Human Genome Project, establishing extensive resources and expertise in the area of genomics that have subsequently been complemented by pioneering functional genomics and model organism approaches. Taking advantage of unique infrastructure and capabilities, Berkeley Lab has also become a world leader in advanced imaging and computation for human biology applications.

A fundamental challenge associated with the functional interpretation of the human genome that became evident from the Human Genome Project is its massive size and complexity, coupled with a limited understanding of the function of genes and gene regulatory sequences residing within the genome. These challenges have become particularly pressing as it is now possible to determine the sequence of human individuals at scale (whole genome sequencing), yet the interpretation of sequence variation discovered through this approach and linking it to





phenotypes including diseases, disease susceptibility, or inter-individual differences in responses to environmental cues, remain difficult. To address this grand challenge, we will develop and apply advanced technologies that enable such functional interpretation of genomic sequence and variation therein. Specifically, the development of advanced epigenomic profiling methods and their application to relevant sample types, such as human tissues associated with diseases or environmental exposures, or tissues derived from purpose-engineered animal models will enable us to link genome sequence to organismal function and responses to endogenous and exogenous stimuli.

Advanced imaging and sensor technologies developed at Berkeley Lab have also opened new avenues into exploring the most complex organ system found within the human body: the brain. A key aspect of these successes has been the integration of multiple orthogonal technologies, including experimental and computational techniques. One challenge in the interpretation of advanced measurement of brain function is our lack of understanding of the physiological and cellular basis of signals that are observed by intracranial brain monitoring technologies such as electrocorticography (ECoG), providing barriers to their interpretation. We aim to gain an improved understanding of the biology underlying these signals and, in turn, apply this to gain an improved understanding of brain function. One area of particular interest is the understanding of distributed circuits within

the cortex and their role in coordinating complex perceptions and behaviors, using a combination of experimental measurements and novel mathematical frameworks to understand the response dynamics of neuron populations. Beyond brain physiology, it is increasingly recognized that disorders of the brain, including major neuropsychiatric disorders, have a component in prenatal development and abnormal gene regulation. Building on functional genomic technologies, we will explore the role of genes, regulatory sequences, and variation in these processes.

Developing new computational tools and resources to understand functional relationships between processes and responses in model systems and humans will be of central importance. The Gene Ontology and the Human Phenotype Ontology within the Monarch Initiative are essential tools, and these need to be expanded and extended to include environmental contexts, the presence of microbial communities, and quantitative as well as qualitative and descriptive models. Third-wave machine learning technologies that combine the transparency of statistical models for causal inference with the predictive and descriptive power of deep learning will be important for discovering and mapping these relationships.

Biological responses to environmental challenges: Develop and deploy model systems together with human population studies to understand how individual genetic, epigenetic and microbiome variation impacts molecular, cellular and organismal responses to environmental challenges, and to identify biomarkers for disease risk assessment

This strategy aims to develop a mechanistic understanding of how environmental challenges—specifically chemicals, radiation, nanomaterials, diet and energy production and use—impact the fitness of metazoan organisms. Understanding how organisms function in adverse environments is an essential component of biodefense. Here we intend to identify, understand, and reduce threats posed by environmental conditions and exposures.

An important first step is defining how exposures affect biological systems, from macromolecular complexes to biological outcomes. Using diverse model systems will enable us to leverage evolutionary biology and map multiscale models of chemical exposures onto the phylogenetic tree. A critical advance will be the development of quantitative models linking responses measured in one or a few species or systems to others. A central aim is to predict the effects of these exposures and additional challenges on organism fitness, and to develop solutions that prevent or mediate negative impacts. Technologies for measuring the types, levels and distributions of biomolecules and cellular and organismal phenotypes have advanced considerably. However, the robust application of these technologies in an integrative, comprehensive manner that reveals impacts on interrelated levels of biological mechanisms and functions has lagged considerably.

Due to the enormous complexity and multifaceted nature of biological systems, a holistic approach is needed to achieve the 10-year Goal. Environmental factors can affect different types of molecules, cells, tissues and organisms independently. How biological functions are affected by interactions among these components must be incorporated to achieve a true and effective understanding of impact on the fitness of an organism. For example, understanding how toxicants impact fitness requires integrating measurements of changes in the levels and distributions of diverse biomolecules (eg., DNA, RNA, metabolites and their fluxes, proteins) in different cells and tissues, as well as information about alterations in the composition and activities of resident microbiome communities, combined with quantification of the impact on host cells, tissues and organismal phenotypes and functions.

By identifying key biological mechanisms in model organisms, instead of merely cataloging relationships between input and output, responses of biological organisms in hypothetical situations can be predicted based upon mechanistic knowledge. For example, can the presence of specific microbiome communities protect against specific toxicants? Can protection be modulated by the genomic makeup of the individual organisms and the host? To answer these questions, manipulatable, isogenic experimental systems that allow control of genotypes and environments are required to enable precise measurements of responses in a way that is impossible with studies of humans. The results will provide a scientific and mechanistic foundation that will inform microbial and human fitness studies.

This research strategy utilizes three well-established model systems, i.e. *Drosophila melanogaster* (fruit fly), *Mus musculus* (mouse), and *Rattus rattus* (rat), whose strengths

have been used to generate deep biological insights over the last century. Most importantly, these model systems provide the ability to leverage sophisticated genomic, genetic, epigenetic, cell biological and developmental tools for multimodal manipulation and measurements of biological responses to environmental challenges, under conditions where the environment, genetic variation and microbiome composition can be controlled. For example, the rat has been foundational for understanding changes in brain function and behavior after environmental challenges. It is also the preferred model for testing innovative biosensing and neuroimaging technologies for measuring brain function, neural networks and communication, and behavior at multiple scales. In these tractable systems, Biosciences aims to quantify responses at the molecular level by developing and applying advanced "omics" tools and expertise focusing on changes to the types, structures, and levels of DNA, RNA, proteins and metabolites and their fluxes in both the host and resident gut microbiomes.

The manipulability of these models also will allow us to correlate molecular changes with cell, tissue and organismal phenotypes, as well as individual variations in responses, ranging from changes in molecular machines and pathways to cellular structures, tissue phenotypes, and animal behaviors. To provide a comprehensive understanding of the health effects of environmental challenges the advanced computational infrastructure at Berkeley Lab will be employed to manage, analyze, visualize and integrate the "big data" generated by these studies. Lessons learned from these controllable model systems will also provide mechanistic insights that will inform approaches to elucidating environmental impacts on human biology and health.

2023 10-year achievement measured by:

Multiscale understanding of human biology

- Develop and apply large-scale functional genomics technologies to elucidate genome functions relevant to human biology, evolution, and disease:
 - Understand how noncoding sequence variation impacts human development, evolution, phenotypic variation, and disease susceptibility through integrative use of genomic, computational, and model organism strategies
 - Develop and harness innovative CRISPR-Cas genome editing strategies in model organisms to inform structural and functional analyses of the human genome
- Integrate multiple approaches to understand brain functioning in health and disease:
 - Elucidate the cellular basis of the signals measured by human brain monitoring technologies (e.g., ECoG)
 - Understand how distributed cortical circuits are coordinated to give rise to complex perceptions and behaviors
 - Create mathematical frameworks to recover networks that give rise to neural population dynamics
 - Use combined epigenetic profiling and model organisms to link psychiatric genetic findings to cell type-specific gene regulation
- Use of computational methods to translate knowledge from model organisms and model systems for predictive disease diagnosis, prevention, therapy, and risk management in order to advance human health and bioresilience:

- Developing methods to predict the pathogenicity of variants in the human genome using machine learning
- Develop computational learning frameworks to discover the epistatic architecture of complex traits from population data
- Predicting causal associations between genomic variants and cellular and whole-organism phenotypes, leveraging experimental data from model organisms
- Developing causal models of gene function in the context of metabolic and signaling pathways in human cells and organ systems
- Predicting relationship between exposome and bioresilience of humans and other organisms
- Develop transfer learning strategies to translate and unify knowledge across model systems

Biological responses to environmental challenges

- Identify the keystone components that mediate the impact of environmental challenges (prioritized based on human epidemiological studies and model system discoveries) on molecules, cells, microbial communities, tissues and organisms in tractable model systems, including:
 - Phenotypic responses to environmental challenges exhibited by cells, microbial communities, tissues, and organisms
 - Dynamic responses of molecules (DNA, RNA, protein, metabolites and their extracellular and tissue fluxes) and phenotypes using advanced imaging, genomics, phenomics and computational approaches
- Determine how the effects of environmental challenges are modulated by genetic, physiological and epigenetic variation in model systems, as measured by:
 - Genetic mapping to identify DNA sequence differences that affect individual responses to environmental challenges
 - Cellular and tissue damage responses that vary by physiological status (gender, age, diet, etc.)
- Elucidate the role of prototypic community interactions within biological systems, and how they are reciprocally affected by environmental challenges, as measured by:
 - Model host/microbial community composition and functional responses to environmental challenges that identify conserved and species-specific mechanisms
 - Identifying microbial communities that impact fitness and define their interactions with each other and hosts
 - Identifying changes in brain function and behavior due to reciprocal flux of gut-microbial and metabolites along the gut-blood-brain axis
 - Demonstrate the ability to manipulate reciprocal interactions between microbes and model organisms that produce benefits to fitness in response to environmental challenges
- Determine if environmental challenges shown to impact model organisms similarly affect human cells and populations and tissues, as measured by:
 - Validation of human biomimetic tissue culture systems (e.g. breast, skin) fabricated from normal primary human cells and extracellular matrices, with respect to relevance to *in vivo* tissues at the levels of architecture and gene and protein expression patterns

- Quantifying the biological effects of exposures (identified as having impact in model systems) on human biomimetic tissues
 - Integrating discoveries about specific environmental challenges into exposome/epidemiological studies in human populations
 - Initiation of strategies for disease prevention, therapy, and risk management based on individual predispositions and systemic responses to environmental challenges
- Integrate mechanistic insights from human and model system studies to develop computational models to predict the effects of environmental challenges on human health, as measured by:
 - Generating a list of bioindicators for human health and disease that includes genes, epigenetic markers, proteins, metabolites and their fluxes, microbiome, and physiological components
 - Through applying standardized ontologies and workflows, link human microbiome data to environmental microbiome data within NMDC to enable cross-study investigations and modeling to extend understanding of microbial processes.
 - Successfully predicting how manipulating host and microbiome properties positively or negatively impact human bioresilience
 - Use knowledge from studies in humans and model systems to design strategies for disease prevention, therapy, and risk management based on individual predispositions and systemic responses to environmental challenges, as measured by:
 - Using phenotypic data, genetic predispositions, and other biological assays to identify individuals at risk for harm from specific environmental challenges,
 - Deploy deep learning algorithms that integrate omics, phenotypic and exposome data that more effectively identify sensitive populations and biomarkers,
 - Developing personalized therapeutic and prevention strategies by utilizing links between genetic, microbiome composition, and environmental responses

2018 Five-year milestones for Health Research Strategies

Multiscale understanding of human biology

- Understand how noncoding sequence variation impacts human development, evolution, phenotypic variation, and disease susceptibility through integrative use of genomic, epigenomic, computational, and model organism strategies
- Develop and harness innovative CRISPR-Cas genome editing strategies to enable high-throughput interrogation of mammalian genome structure and function
- Develop methods to predict the pathogenicity of variants in the human genome using machine learning

Biological responses to environmental challenges

- Identified genetic, transcriptomic, metabolomic, microbiome and phenotypic responses (molecules, cells, tissues, organisms) to environmental challenges (e.g. anthropogenic pollutants such as heavy metals and synthetic organic compounds) in model microbial communities, biomes and eukaryotic organisms

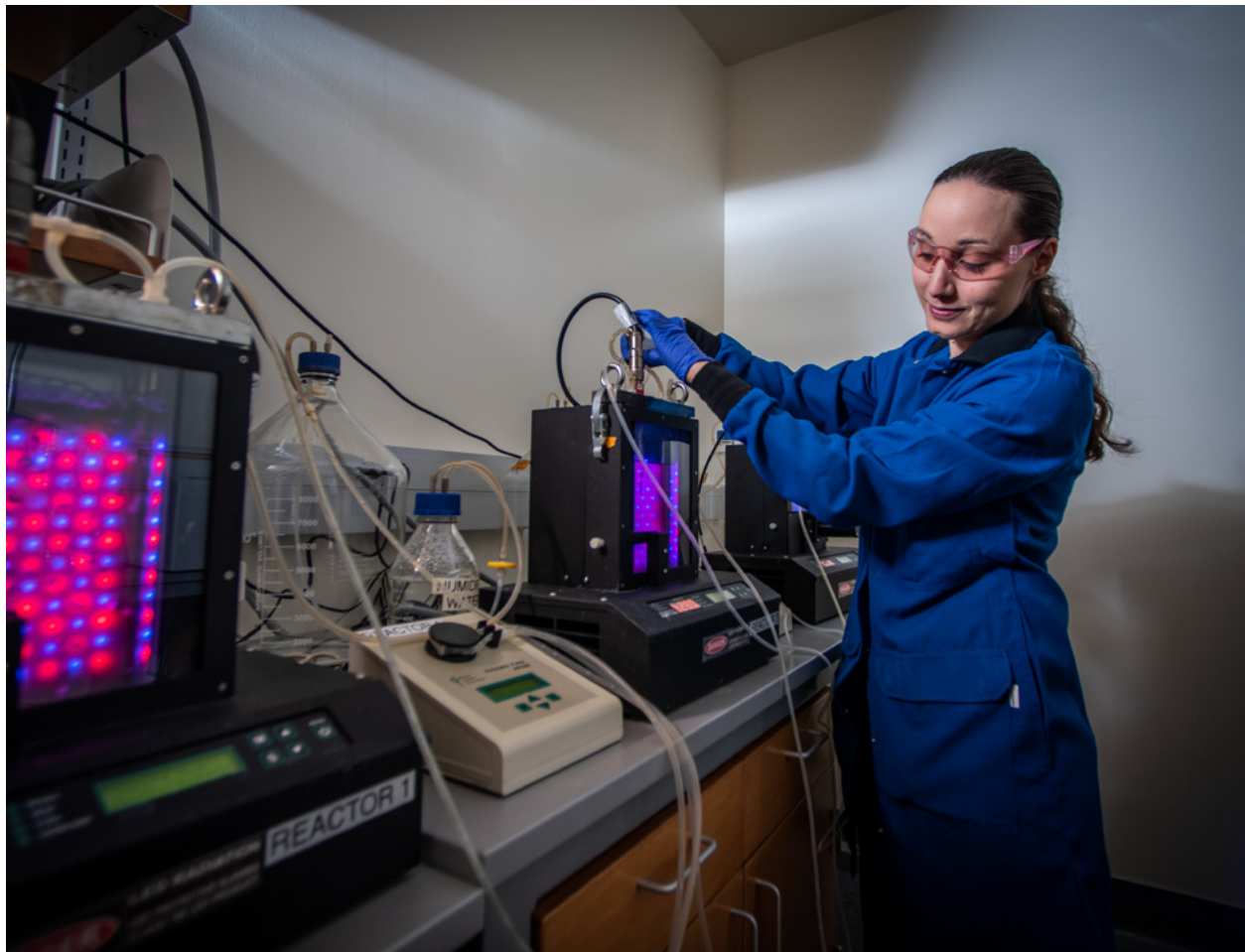
- Demonstrated technologies for high-throughput characterization of macromolecular complexes acting in environmental responses
- Integrated data to develop and test mechanistic and predictive models of two environmental responses, including microbiome/host interactions
- Formulated a list of predictive bioindicators for fitness that includes genes, metabolites, and microbiome components

A look ahead to 2028

Multiscale understanding of human biology

How do we enhance human bioresilience by systematically translating multi-scale models that integrate genetic variation, gene regulation, metabolic fluxes, and microbial inputs?

- From foundational studies reconstruct individual cells, tissues, organisms and their microbiomes

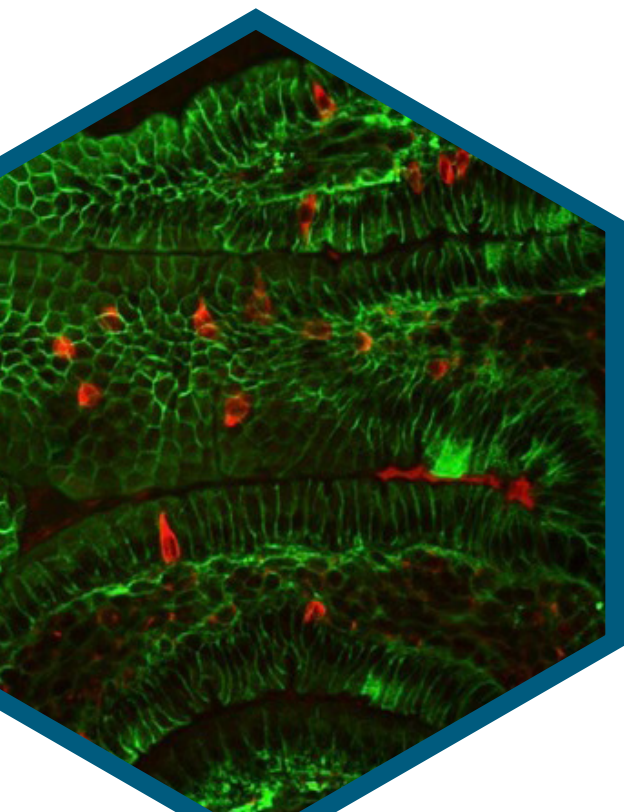


- Develop integrative multi-scale biomarkers to predict health, performance and resilience
- Develop quantum sensors to enable non-destructive monitoring of neurons, and connect gene regulatory and metabolic models to behavior and decision-making
- Computationally predict the structures and functions of the vast majority of proteins in a given genome beyond the exascale (quintillion calculations per second)
- Develop full holobiont metabolic flux simulations for use in *in silico* chemical and drug screening

Biological responses to environmental challenges

How do we prevent, ameliorate or mitigate the adverse health effects of environmental challenges?

- Design and integrate experimental and computational approaches to understand how individual genetics, gene regulation, metabolic fluxes and microbiomes impact molecular, cellular and organismal responses to environmental challenges
- Integrate technologies to advance therapeutic discovery and design
- Develop new approaches to improve human bioresilience to chemical and physical environmental challenges
- Develop biomarkers for risk assessment, prevention, mitigation strategies and biodefense







RACK 500282 FOR 6mm TU
FOR USE IN THE COVARIS I

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24



BIOMANUFACTURING

10-year Goal: Develop and demonstrate accessible, scalable, flexible, cost-effective, and sustainable biology-based manufacturing infrastructure and expertise driven by applications in energy, health, materials, environment, and agriculture.

Background and Motivation

Forty years ago, the development of recombinant DNA revolutionized biotechnology. The unprecedented ability to engineer bacteria birthed multibillion-dollar industries in pharmaceuticals, materials, chemicals, foods, and fuels. The promise of biomanufacturing has also inspired decades of improvement in our basic understanding of biology. Sequencing technologies have been developed to quickly and reliably identify useful genes across ecosystems. Computational and analytical technologies have been developed to predict and measure the presence of metabolites. A detailed understanding of biomolecular function and cellular physiology has been developed, enabling methods to effectively express pathways and engineer hosts. Synthesis of arbitrary DNA sequences has advanced, obviating the need for physical templates. Knowledge systems have been built to learn from past experiences and to computationally design and control biomanufacturing processes.

Despite these advances, there remain significant technological barriers to the predictable development of biomanufacturing solutions. Berkeley Lab's Biosciences Area can contribute to overcoming these barriers through the development of science and technology programs that create scalable infrastructure and knowledge systems, and foster a network of expertise and innovation in engineering biology.

10-YEAR Biomanufacturing Goal


Develop and demonstrate accessible, scalable, flexible, cost-effective, and sustainable biology-based manufacturing infrastructure and expertise driven by applications in energy, health, materials, environment, and agriculture.

Biomanufacturing Research Strategies to Achieve Goal

Tools to design, construct, and debug biology. Develop computer-aided design and fabrication tools, computational and analytical methods to model and learn from engineered biological systems and processes, and retrosynthetic infrastructure providing optimized pathways to key molecular hubs.

Designed biological systems. Engineer and scale-up predictable, controllable, trackable, and robust biological systems (prokaryotes, archaea, eukaryotes, microbiomes) for key energy, health, and environmental biomanufacturing applications.

Biodirected materials and bionanoscience. Couple biological components to chemical and physical systems to biosynthesize desired mineral/metal nanostructures.



Biomanufacturing Research Strategies

- **Tools to design, construct, and debug biology.** Develop computer-aided design and fabrication tools, computational and analytical methods to model and learn from engineered biological systems and processes, and retrosynthetic infrastructure providing optimized pathways to key molecular hubs.
- **Designed biological systems.** Engineer and scale-up predictable, controllable, trackable, and robust biological systems (prokaryotes, archaea, eukaryotes, microbiomes) for key energy, health, and environmental biomanufacturing applications.
- **Biodirected materials and bionanoscience.** Couple biological components to chemical and physical systems to biosynthesize desired mineral/metal nanostructures.

To achieve the Biomanufacturing Goal by 2023, Berkeley Lab developed an approach that employs three areas of strategic focus: the development of tools to design, construct, and debug biological systems; the design and scaling of biological systems; and the creation of biodirected materials and bionanoscience. These areas are scientifically tractable within a 10-year span, will meet the long-term national need for novel biomanufacturing solutions, and leverage specific facilities, organized research groups, and core competencies within the Biosciences Area at Berkeley Lab. These strategies will be executed in parallel.

Tools to design, construct, and debug biology: Develop computer-aided design and fabrication tools, computational and analytical methods to model and learn from engineered biological systems and processes, and retrosynthetic infrastructure providing optimized pathways to key molecular hubs

The pace, cost-effectiveness, and scope of biomanufacturing process development is severely limited by our imperfect ability to design and control sophisticated gene networks and metabolic pathways. Once a biological system has been constructed, there are limited tools to debug the system and improve upon it. We also cannot efficiently and predictably scale processes from the lab bench to relevant pilot, demonstration, and commercial scales.

Biological Computer-Aided Design and Manufacture (BioCAD/BioCAM)

Advanced engineering relies on sophisticated mathematics, informatics, and computing infrastructure to guide all aspects of system design and manufacture. Electronics design automation revolutionized the electronics industry with tools spanning simulation of silicon materials, physical and logic design of circuitry, and physical layout and manufacturing optimization engines. These billion dollar-scale infrastructure assets are based on libraries of knowledge and models of physical principles, standard manufacturing protocols and design elements (physical parts), and design templates for standard applications. Standards for information interchange, algorithmic update and testing, and form factors and interconnects have been specified so that multiple horizontal industries can compete to serve various vertical application industries.

We need the equivalent infrastructure to industrialize engineering biology.

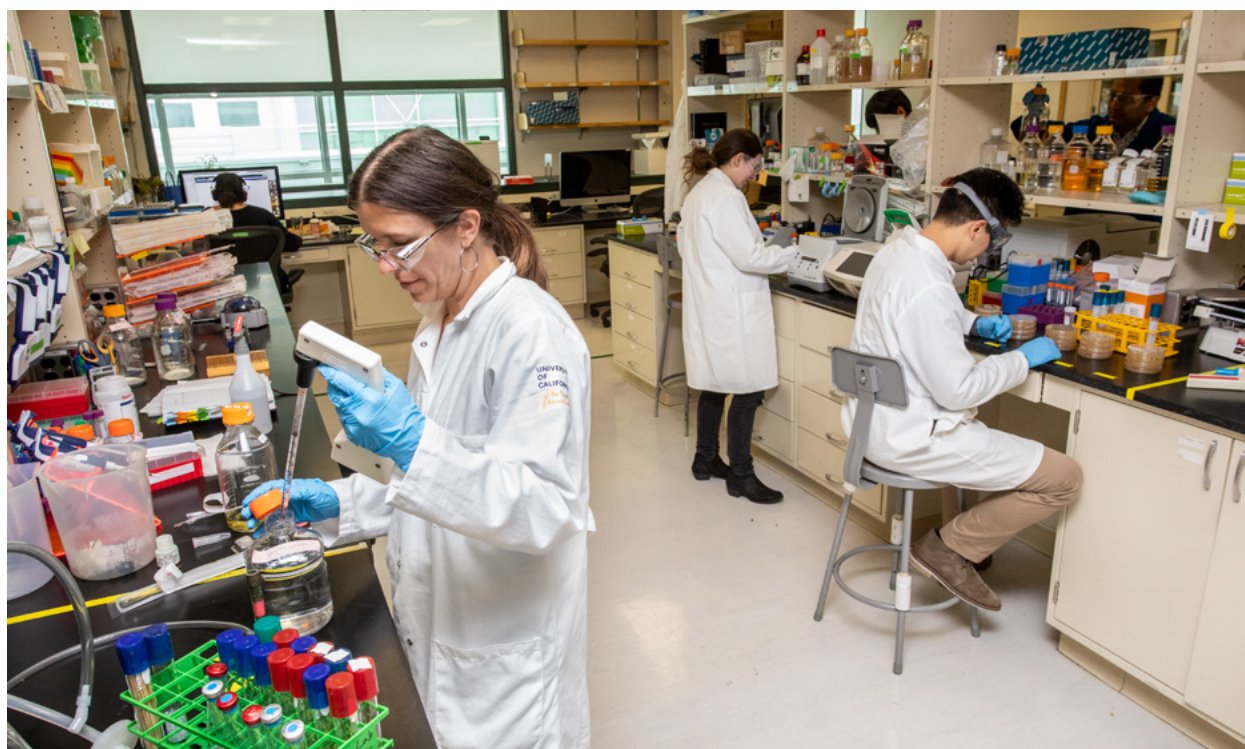
While sequence databases (e.g. those at NCBI and JGI), functional genomics databases, and modeling tools (e.g., those accessible through KBase) serve as foundational infrastructure, tools and standards for the simulation and automated design, construction, and iterative improvement of biological systems and processes remain in their infancy. For materials applications in particular, structural biology and bioimaging present as new opportunities. Computational learning frameworks and algorithms amenable to complex, multi-modal, and often multi-scale biological data are needed to discover mechanistic models from experimental data. Automation that radically scales (from tens to thousands of samples) biological system fabrication and the subsequent collection of imaging and multi-omics data, would have immediate and transformative impacts, as it would enable the application of modern machine learning and artificial intelligence algorithms to engineering biology.

Biological debugging tools

When engineered biological systems fail, it is often difficult to determine why. Beyond continually improving and innovating multi-omic, microfluidic, and cytometry methods coupled with systems biology and computational learning frameworks (e.g., mechanistic modelling as well as statistical machine learning and artificial intelligence), emerging *in situ* sensing technologies may empower us to more rapidly diagnose what went wrong and prescribe new designs to mitigate the identified deficiencies.

Retrosynthetic infrastructure and key metabolic intermediate hubs

Sophisticated retrobiosynthetic design tools enable biomanufacturing access to combinatorially large sets of target molecules. The power of this retrobiosynthetic infrastructure is compounded when synergistically combined with optimized routes to key metabolic intermediate hubs that greatly facilitate biosynthetic access to the downstream target molecules.





Designed biological systems: Engineer and scale-up predictable, controllable, trackable, robust biological systems (prokaryotes, archaea, eukaryotes, microbiomes) for key energy, health, and environmental biomanufacturing applications

Humans have an established history of modifying the natural world. Child-friendly dogs in the pet store and huge juicy carrots in the grocery store are the direct result of human will impressed upon the DNA of promising life forms. The process of selecting for pliability, safety, and utility in organisms is termed domestication, and the only apparent downside so far has been the amount of time it takes to accomplish.

Certain microorganisms have been domesticated. Bacteria and fungi involved in fermentative food, feed, beverage, fuel ethanol, and pharmaceutical production are the best understood, safest, and most manipulated microorganisms on the planet. In the last few decades, sequencing technology has begun to identify exactly what genetic changes are correlated with domestication, and genetic manipulation technology has enabled the ability to make those changes directly. Biological engineers have taken advantage of these modern tools to rapidly customize the domesticated organisms further, engineering and scaling them to produce a broader range of natural products than ever before.

However, successes have been constrained by the limited set of organisms in play. Decades of *Escherichia coli* domestication have provided an impressive array of genetic and genomic engineering tools—most of which only work in *E. coli*. When engineers plan the manufacturing of a natural product, they are limited to organisms that are well understood and genetically pliable for historical reasons but are not necessarily inherently capable of the biochemistries required for production. This has a dramatic impact on design time and difficulty, scalability and product titer, and the range of products we can actually engineer.

There is an alternative to relying solely on model organisms. Every desirable natural product comes from a living organism, often plants or microbes, some of which may be quite amenable to domestication. That is, rather than re-engineering desirable pathways

into a pliable but basically incompatible host, suitability could be engineered into already productive organisms. Adding new bacteria, archaea, fungi and plants that are generally regarded as safe to the stable of tractable hosts would not only advance the science of genetic manipulation, it would significantly broaden the range of products for biomanufacture.

If domesticating non-model organisms is prone to increase our biomanufacturing capabilities, engineering full microbial communities could catapult them. Indeed, most ecosystem services (e.g. crop productivity increase or phosphorus removal) are provided by diverse microbiomes rather than pure cultures. Furthermore, the challenge of engineering full microbial communities will force microbial ecology to move beyond descriptive approaches in order to generate the mechanistic, predictive, and actionable understanding of microbiome function that enables rational engineering. Hence, an investment in microbiome engineering provides a substantial return in practical applications as well as in scientific understanding.

Microbiology, botany, microbial ecology, synthetic biology, and process development and scale-up are Berkeley Lab strengths. For example, Berkeley Lab scientists have been central in developing CRISPR-Cas9 technologies for scalable, cross-kingdom engineering and regulation of microbes, plants, and mammalian cells. Hundreds of metagenomes and metatranscriptomes have been mined at JGI to improve our understanding of microbial community activity. Microbes have been engineered to produce a vast array of active pharmaceutical ingredients, advanced biofuels, and commodity and specialty chemicals and their intermediates. And microbes have been produced that are holistically amenable to the biomanufacturing processes of plant biomass deconstruction and raw cellulosic hydrolysate consumption. The ABPDU, in conjunction with industry sponsors, has demonstrated commercially relevant bioprocesses for the production of biofuels, bio-based chemicals, food ingredients, and materials.

The engineering of predictable, controllable, trackable, and robust biological systems for biomanufacturing applications will depend upon the identification of likely organisms, the development of domestication protocols, and the creation of novel and potentially idiosyncratic genetic toolkits. The engineering of microbiomes to a given specification will require the blend of microbial ecology, synthetic biology, mathematical modeling, high-throughput phenotyping, and automation. No institution is better poised than the Berkeley Lab to make these massive contributions to biomanufacturing.

Biodirected materials and bionanoscience: Couple biological components to chemical and physical systems to biosynthesize desired mineral/metal nanostructures

Molecular and cellular self-assembly, the process by which molecules or cells spontaneously adopt a desired arrangement without external guidance, underlies the construction of macromolecular assemblies that enable cells to function. Additionally, a growing body of evidence points to genetically encoded processes that lead to biomineralization events, such as nucleation, amorphous stabilization/transformation and polymorph selection in the formation of biomaterials.

Because of this inherent “programmability,” molecular self-assembly has also become fundamental to certain aspects of nanotechnology and mesoscale science, and there has been a recent bloom in the areas of programmable biomolecular assemblies and biodirected materials. For example, researchers have developed highly sophisticated

drug-delivery vehicles that decorate and are encapsulated by cell-mimetic materials and allow an unprecedented degree of control over the localization, specificity, timing, and dose of pharmaceuticals to specific disease sites. In another case, a viral platform has been created for biologically assembling sophisticated materials at the nano-scale such as gold and silver noble-metal wires with high aspect ratios and diameters below 50nm that can be used as cathodes for lithium ion batteries. Nucleic acids have also arisen as nanoscale supramolecular building blocks and so-called "DNA origami" can self-assemble into arbitrarily shaped 2D and 3D nanomaterials. These applications rely on the long-term development of macromolecular and viral engineering frameworks that provide a foundation for developing new molecular components in new arrangements for new applications. Just a handful of laboratories have technological expertise in using these systems. Cell self-assembly has led to genetically encoded materials. For example, diatoms engineered to deliver chemotherapeutic drugs to cancer cells, or renewable Styrofoam replacements composed of mycelium cells.

The biological processes governing the formation and assembly of genetically encoded composite materials are poorly understood. Several processes have been described, such as biologically influenced *in situ* precipitation of carbonates in the formation of microbialites, sedimentary carbonate structures created by microbes, and the synthesis of magnetosomes, specialized organelles containing iron. Indeed, it is likely that only a small fraction of the range of materials and the extent of their genetic origins have currently been discovered to date. Therefore, there is a need for discovery of new organisms, the materials produced by them, and the genetic and metabolic bases that underlie them. Genes encoding biomineralization-involved proteins have been characterized (e.g., those for magnetite and magnetosome formation). However, systematic genotype-phenotype studies have not been performed at scale. The use of iterative profile hidden Markov Models could be used to mine large genome information repositories, for example the JGI's IMG, MycoCosm, and Phytozome systems for genes involved in biomineralization and correlate these to known organisms with biomineralization functions. Such biomineralization mining approaches could be used iteratively to identify genes in unknown organisms to identify novel pathways and molecules. Needless to say, the regulatory mechanisms that govern the expression of these genes is also unknown.

We also predict that novel inorganic biomaterial genes, identified by omics-based approaches, may be useful in generating new, synthetic biomaterials by heterologous expression in non-native hosts. As synthetic biology toolboxes for broader host engineering, and expression and control systems grow, functional characterization of genetically encoded systems for biomaterials will be possible.

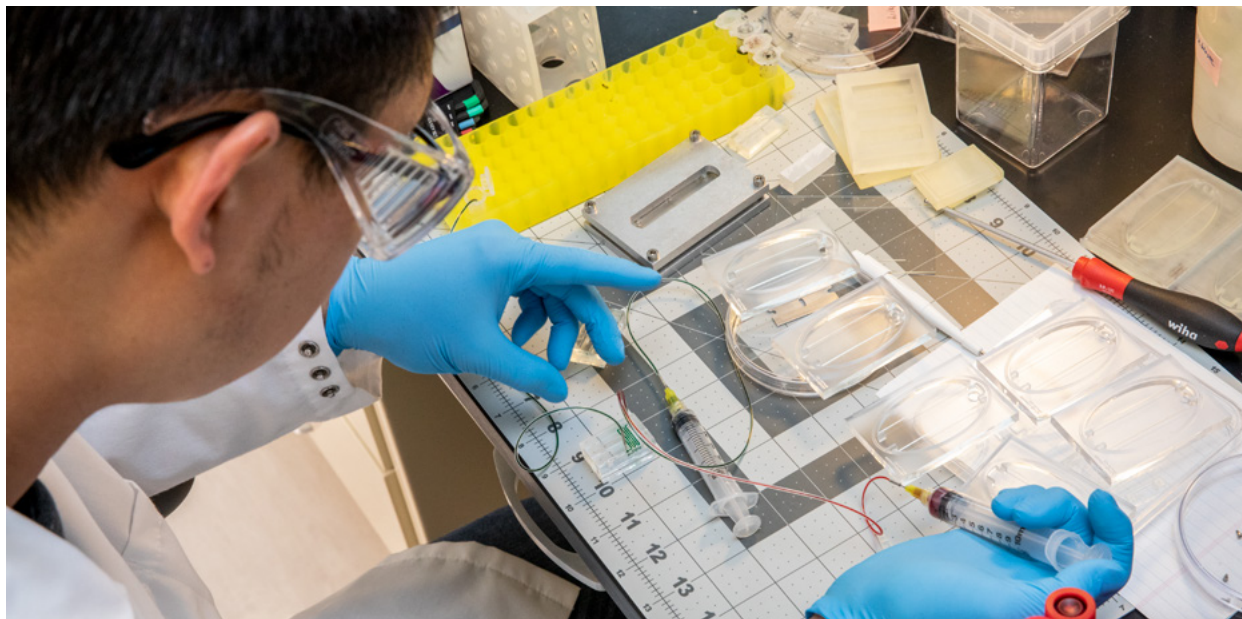
The rate of new innovations in these areas is high but the translation of the results into industrial application and scaling beyond a few highly skilled laboratories has lagged. Increased investment in macromolecular design, scaling manufacture of cell-mimetic systems, and computational and experimental methods for supramolecular assembly design will greatly enhance biodirected manufacturing capabilities. At Berkeley Lab, the intersection of biology and nanotechnology is strong and has potential to do what no other research entity can do in this area. Berkeley Lab's Molecular Foundry, which has assembled state-of-the-art tools for doing nanoscience and hosts users, has a particular strength in the integration of biology and nanotechnology. Berkeley Lab's combination of expertise in synthetic biology and nanotechnology and available tools and hardware for synthesizing and characterizing biomaterials position the Lab as a strong leader to drive national-scale scientific advances for biomanufacturing.

2023 10-year achievement measured by:

Tools to design, construct, and debug biology

- Develop a BioCAD/CAM infrastructure comprising tools for:
 - Pathway retrosynthesis and host engineering for optimized production titers, rates, and yields under industrially-relevant conditions
 - Integrating functional genomics data into the design process
 - Learning from characterization results to inform the design process
- Develop simulation and control capabilities for biomanufacturing systems:
 - Small-scale (< 2 L) physical and computational simulations of large-scale (> 100 L) reactors to aid in strain optimization
 - Sensors to measure spatially different conditions in a bioreactor, for application in the computational simulations
 - New control modalities (physico-chemical or metabolic) for real-time optimization of fermentation performance
- Develop detailed and extensible techno-economic models for manufacturing processes and applications:
 - Model-informed selection of strategies for synthesis and characterization
 - Holistic consideration of bioprocesses, from feedstock to reactor to downstream processing, integrating chemical and biological process steps, and the path to production at scale
- Develop retrosynthetic infrastructure:
 - Establish optimized routes to 30 key retrosynthetic molecular intermediates
 - Create automated workflows to test retrosynthetic computational predictions





Designed biological systems

- Establish a robust protocol for host organism domestication
 - Identify common barriers to genetic pliability and laboratory cultivation
 - Develop broad-range tools for host manipulation (i.e., plasmids, CRISPR-Cas systems)
 - Create models of physiological changes induced by domestication and scale-up of manufacture drawn from functional genomics analyses of previously domesticated hosts
 - Develop means for identifying and tracking engineered organisms in the complex environments
 - Develop means for removing engineered organisms from complex environments
 - Create models for predicting and platforms for assessing the environmental impact of engineered organisms

- Nominate a range of key hosts for biomanufacturing, agriculture, environmental remediation, water support, and human health
 - A host database serving bacterial information and protocols related to cultivation and transformation
 - Determine active and potentially tamable hosts compatible with key application environments
 - Identify host properties that can be engineered for improved robustness, productivity, and optimized energy throughput
 - Develop robust genetic toolkits for each host: expressing and integrating plasmids, media recipes and transformation protocols, characterized regulatory parts, selection and screening markers

- Establish robust tools for plant engineering
 - A host database containing genome information (size, ploidy, gene expression, etc.) and protocols related to cultivation and transformation
 - Development and characterization of “universal” plant expression tool kits
 - Generation of public part libraries and associated databases
 - Development of a pipeline for rapid and efficient trait stacking

- Develop and demonstrate bench- and pilot-scale fermentation processes for novel bioproducts
 - Quantify the impact of gene expression and metabolism on the scale-up of biomanufacturing fermentations
 - Identify, prototype, and deploy new protocols for more flexible and efficient production of precursor biomolecules and bioproducts
- Develop infrastructure to design and engineer microbiomes:
 - Tools to construct microbiomes to a given functional or compositional specification
 - Algorithms to predict expected microbiome behavior

Biodirected materials and bionanosciences

- Invent new routes for the design of biohybrid systems that mechanically or electronically interface active biological elements with polymeric and inorganic materials
 - Demonstrate electron transfer between intracellular and extracellular redox active species along a molecularly defined path
 - Create at least two molecularly defined pathways that operate at different redox potentials
- Achieve ability to interface biological components to electronic apparatus to control their activity
 - Develop methods to electronically modulate catabolic fluxes
 - Develop methods to electronically stimulate intracellular reactions
 - Interconvert electrical energy with chemical and/or light energy
- Build infrastructure to understand charge transfer at the abiotic/biotic interface
 - Develop methods to structurally characterize redox active molecules at the abiotic/biotic interface
 - Develop spectroscopic methods to characterize charge transfer and dynamics energetics
 - Create basic framework relating structure, energetic and dynamics of charge transfer
- Discover the biological basis underlying the biosynthesis of naturally-occurring inorganic materials and inorganic-organic composites
 - Develop computational pipelines for mining of genomes and metagenomes for genes involved in the transport, biosynthesis, and assembly of inorganic biomaterials
 - Build out a data platform for predictive methods to identify novel pathways and biomaterials
 - Conduct multi-omics studies, at the unicellular and nanostructural levels, on native biological systems that produce or modify inorganic materials to develop our understanding of the plant and microbial processes that direct the synthesis, transport, modification, assembly, and storage of inorganic biominerals
- Achieve ability to biosynthesize architecturally specified, possibly self-healing, mineral/metal nanostructures and mesostructures on demand, using biological entities
 - Develop computational systems biology and biodesign tools for a systems-level understanding and forward engineering of inorganic material synthesis
 - Develop synthetic biology tools for engineering biomineralization including controlling transport, synthesis, spatial patterning, and timing

- Invent new technologies to support high-throughput or massively parallel determinations of the function of inorganic material biosynthetic pathways
- Intentionally align structural and functional tools to characterize inorganic biominerals

2018 Five-year milestones for Biomanufacturing Research Strategies

Tools to design, construct, and debug biology

- BioCAD/CAM infrastructure developed comprising one or more tools each for:
 - Pathway retrosynthesis and host engineering for production titers, rates, and yields under industrially relevant conditions
 - Integrating functional genomics data into the design process
 - Learning from characterization results to inform the design process
- Small-scale (< 2 L) physical simulation of a large-scale (> 100 L) reactor developed to aid strain optimization
- Biomanufacture of two key product molecules designed, implemented, and optimized
- Biological routes to ten key retrosynthetic molecular intermediates established

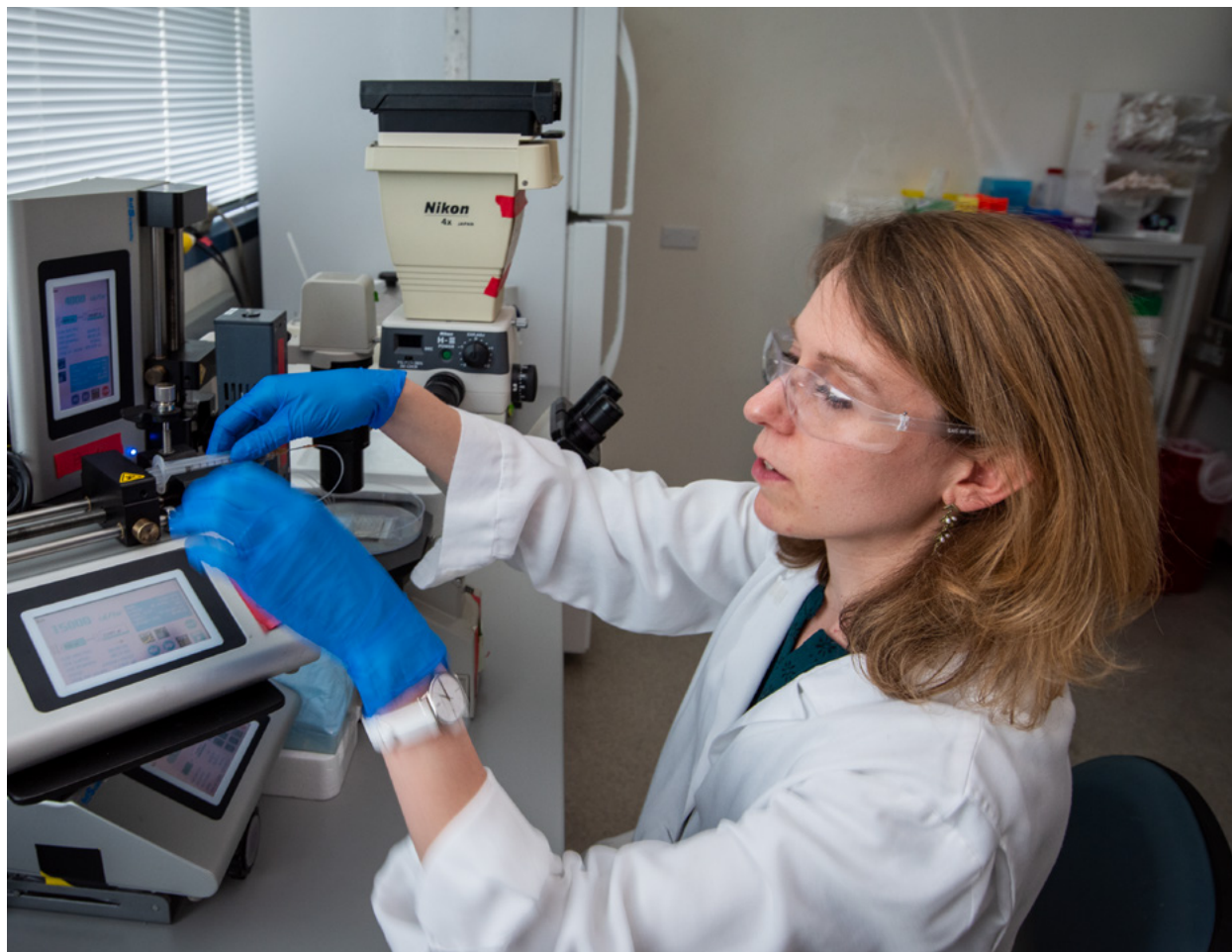
Designed biological systems

- A broad host range domestication protocol designed for use in non-model organisms
- Domesticated one previously intractable host
 - Used functional genomics to identify host systems responsive to domestication and manufacture scale up
 - Proposed a means for identifying and tracking the engineered host in different environments and assessing its impact
 - Improved biosynthesis of at least one product at previously unattainable yields
 - Demonstrate that lessons learned apply to an unrelated host
 - New bench- and pilot-scale unit operations and integrated processes developed and demonstrated for the production of a novel bioproduct
- Engineered plants that produce modified lignin that can be easily transformed into a useful commodity chemical
 - Identified aromatic precursors that can be used in metabolic pathways to efficiently produce chemicals of interests

- Manipulated plant metabolism to redirect metabolic flux toward desired products (e.g. commodity chemicals, aromatic precursors, novel monolignols)
- Engineered monolignol pathways to produce novel lignin that are efficiently deconstructed and are more readily processed to commodity chemical

Biodirected materials and bionanoscience

- Prototyped a biohybrid system that interfaces active biological elements with other chemical, physical, or electronic materials
 - Demonstrated molecularly-defined extracellular electron transfer to metals, metal oxides, and electrodes
 - Demonstrated molecularly-defined electron transfer from an electrode to intracellular species
 - Identified at least two electron transfer pathways to extracellular acceptors that operate at different redox potentials
- Structurally characterized an example of a redox active molecules that functions at the abiotic/biotic interface
- Developed a prototype multifunctional platform intended for manufacturing of two or more biodirected materials



A look ahead to 2028

Tools to design, construct, and debug biology

How to innovate and extend capabilities beyond the stated 10-year achievement Goals?

Research areas:

- Develop approaches and infrastructure to bioengineer multicompartiment organisms and multicellular organisms through engineered cellular self-assembly
- Couple engineering biology with temporal and spatial imaging and predictive design algorithms
- Extend retrosynthetic infrastructure with algorithms to predict the properties of prospective molecular targets

Designed biological systems

How to innovate and extend capabilities beyond the stated 10-year achievement Goals?

Research areas:

- Develop a predictive control system that can optimize fermentation processes in real-time
- Design and test at least three stable synthetic microbiomes to a given specification

Bidirected materials and bionanoscience

How to innovate and extend capabilities beyond the stated 10-year achievement Goals?

Research areas:

- Understand the biological basis of multicellular self-assembly so as to genetically encode different shapes of biomaterials







91



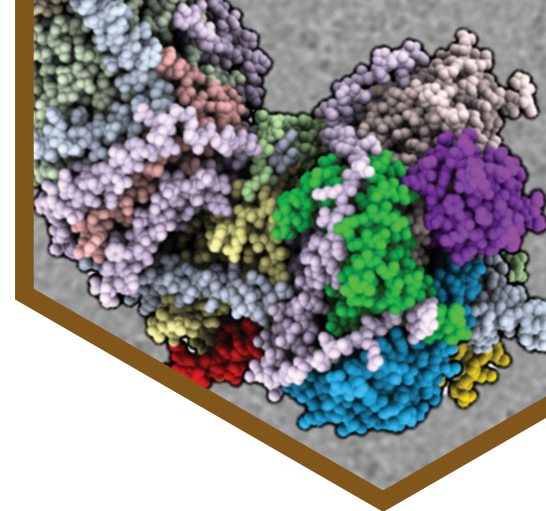
TECHNOLOGIES

10-year Goal: Develop a technology infrastructure to measure, predict and control biological systems for solving energy, environmental, and health challenges.

Background and Motivation

Science is not possible without methods for measurement, visualization, and modeling. The combination of experimental data and computation is what enables prediction and manipulation of the world around us. From the light microscope of the seventeenth century to the particle accelerator of today, the tools of science have grown exponentially in sophistication and capability. The cornerstones of the Biosciences Area technology platform are: imaging biological systems from atoms to cells to organisms; functional characterization; decoding and understanding the function of genomes; and computational methods for analysis and prediction.

Today instruments exist to view the molecular world in atomic detail, with over one hundred thousand protein and nucleic acid structures solved. But molecular structure is just the beginning of understanding and controlling biological systems. Rarely do proteins in cells work alone, or in a static configuration. Tools for kinetic analysis, imaging of larger order structures such as complexes or subcellular structures, measuring cellular physiochemistry in situ, assembly and analysis of complex plant feedstock genomes, and sequence-based analysis of the composition and function of complex environmental communities of organisms must all be brought to bear in order to gain a complete picture and true predictive power. Macromolecules are in constant motion, partnering with other molecules in a constant, complicated dance, and thus no matter the “zoom” level, no two complexes, organelles, cells, or organisms are identical. Therefore, novel methods must be developed to distinguish and quantify the similarities and differences across systems. The major challenges of this century are multiscale and



10-YEAR TECHNOLOGIES Goal

Develop technologies infrastructure to measure, understand, predict, design, engineer, and control biological systems for solving energy, environmental, health, and biomanufacturing challenges.

Technologies for Biosciences to Achieve Goal

Structural biology
Bioimaging across scales
Functional genomics
Mathematics, informatics, and computing



thus require cross-cutting approaches. No one technology can be used to single-handedly address these challenges. Application of these tools requires an integrated approach, both in instrumentation and analysis.

Berkeley Lab houses both the expertise and advanced instrumentation to make this goal a reality. Over the last two decades at the Advanced Light Source crystallography has been developed for atomic-level structural information of macromolecules, automated small-angle X-ray scattering (SAXS) to provide accurate shape information about macromolecules in solution, tomographic imaging of whole cells, and infrared spectroscopy of live biological samples. The DOE JGI provides high-quality DNA sequencing, genome assembly, analysis and functional annotation, and DNA synthesis capabilities. The ABPDU includes reactors for biomass pretreatment, controlled-environment fermentation capacity from 3 L to 300 L, and product analysis capabilities. The Molecular Foundry enables research on the nanoscale in a multidisciplinary, collaborative environment. NERSC provides large-scale, state-of-the-art computing, storage and networking for unclassified research programs. Berkeley Lab is developing a new cutting-edge electron cryo-microscopy (cryo-EM) facility that will take the high-resolution imaging capability of Berkeley Lab to new heights, enabling analysis of samples in their native environment without staining or fixation. Cryo-EM is now capable of generating structures at near atomic resolution. New resources for mass spectrometry imaging and scanning ion optical imaging will provide new frontiers for integrating imaging with functional properties of cells.

Structural biology

The workhorse technique of crystallography is considered a mature and accessible method; it has been used to delineate over a hundred thousand molecular structures to date, and has additionally benefitted from internationally accepted databases to store solved structures, as well as numerous computational programs designed to aid in structure solution and visualization. Single-particle cryo-EM is rapidly adding to the scope of what can be done by crystallography, providing structures of flexible macromolecules, very complex assemblies, and low copy-number structures that had proven to be difficult to crystallize. This technological advance, much of which was pioneered at Berkeley Lab, has truly revolutionized the field of structural biology. Advances in detectors, phase contrast, automation for sample delivery and data processing are now critical in order to keep pace with the demand for instrument time. Further advances in the method of electron cryotomography (cryo-ET) will enable even higher-order cellular structures, such as organelles, to be resolved in their native environment.

The technique of small-angle scattering (SAXS) is also approaching maturity, with numerous software analysis packages and high-throughput instruments now available. Other structural methods such as X-ray tomography and infrared microscopy are being applied to ever more diverse biological systems and they are gaining wider access by the national scientific community. These techniques inform and enhance one another. For instance, crystallographic structures are often used to help refine cryo-EM structures; scattering data are often essential in determining the molecular envelope for complexes in which crystallography data has supplied the individual subunits.

Our 10-year achievement is to link these structural methods into a unified platform in both shared instrumentation and in data analysis programs and knowledge databases to gain a full picture of the conformational and functional flexibility of biological components. As has been proved with the crystallographic databases, as more data is shared more widely, more science

is enabled more broadly. In the case of protein crystallography, access to a greater number of previously solved structures gives a higher probability of success for future structures.

The first step toward a more integrated infrastructure is the co-utilization of techniques, beginning with the establishment of a cryo-EM resource for Berkeley Lab researchers. Within the synchrotron facility, instrumentation will be developed to pair methods in which data can be collected simultaneously, as in the case of scattering and footprinting, or spectroscopy and high-resolution structure determination. Further advances will be made in beamline development for crystallography, such as the buildout of the Gemini beamline, which integrates advanced methods in sample handling and data processing to provide exceptionally high throughput. For samples requiring different preparation states, such as X-ray tomography and infrared spectromicroscopy, for example, user programs will be more integrated in order to enable serial data collection and collaborative data sharing. This will be especially important for leveraging new capabilities as the ALS transitions to ALS-U, a diffraction-limited light source with even higher brilliance in the X-ray regime. In the area of data integration, common formats will be developed in order to share cross-technique structural information and to enable teams of scientists with diverse areas of expertise to communicate effectively about a common problem.

Bioimaging across scales

All biological systems display unique behaviors, including self-organization across temporal and spatial scales ranging from atoms to organisms and ecosystems. Visualization provides, perhaps, the most powerful mechanism for understanding a range of behavior, from how cells function at a molecular level to how communities are formed through the interaction of cells and organisms with each other and with the surrounding environment. For example, the response of microbial communities to environmental stress impacts the cycling of elements such as carbon and nitrogen. Information about the composition of these communities and their physical interactions can come from the development of advanced imaging approaches. In higher organisms, brain cells operate within a complex microenvironment comprising multiple connections and signaling systems, which are altered by environmental toxicants. The responses, however, depend on which network



connections are targeted, which neurotransmitters are affected, how fast they respond, the pathways that regulate neurotransmission, and blood-gut-biome communication.

Visualization requires zooming out to an entire organism or zooming in the focus on a molecule within a single cell. Despite obvious complexity in biological systems, most advances in bioimaging technologies have been focused on a single modality, which often fails to adequately address the spatial and temporal realities of biological systems, their cell-cell communication, and their ability to respond to challenges. Berkeley Lab plans to address these challenges through advances in instrumentation including spectromicroscopy and integrated computational strategies for large and diverse data arrays. Advances will build on available expertise and instrumentation for imaging cells, tissues and organisms, and novel probes based on chemistries and sensors. Fluorescent probes, radioisotopes, and electromagnetic radiation provide signals to observe molecules within biological samples. Non-invasive methods use molecular vibrations as the basis for visualization of living cells in their natural states. Multimodal imaging combined with artificial intelligence and computational methods integrate different modes of visualization across a range of length, time, and resolution scales.

Functional Genomics

Functional genomic technologies are central to all of the scientific research strategies in the BSP including understanding the functions of genes in diverse species, dissecting interactions within microbial communities, and enabling synthetic biology approaches for sustainable energy and materials. These functional genomic technologies will also supply rapid feedback on the performance of engineered and natural biological systems providing including detailed insights into the genetic and physiological states. These data can then be used in conjunction with structural biology and bioimaging data using our developing mathematics, informatics and computing capabilities, to understand and then generalize the molecular mechanisms controlling important biological processes, from sequence to function within specific spatiotemporal contexts.

A molecular understanding of organismal interactions, for example microbe-microbe and microbe-host interactions, is necessary for predictably harnessing microorganisms for beneficial energy and environmental outcomes. To enable this understanding, Biosciences is coupling multidisciplinary experimental approaches including metabolomics with computational modeling to uncover the functions of genes and small molecules that mediate organismal interactions in diverse environments including sediment, groundwater, soil, and higher eukaryotes including the rhizosphere.

Rapid advances in DNA sequencing have revealed millions of genes, many of which are predicted to encode for proteins. However, most of these proteins are of unknown function because they are too distant from an experimentally characterized protein to be accurately annotated by sequence similarity. To meet this challenge, Biosciences is developing experimental technologies, including high-throughput genetics, proteomics, and metabolomics, to bridge this gap between sequence and function. Importantly, these tools are being developed such that they can be flexibly applied across diverse taxa at low-cost. This will enable Biosciences to rapidly discover new protein functions and increase the accuracy of computational gene annotations.

Metabolomic analysis provides a direct measure of the activities and functions of biological systems. This makes it a powerful complement to DNA sequencing and is

central to understanding biological processes relevant to the Biosciences Strategic Plan. Currently, coverage in metabolomics experiments is low compared to other omics approaches, especially from complex samples (e.g. soil, one-pot fermentations). To address this gap, Biosciences is developing integrated experimental and computational approaches to both dramatically increase the number of metabolites routinely identified and explore yet-unknown metabolic processes.

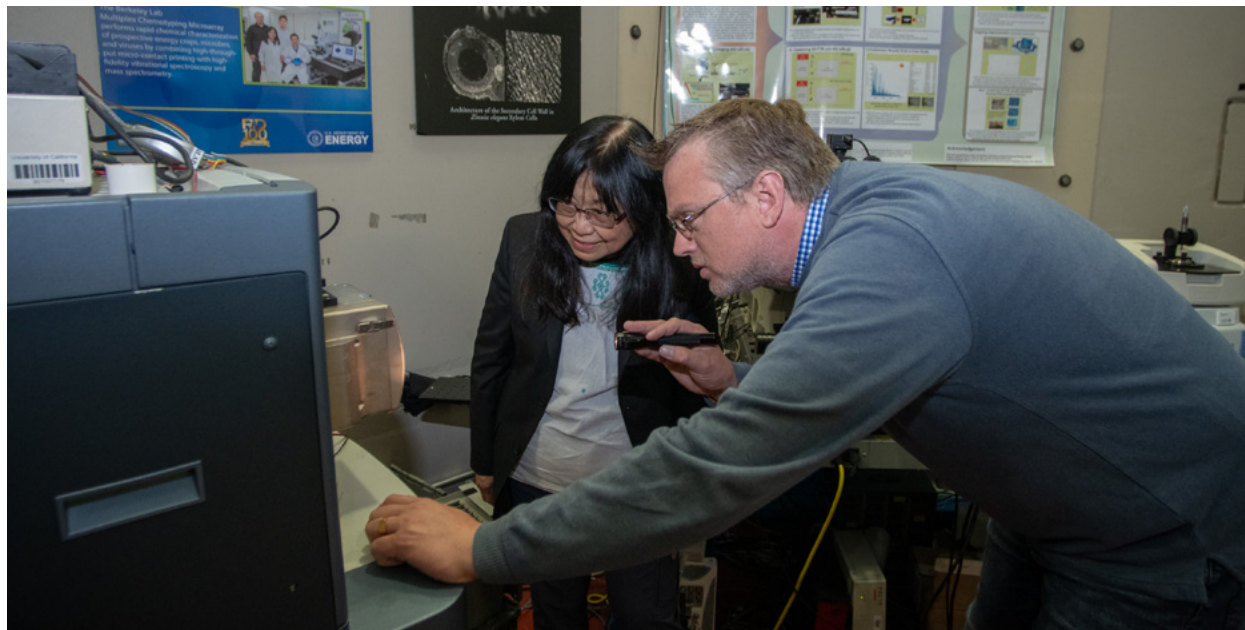
Overall, our functional genomic strategy will develop and integrate important new enabling technologies for use by large-scale research teams to achieve these 10-year Goals. The major technology thrusts in our strategy are: 1) rapidly advancing our functional understanding of genes, proteins, metabolic pathways, and metabolites; 2) deconstructing activities and interactions within complex cellular communities; 3) improving the performance of metabolomics; and 4) using imaging- and spatial-defined omics to measure complex biological processes at relevant scales and environmental contexts.

Mathematics, informatics, and computing

Mathematics, informatics, and computing play a central role as biology continues to evolve from an anecdote-based, descriptive field of science into a quantitative discipline relying on—and driven by—large data sets. Advanced data analysis tools are the foundation of modern genomic analyses, drug development, and a fundamental understanding of cells, tissues, and microbial communities. The Biosciences Area has a long track record in coupling biological experimentation with analytics, having developed a number of key technologies in a variety of disciplines that are in use around the globe. Various entities within the Biosciences Area, such as the JGI, KBase, the Computational Crystallography Initiative, as well as a number of other teams, are developing new computational methods to advance biosciences research across a wide variety of sub-disciplines.

To ensure that JGI can fulfill its mission of advancing energy and environmental sciences, strategic development directions at JGI in the computing arena include increases in algorithmic and pipeline efficiency, and moving towards an exascale-aware configuration of hardware and software solutions. KBase provides an open-science data-sharing platform for collaborative analytics that encompasses access to public database repositories, public JGI data and tools, and high-performance computation through NERSC. The KBase platform enables the scientific community to explore, model, and predict interactions of genes, proteins, organisms, and their environment, ultimately allowing researchers to understand, engineer, and control biological systems from the micron scale up to ecosystems.

Future developments will continue to leverage the strong Berkeley Lab computational environment. High-performance computing solutions will be provided by NERSC to the scientific community, supported by the Energy Science Network (ESNet), which provides a high-performance unclassified network built to support scientific research, providing computer network capabilities to over 40 DOE research sites, including all national laboratories, several supercomputing facilities, and major scientific instruments. Efforts in the Center for Advanced Mathematics in Energy Research Applications (CAMERA) are focused on the development of novel mathematics and algorithms that either significantly improve current approaches or enable the analyses of novel experimental techniques. Finally, the Computational Biosciences Group is a collaboration between the Computing Sciences and Biosciences Areas that combines data analytics and statistical machine learning with theory and mechanistic models enabled by high-performance computing towards addressing our nation's energy, environment, and health needs.



Research in mathematics, informatics, and computational sciences are an integral part of the technology strategies. Methods will be developed to extract maximal information from weak data, enabling the analysis of experiments from free-electron lasers, electron microscopes, super resolution microscopy and other experimental modalities. A key advance will be new algorithms that integrate experimental information of multiple types. Integration of experimental data with simulations will be critical for bridging the several orders of magnitude in scale that separate the smallest and fastest processes and the largest and slowest processes. Conceptually, we must be unyoked from the degeneracies of phenomena at atomic scales to study emergent properties at meter scales. Finally, the most significant breakthrough will be methods that enable automated knowledge generation from complex data types, constructing mathematical models that explain a wealth of experimental observations. The development of novel statistical machine learning methods that lead to interpretable models from which a predictive understanding can be abstracted will be essential.

2023 10-year achievement measured by:

Structural biology

- Constitution of methods for correlating and analyzing structural and functional data from the molecular to the cellular level
- Creation of computational analysis programs capable of data structure input from synchrotron X-ray and electron microscopy methods
- Development of methods to integrate advances in functional genomics with structural biology methods

Bioimaging across scales

- Establishment of a center for integrated bioimaging

- Availability of probes, labeling chemistries, and label and label-free approaches that provide contrast across multiple imaging modalities
- Creation of visualization, modeling, and interaction systems for experimenters to efficiently extract knowledge from data
- Development of new sample preparation methods compatible with multi-model imaging

Functional Genomics

- Demonstrate technologies for rapid phenotyping, especially microfluidic chip-based automation and mass spectrometers with enhanced analytical chemistry
- Vastly improve performance of metabolomics for energy, materials, and environmental research
- Develop and apply at-scale, new genetic approaches to determine the function of genes from diverse organisms (bacteria, archaea, viruses, fungi, algae, plants)
- Pioneered community editing technologies (CRISPR, phage, optogenetic, etc.) to determine the functions of cells, organisms, proteins, pathways, and metabolites within specific environmental contexts
- Develop, integrate, and apply new nondestructive modalities with functional genomic measurements to gain insights into functional genomic processes within relevant spatiotemporal contexts and scales
- Develop isotope-based methods, coupled to systems biology tools, for *in situ* characterization of important metabolic processes
- Demonstrate utilization of genetically targeted, *in vivo* cellular manipulation techniques (e.g., optogenetics) to control biological function

Mathematics, informatics, and computing

- Improve mathematical and computational approaches for addresses challenging problems in structural biology, genomics, bioinformatics and multiscale modeling
- The development of novel mathematical and computational approaches enabling new scientific approaches or solving outstanding problems in biophysics, biochemistry and genomics
- In collaboration with NERSC and the Computing Research Division, redesign existing computational approaches that can make optimal use of novel hardware, such as exascale computers
- The development of mathematical and algorithmic approaches that allows the joint analyses of multimodal methods, operating on length scales from atoms to cells to organisms

- Innovation of dynamical systems approaches to understand biological time-series data towards revealing the governing equations of biological systems
- Creation and application of statistical machine learning methods to extract small sets of features predictive of complex biological phenomena to enable understanding and causal interventions
- An increase in the integration of data analytics to drive the design and adaptive control of experiments in the Biosciences

2018 Five-year milestones for Technologies Strategies:

Structural biology

- Establishment of an LBNL cryo-EM resource to support biosciences, geosciences and soft matter research
- Integration of scattering and diffraction structural methods at the Advanced Light Source
- Advanced automation in crystallography and microscopy methods

Bioimaging across scales

- Multiple imaging modalities applied to imaging biological systems from the nm to mm length scale and the msec to hour time scale
- New light microscopy technologies, including new contrast probes and labeling chemistries, integrated into existing imaging systems
- Model-based algorithms developed for combining information across multiple imaging methods and integrating functional data
- Collaborations with Berkeley Lab mathematics, informatics and computing researchers result in big data analysis methods applied to several challenging biological problems
- Establish a virtual institute to share methodology and computational approaches across scales

Functional Genomics

- Demonstrated effectiveness of an automated chip-based mass spectrometry platform with enhanced throughput and analytical chemistry capabilities
- Demonstrated advanced technologies for high-throughput functional genomic analyses that are tightly integrated with computational resources
- Applied comparative gene expression coupled with metabolomic, proteomic and fitness data for biochemical discovery and testing of genomic predictions
- Demonstrated technologies to identify active metabolic pathways within complex multicellular systems

Mathematics, informatics, and computing

- The development of novel data analysis methods for existing methods such as X-ray crystallography, Small and Wide Angle X-ray Scattering (SAXS / WAXS), X-ray tomography
- The development of new applied mathematical and algorithmic methods for the analysis of the emerging methods that make use of X-ray Free Electron Lasers (XFEL), near atomic cryo-electron microscopy, electron tomography, and correlated X-ray scattering
- The development of new methods that bring together genomic, imaging, and functional genomic data types across multiple resolution ranges through computational models, in the process generating new knowledge of biological systems
- Prediction of biological activities, at the atomic, cellular and organismal level, through the use of integrated analysis and computation

A look ahead to 2028

Structural biology

How does use of X-ray and electron-based techniques become routine for genome annotation of biological systems?

- Greatly improve throughput for electron microscopy imaging of molecules
- Develop pipelines that link genomics, gene-synthesis, proteomics, and light source techniques
- Integrate many data types coupled with structurally informed analysis tools in KBase

Bioimaging across scales

A mechanistic understanding of multiscale phenomena in cellular systems will require:

- Sample preparation technologies that enable multiple imaging modalities
- Streamlined, high-throughput data collection technologies
- Visualization, modeling, and interaction systems for experimenters to efficiently extract knowledge from multimodal data

Functional Genomics

How do specific genes, biopolymers, and metabolites determine the spatial assembly of microbes?

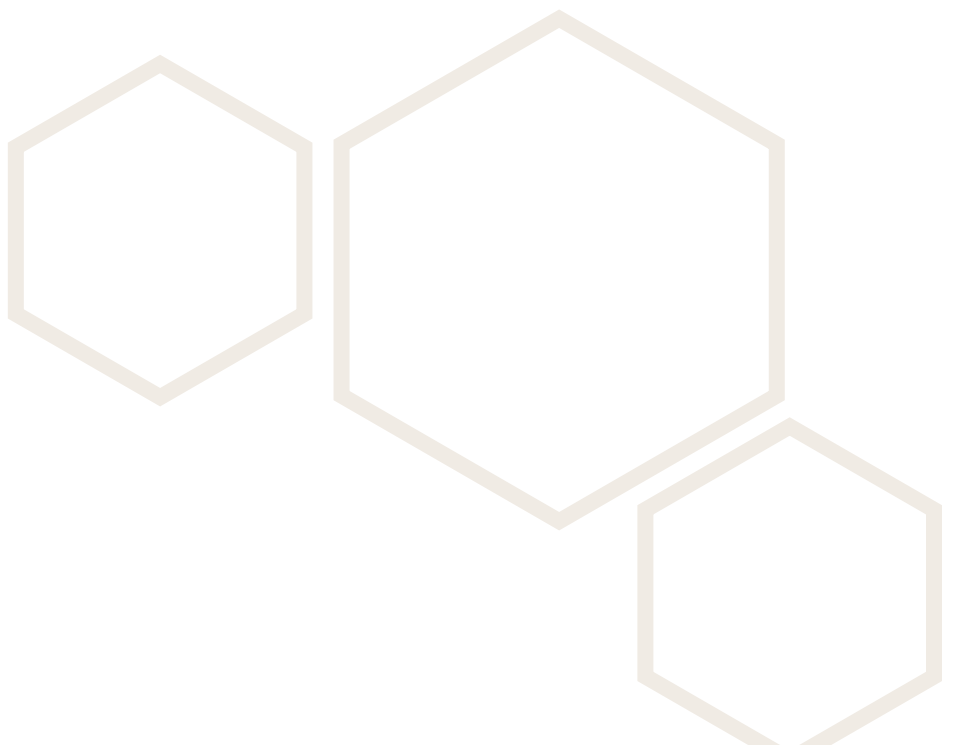
- Develop high-throughput genetics and omics approaches to greatly improve our understanding of microbial genomics
- Develop advanced technologies that enable spatially defined measurement of microbes, biopolymers, and metabolites

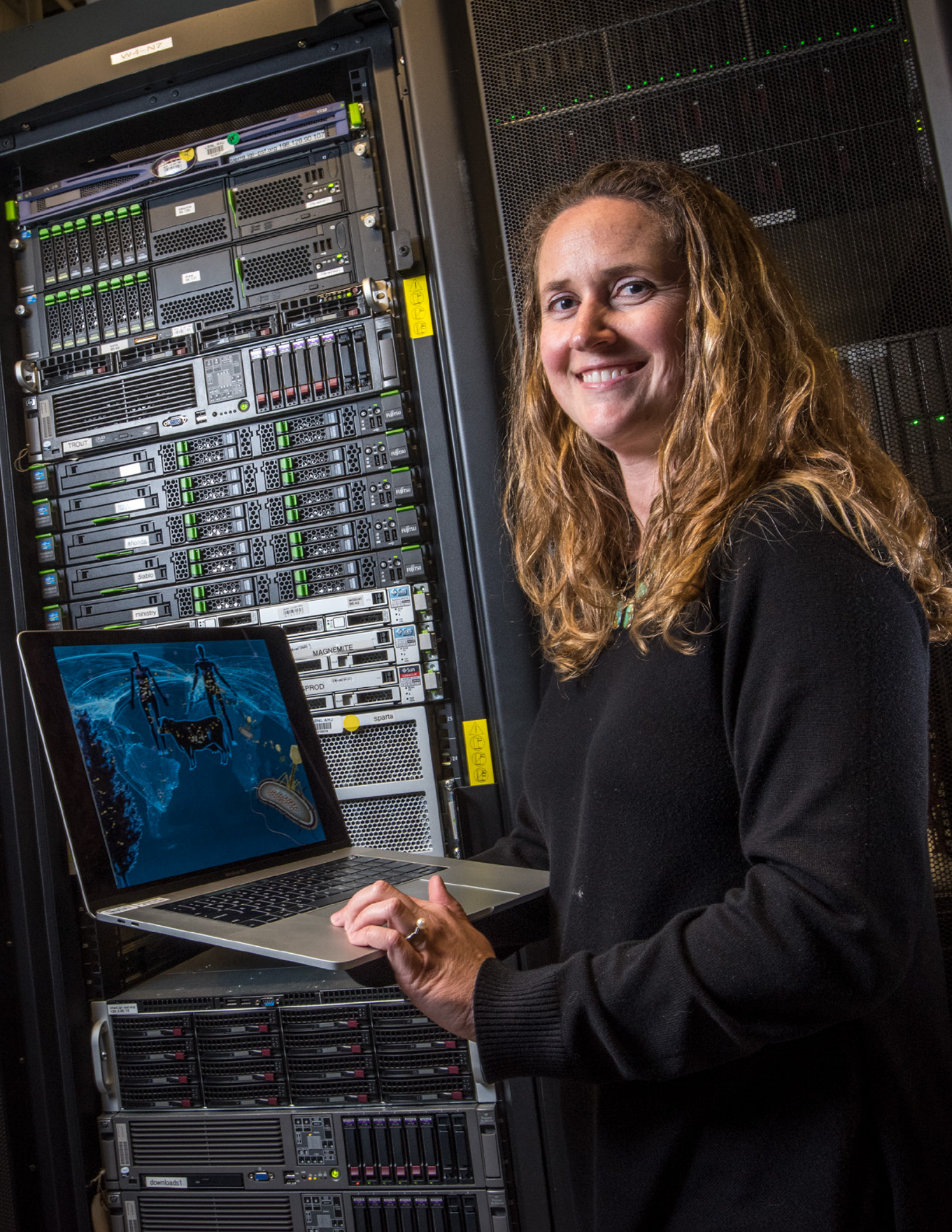
- Develop the ability to edit microbial communities *in situ* to deconstruct molecular mechanisms

Mathematics, informatics, and computing

How can we apply computational modelling approaches that enable prediction of complex behaviors in biological systems?

- Develop mathematical and algorithmic approaches for the joint analyses of multimodal methods
- Create new methods to determine the governing equations of biological systems from time-series data
- Create and apply statistical machine learning methods to extract predictive features of complex biological phenomena





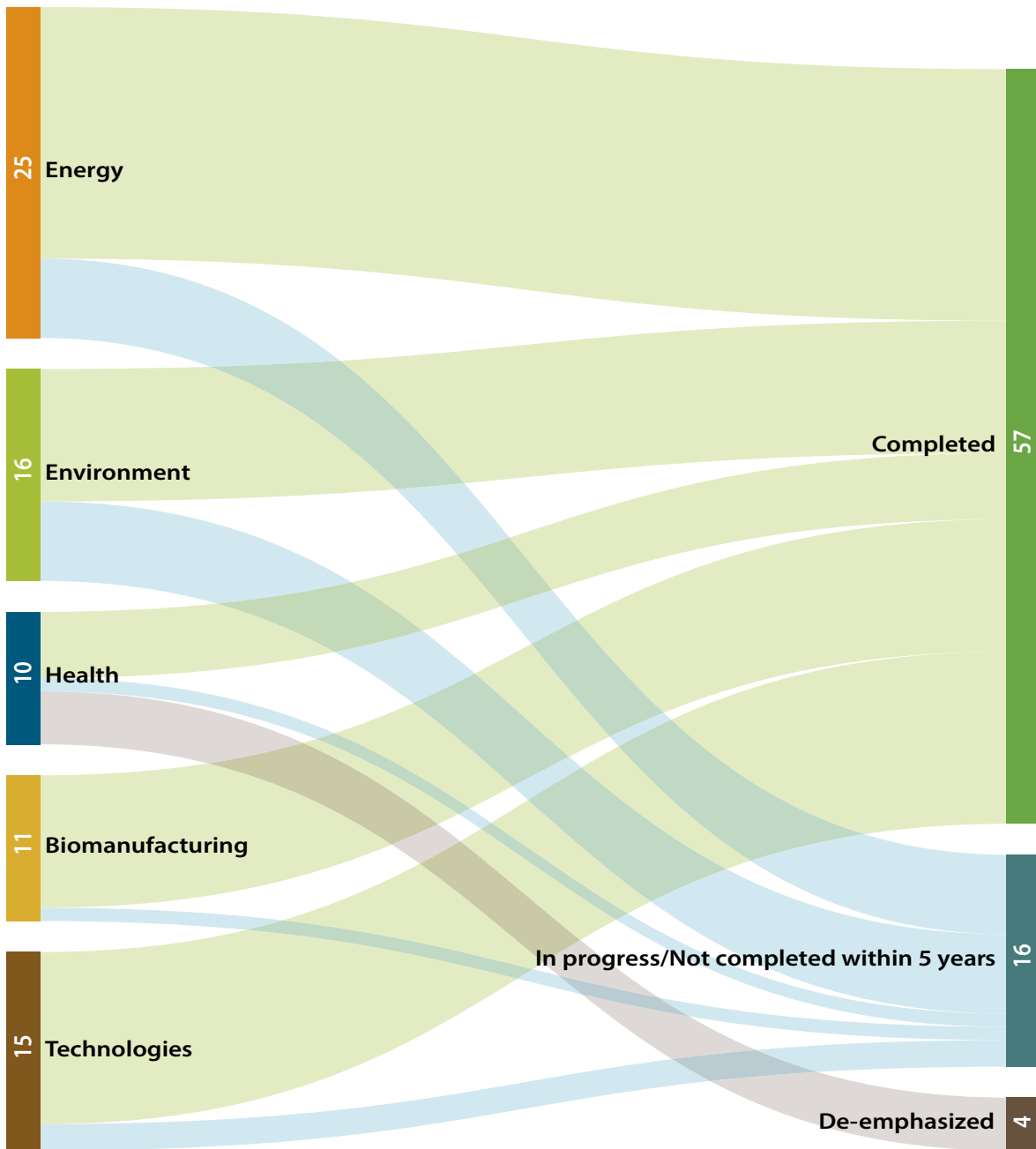
FIVE-YEAR PROGRESS REPORT SUMMARY

In 2018, Biosciences assessed its progress towards meeting the five-year milestones outlined in the Biosciences Strategic Plan. Biosciences staff were asked to determine if milestones were completed based on a variety of metrics appropriate for each strategy. In many cases, completion was evaluated by publication of scientific results related to the milestone. For other milestones, the establishment of a new research program or strategic personnel hire was considered as a metric for completion of the milestone. These assessments were compiled into the Biosciences 10-year Strategic Plan 5-year Progress Report that described significant accomplishments towards meeting the five-year milestones. During this process, Biosciences staff determined that 74% of the milestones were met and another 21% were in progress but had not yet been completed. Only 5% of the milestones were not on track for eventual completion and were de-emphasized for implementation of the plan.



This progress report is a demonstration of Biosciences' commitment to implementation of our science strategies and tracking the metrics for realizing those aspirations initially developed in 2013. In the six years since the publication of the first Biosciences Strategic Plan, the Area has committed to program development activities to build new research programs that put scientists on track to meet the 10-year Goals. Some of these new research programs, such as the Agile BioFoundry, the Microbial Community Analysis and Functional Evaluation in Soils (mCAFES), Trial Ecosystem Advancement for Microbiome Science (TEAMS), and National Microbiome Data Collaborative (NMDC) have been and will be critical for meeting the metrics laid out for the next four years. In addition, Biosciences completed a comprehensive scientific and operational reorganization in 2015 that organized researchers based on scientific disciplines and research topics, building new collaborative networks to advance bioscience.

5-Year Milestone Assessment



The Biosciences Area’s progress towards completion of 5-year milestones by strategic goal and by milestone status after 5 years of implementation.

Left, number of 5-year milestones for each strategic goal.

Right, number of 5-year milestones that were completed (green), in progress but not completed within 5 years (blue) or de-emphasized (brown).

BIOSCIENCE CAPABILITIES

Berkeley Lab's integrated biosciences program benefits from the expertise of a large staff of leading researchers, access to world-class facilities, and the organizational strength of the Laboratory's Divisions and affiliated research institutes. Capabilities are vast, and span Biosciences Divisions, mission-focused centers, national user facilities, and the synergistic teams that collaborate on national-scale biosciences research efforts.

Berkeley Lab Biosciences Division

Biological Systems and Engineering Division—Advances mechanistic and predictive understanding of complex biological systems as a function of environment and scale and then translate this knowledge using engineering principles to advance resilient and efficient solutions in synthetic biology, oncology, agriculture, biofuels, renewable chemicals, materials, and abiotic-biotic systems.

Environmental Genomics and Systems Biology Division—Links genome biology to ecosystem dynamics through deep understanding of the genetic and molecular mechanisms governing the activities and ecology of organisms and multispecies communities to predict responses and harness microbes and plants for energy and environmental solutions.

Molecular Biophysics and Integrated Bioimaging Division—Studies the fundamentals of biology with the goal of learning how to manipulate, control, and create biological functions in order to solve national challenges in energy, environment, health and biomanufacturing. By using advanced imaging technologies that probe the structure, chemistry, physics and dynamics of complex systems the secrets of critical biological processes are being revealed.

Mission-Focused Efforts

The **Joint BioEnergy Institute** (JBEI) is one of four Bioenergy Research Centers created by the DOE in 2007 to advance the development of transportation fuels from lignocellulosic biomass. Key capabilities at JBEI include: basic gene discovery in plants, microbes, and microbial communities; process development for cellulose extraction from biomass; engineering fuel synthesis in microbes; and synthetic biology/biodesign.

The **Joint Center for Artificial Photosynthesis** (JCAP) is the nation's largest research program dedicated to the development of an artificial solar-fuel technology. Established in 2010 as a DOE Energy Innovation Hub, JCAP employs capabilities in physics, chemistry, materials science, and nanotechnology to find a cost-effective method to produce liquid fuels using only sunlight, water, and carbon dioxide. JCAP is led by the California Institute of Technology with Berkeley Lab as its lead partner.

The **Systems Biology Knowledgebase** (KBase) is an extensible and scalable open-source software framework and application system to support the analysis of microbes, microbial communities, and plants. KBase offers free and open access to data, tools, and models, helping scientists and researchers integrate various data types to build new knowledge and share their findings with others. KBase is designed to connect relationships across all shared and public reference data with the goal of generating multiple lines of evidence to support community-driven inferences of biological function.

The **Agile BioFoundry** (ABF) is a multi-lab effort created by the DOE in 2016 to enable the realization of a thriving national bioeconomy through advanced biomanufacturing. At the ABF, product development is conducted in an automated design-build-test-learn integrated fashion that generates robust organisms, genetic circuits, and conversion pathways tunable to industry demand, with predictable process scaling and potential downstream process development impacts addressed at the outset.

The mission of the **National Microbiome Data Collaborative** (NMDC) is to empower the research community to harness microbiome data exploration and discovery through a collaborative integrative data science environment. To accomplish this the NMDC seeks to address fundamental roadblocks in microbiome data science, including 1) implementing FAIR data principles (making data findable, accessible, interoperable, and reusable); 2) connecting data resources and compute resources; and 3) developing a framework for community engagement that supports open science and shared ownership. The NMDC will enable advanced data analysis and tool development, enhancing cross-study comparisons, and ensuring that new and updated information can be incorporated.

ENIGMA, Ecosystems and Networks Integrated with Genes and Molecular Assemblies, seeks to understand how microbes and microbial communities impact their ecosystems through detailed field and laboratory studies, with an ultimate goal of a mechanistic interpretation of microbial ecology. ENIGMA researchers study environmental bioprocesses and their impact on cellular fitness under different conditions to link microbial processes to different ecological theories to explain the observed differences in identity and diversity of microbes along transects (gradients) of geochemical parameters.

The **Microbial Community Analysis and Functional Evaluation in Soils** (mCAFES) project develops complementary approaches to precisely edit microbiomes with foundational tools and molecular ecosystems biology-based understanding required to predict, manipulate, and design rhizosphere microbial communities. mCAFES seeks to interrogate the function of soil microbiomes with critical implications for carbon cycling and sequestration, nutrient availability and plant productivity in natural and managed ecosystems. The project targets molecular mechanisms governing carbon and nutrient transformation in soil, with a focus on microbial metabolic networks, developing functional genomics tools to explore and understand ecosystem biology.

Berkeley Lab's **Computing Sciences Area**—consisting of the National Energy Research Supercomputing Center (NERSC), the Computing Research Division (CRD), and the Energy Sciences Network (ESnet)—similarly provides both the infrastructure and the opportunity for collaborative relationships. NERSC and ESnet are also designated DOE user facilities.

National User and Collaboration Facilities

The **Advanced Light Source** (ALS) is one of the world's premier synchrotron radiation facilities and hosts hundreds of investigators from the biosciences community every year on its beamlines. Capabilities of the ALS especially valued in biosciences research include macromolecular crystallography (MX) for atomic level structural information of proteins and nucleic acids singly or in complexes, automated small-angle X-ray scattering (SAXS) capabilities that provide accurate shape and assembly information about functioning macromolecules in solution, tomographic imaging of whole cells (the National Center for X-ray Tomography), and infrared spectroscopy of live biological samples (the Berkeley Synchrotron Infrared Structural Biology Program). Collectively these resources are essential to research carried out by Biosciences scientists.

The Department of Energy **Joint Genome Institute** (JGI) is the world's leading user facility for energy and environmental integrative genome science. The JGI's programs focusing on microbes, fungi, algae, plants, metagenomes, metabolomics and DNA synthesis science enable an international user community to derive biological insights that can be used to harness biology for societal impact at an unrivaled scale and complexity.

The **Advanced Biofuels and Bioproducts Process Development Unit** (ABPDU) is a state-of-the-art user facility for testing and developing emergent biofuels technologies. It was funded by DOE to allow laboratory-scale processes for fuels synthesis to be scaled-up and commercialized. The facility includes reactors for biomass pretreatment, controlled-environment fermentation capacity from 3 L to 300 L, and product analysis capabilities. The 15,000-square-foot facility is available to companies, Bioenergy Research Centers, DOE-supported researchers, academic institutes, and nonprofit research organizations involved in biofuels and bioproducts R&D production.

The **Molecular Foundry**, one of five DOE nanoscale science research centers (NSRCs), provides users with access to expert staff and leading-edge instrumentation to enable research on the nanoscale in a multidisciplinary, collaborative environment. Selected through an external peer-reviewed proposal process, nearly 700 users come each year from academic, industrial and national laboratories, both domestic and international, free of charge. Research that results in over 300 publications each year is organized into seven closely coupled facilities: Inorganic, Organic, and Biological facilities for material synthesis, preparation, and assembly; Nanofabrication, for processing and integration; the National Center for Electron Microscopy and Imaging and Manipulation, for characterization; and Theory, for understanding and predicting material properties.

The **National Energy Research Scientific Computing Center** (NERSC) is the mission high-performance computing facility for the Department of Energy's Office of Science (DOE SC); it has about 6,000 users from universities, national laboratories, and industry. NERSC's primary goal is to accelerate scientific discovery at the DOE SC through high performance computing, related technology development and data management and analysis. Toward this end, NERSC provides large-scale, state-of-the-art computing, storage and networking for DOE SC's unclassified research programs in alternative energy sources, climate change, energy efficiency, environmental science, and other science areas.

The **Energy Sciences Network** (ESnet) is the ultrafast data highway for all DOE national laboratories. If supercomputers like NERSC provide the horsepower for data-intensive science, ESnet provides the connectivity. Large-scale collaborative research is the heart

and soul of the modern scientific enterprise. Researchers today share data sets in the petabyte range—a million times the size of files familiar to consumers. The network links tens of thousands of researchers at more than 40 institutions, at high speed and securely. ESnet engineers are developing a new network technology that will boost data transmission rates to 100 gigabytes per second—ten times faster than today's.

Collaborative Research and Resources

Biosciences research teams at Berkeley Lab have an exceptional history of productive collaborative interaction. For example, the DNA sequencing capabilities at DOE JGI are used to support JBEI and ENIGMA, as well as environmental research by EESA. JBEI is also a user of the ABPDU for development of novel biomass deconstruction methods. DOE JGI is partnered with KBase to develop and serve analytical tools and public data to scientists studying DOE-relevant problems. Fabricated ecosystems, from the bench-scale (EcoFABs) to the phone booth-size (EcoPODs), allow scientists to reconstruct real-world environmental interactions in laboratory settings and transfer their findings to field-scale research. The community of EcoFAB users outside of Berkeley Lab is growing and ring trials have demonstrated that experiments are reproducible across states, countries, and continents. In 2019, Biosciences expanded its computational capabilities, including machine learning for biology, to apply new mathematic, statistical, and computational methods to problems ranging from production of new bioproducts in microbes to understanding application of microbes in agriculture for academic and industry partners. Since 2013, Biosciences has expanded its interactions with industry in order to better translate and transfer fundamental scientific discoveries to the U.S. innovation ecosystem. The ABF, ABPDU, JBEI, and JGI all interact on a regular basis with industry members, through formal industry engagement programs that seek to match the unique capabilities of the national laboratories with companies that have challenging problems that could be solved by the expertise of researchers and the resources at the labs.

Expertise and Tacit Knowledge

The collective experience of Biosciences, Berkeley Lab, and other national lab staff is an important resource for knowledge and expertise that is challenging to transfer through traditional methods of dissemination of scientific results. Data, inferences, and knowledge generated from research studies are often shared through peer-reviewed publications. However, most of these publications do not contain raw data and are devoid of the know-how associated with experimental setup and execution. Sharing a comprehensive description of the experimental setup is necessary for biological studies due to the multitude of variables, including those like the altitude of the laboratory location, that can affect the outcome of a particular experiment. This gap leads to poor replication of experiments and the associated inferences. Also, with limited opportunity to scour all publications in their respective areas, researchers design experiments with either less relevant or well-studied variables, leading to a rediscovery of inferences. To avoid loss of precious resources for research in irreproducible or rediscovery-oriented studies, expertise generated from individual research studies should be pooled and disseminated in the most comprehensive manner. Knowledge gathering and dissemination is essential, particularly for studies where expense limits replication and where private-sector innovation is hindered from a lack of public source of data. Proactively and continuously designing and improving databases to share pooled expertise with researchers is essential to guiding them through thought-experiments that empower them in designing best possible research studies.



Biocampus

During the next 10 years, Berkeley Lab understands that new collaborations and capabilities will emerge. To enhance collaboration among Bioscience research teams, we plan to relocate Biosciences researchers to the Berkeley Lab campus. The first building, the Integrative Genomics Building, was completed in 2019 and is the new home of JGI and KBase, collocating together these two complementary and synergistic DOE programs. Berkeley Lab is actively planning for a second building, the Biological and Environmental Program Integration Center (BioEPIC), that will bring together researchers across disciplines to investigate key biological processes in an environmental context, bridging laboratory and field studies through the use of fabricated ecosystems. A proposed third building is envisioned as a future home for Biosciences' biofuels and bioproducts programs including JBEI, the ABF, and the ABPDU to advance bioenergy and biomanufacturing research from fundamental proof-of-principle to industrially-relevant scales of production.

