

# Biosystems Design

Report from the July 2011 Workshop

Convened by  
U.S. Department of Energy  
Office of Science  
Office of Biological and Environmental Research



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

Office of Biological and Environmental Research

# Biosystems Design Workshop

July 2011

Convened by

**U.S. Department of Energy**  
**Office of Science**  
**Office of Biological and Environmental Research**

**Co-chairs**

**Edward Eisenstein**

University of Maryland, College Park

**James Liao**

University of California, Los Angeles

**Jennifer Reed**

University of Wisconsin, Madison

**Organizer**

**Pablo Rabinowicz**

Biological Systems Science Division  
pablo.rabinowicz@science.doe.gov, 301-903-0379

---

**Web address for this document:**  
[genomicscience.energy.gov/biosystemsdesign/](http://genomicscience.energy.gov/biosystemsdesign/)



# Biosystems Design

## Report from the July 2011 Workshop

Published April 2012

Convened by  
U.S. Department of Energy  
Office of Science  
Office of Biological and Environmental Research



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

# Contents

Executive Summary.....	iii
Introduction.....	1
Sidebar 1: DOE Genomic Science Program .....	2
Sidebar 2: Advances in Genome Engineering .....	3
Biological Design Principles.....	4
Sidebar 3: Moving Toward Design-Based Biosystems Engineering .....	6
Biological Design Strategies, Methodologies, Approaches, and Knowledge Gaps.....	7
Sidebar 4: Molecular Communication Systems for Biodesign .....	12
Sidebar 5: Plant Signaling Systems for Biodesign.....	13
Sidebar 6: DOE Joint Genome Institute.....	14
Sidebar 7: DOE Systems Biology Knowledgebase .....	15
Moving Toward Systems-Level Predictive Biology.....	16
Sidebar 8: Visions for the Future of Biodesign .....	18
Conclusions .....	19
Appendices.....	21
Appendix 1: Workshop Agenda.....	21
Appendix 2: Breakout Sessions and Charge Questions.....	22
Appendix 3: Participants.....	23

## Executive Summary

Growing energy demand cannot be met by nonrenewable fossil fuels. Consequently, alternative energy sources are being intensively pursued. Cellulose, a very stable carbohydrate that makes up plant cell walls, is a potential source of renewable fuels. Transforming cellulose into liquid fuels, however, requires substantial chemical and biological processing, first to extract the sugars that comprise the cellulose and then to convert those sugars into fuels. These requirements pose major hurdles to sustainable biofuels production, but overcoming them may be possible by engineering new plants that facilitate the extraction and conversion of their cell walls into liquid fuels. Similarly, newly designed microbes capable of metabolizing plant cell wall components can simplify biomass processing for fuel conversion. Other microorganisms such as microalgae and cyanobacteria can be redesigned to incorporate new functionalities. For example, these photosynthetic organisms, which capture light and convert it into chemical energy, can be re-engineered to produce biofuels, including biodiesel, directly from sunlight.

Technological advances enabling the design of new biological systems are already moving biofuels closer to becoming viable, alternative renewable energy resources. Further advances are necessary for developing useful bioenergy crops that not only allow facile conversion of biomass into biofuels, but also do not compete with food crops for arable land. In other words, crops must be rationally redesigned to produce high biomass yields on marginal agricultural lands and under changing weather conditions.

To explore the current state of the art in the field of biosystems design, discuss new biodesign technologies and approaches, and identify key scientific challenges and knowledge gaps, the Department of Energy's Office of Biological and Environmental Research organized the Biosystems

Design Workshop in July 2011 in Bethesda, Maryland. The workshop's goal was to bring together scientific leaders in microbiology, plant biology, metabolic engineering, systems biology, bioinformatics, computational modeling, and other relevant disciplines to examine fundamental aspects of biosystems design from molecules to organisms to communities. The focus was on the fundamental biological principles that must be harnessed to make biological design possible and the tools and computer-aided testbeds needed to design, prototype, and functionally validate multiscale natural and hybrid biological systems.

Building on the outcomes of the genomic revolution that altered the course of biological research in the last decades, biosystems design research will identify modular components that can be modified, enhanced, and exchanged among different organisms, enabling the manipulation of biological systems. With the help of computational modeling, *de novo* design of new organisms with novel capabilities for defined purposes will be possible. These new biological systems and modules not only will generate new, useful functions, but also will provide powerful tools to further our understanding of the fundamental principles that rule biology.

The time is ripe for bold new research approaches that harness biology's potential. These approaches include designing new biosystems to address critical needs, such as sustainable production of advanced liquid biofuels, while contributing to sequestration of carbon and reduction of greenhouse gas emissions and improving nutrient- and water-use efficiency of bioenergy crops. The field of biosystems design will enable living organisms to be manipulated and tailored in unprecedented ways, leading to a bioeconomy that can meet our energy needs while minimizing impacts on the environment.





# Introduction

The biological sciences have witnessed a “genomics revolution” over the last two decades. Following completion of the human genome sequence, tremendous technological advances and decreases in the cost of nucleic acid sequencing technology have catalyzed an explosion in the number and accuracy of new genome sequences available for a variety of organisms. Consequently, a vast and ever-growing amount of genomics data is now available in public databases, including the sequences of thousands of small microbial genomes as well as many large plant and animal genomes and mixtures of genomes from microbial communities, or metagenomes. Additional layers of genome-scale information are represented by comprehensive profiles of gene expression (transcriptomics), proteins (proteomics), epigenetic modifications (epigenomics), and metabolites (metabolomics), among many other “omics” datasets. Multi-disciplinary systems biology approaches are necessary for analyzing all these types of genome-scale data to enable a deeper understanding of biological systems. Computational biology plays a central role in such approaches, allowing complex biological networks to be dynamically modeled and displayed. Genomics and systems-level predictive understanding of biological systems are uncovering the design rules that govern their behavior to the extent that rational redesign of organisms is becoming possible. The Department of Energy (DOE) Genomic Science program has played a leading role in advancing the science underlying a predictive understanding of biological systems relevant to energy production and other DOE missions (see Sidebar 1, DOE Genomic Science Program, p. 2).

The field of biosystems design emerged as a product of the synergistic integration of advanced biotechnology, biosciences, and nanotechnology, which has enabled the large-scale modification of biological systems and the construction of completely new biological components and systems. The merger of biology, chemistry, physics, and engineering has the potential to transform fundamental and applied science by extending the capability of natural organisms to shed light on the fundamental principles of biological system organization and evolution. This understanding

can then be used to solve practical problems of significant national and global importance.

A number of recent breakthroughs have piqued keen interest in genome engineering and biological design (see Sidebar 2, *Advances in Genome Engineering*, p. 3). Using computational tools, novel proteins can be designed and synthesized *de novo* to achieve new functions. New regulatory circuits can be constructed and evaluated for high-level function and control. Complete metabolic pathways can be assembled, engineered, and introduced into living cells to produce high-value compounds. Entire bacterial genomes can be replaced with modified synthetic counterparts. Orthogonal molecular processes have been developed to incorporate unnatural amino acids into proteins, conferring new functions by codon replacements through directed evolution. Novel nucleases with customizable specificity for genome engineering have been developed from bacterial transcription activator-like effectors (TALEs). Substantial progress also has been made toward constructing synthetic eukaryotic chromosomes, and the vast amount of comprehensive data available for genes, transcripts, proteins, and metabolites under different conditions for multiple individuals has dramatically advanced network analysis and computational modeling of biological systems.

Growing interest in biological design research stems from its wide-ranging potential. Advances in this nascent field will provide new tools for engineers to construct biological systems and organisms that address unique challenges and bring new approaches to enhance our understanding of complex biological systems. Biosystems design technologies promise innovative solutions for bioenergy, carbon management, and the environment. In the field of energy research, for example, new microorganisms could be developed to harness sunlight and environmental nutrients and recycle atmospheric carbon dioxide, producing fuels and chemicals at a high enough efficiency to outcompete petroleum at current exploration and refining costs. Novel energy crops may be designed to achieve high biomass yields while avoiding negative impacts on the environment. In addition, hybrid biological and nonbiological systems that utilize sunlight

more efficiently than current photovoltaic and biological photosystems may become realities. At the same time, a consortium of multiple designed species may be used to monitor the impact of those activities on the ecosystem, ensuring environmental sustainability.

In July 2011, DOE's Office of Biological and Environmental Research organized the Biosystems Design Workshop, gathering leading researchers in the fields of systems biology, metabolic engineering, microbiology, plant biology, bioinformatics, computational modeling, and other disciplines that touch upon this broad and complex topic. Workshop participants explored the current state of biosystems design research, discussed relevant technology and methodological approaches, and identified key scientific knowledge gaps and potential barriers to progress. The workshop's overarching objective was to examine fundamental aspects of biological design, from molecules to organisms to communities. The focus was on the fundamental biological principles that must

be harnessed to make biological design possible and the tools and computer-aided testbeds needed to design, prototype, and functionally validate multiscale natural and hybrid biological systems.

The workshop began with three plenary talks from renowned scientists in plant biology, microbiology, and bioinformatics. Subsequently, workshop participants gathered into three breakout groups to discuss a series of key questions. (See Appendices 1, p. 21, and 2, p. 22, for workshop agenda, breakout sessions, and charge questions, respectively.) The designated leaders of each breakout group reported their group's discussions to all workshop participants (see Appendix 3, p. 23) during a plenary session on the second day. Participants were encouraged to "think beyond any currently existing box" while attempting to frame their vision within DOE's missions in energy and the environment. Such a task is intrinsically broad and challenging, and the opinions expressed were expectedly rich and diverse.

### Sidebar 1

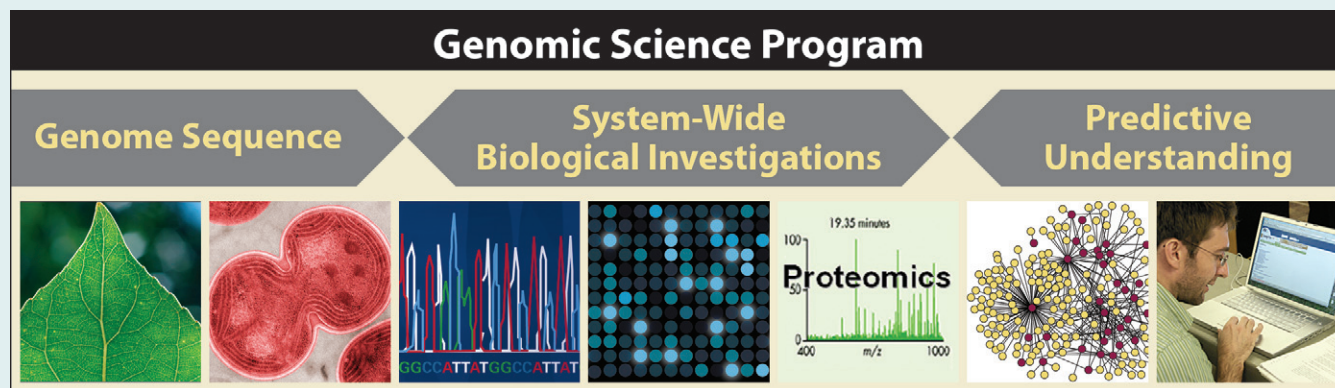
## DOE Genomic Science Program

The Department of Energy (DOE) Genomic Science program supports research aimed at identifying the fundamental principles that drive biological systems relevant to DOE missions in energy, climate, and the environment. These principles guide translation of the genetic code into functional proteins and the metabolic and regulatory networks underlying the systems biology of plants, microbes, and communities. Advancing fundamental knowledge of these systems will enable new solutions to national priorities in sustainable bioenergy production, understanding the fate and transport of environmental contaminants, and developing new approaches to examine the role of biological systems in carbon cycling, biosequestration, and global climate. The major objectives of DOE's Genomic Science program are to:

- Determine the molecular mechanisms, regulatory elements, and integrated networks needed to understand genome-scale functional properties of microbes, plants, and interactive biological communities.
- Develop "omics" experimental capabilities and enabling technologies needed to achieve dynamic, system-level understanding of organism and community function.
- Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance predictive understanding and manipulation of biological systems.

This program is supported by the Office of Biological and Environmental Research within DOE's Office of Science.

[science.energy.gov/ber/research/bssd/genomic-science/genomicscience.energy.gov](http://science.energy.gov/ber/research/bssd/genomic-science/genomicscience.energy.gov)





## Sidebar 2

## Advances in Genome Engineering

Biosystems design is predicated on the ability to introduce large-scale changes into the genomes of a wide range of organisms. Although improvements are needed in designing and constructing large DNA segments that can be effectively inserted into genomes, recent innovations suggest that a variety of methods for genome engineering may help accelerate the development of synthetic biosystems for useful applications.

One groundbreaking approach for building large DNA molecules was demonstrated in the first complete assembly of a synthetic bacterial genome (Gibson et al. 2008). Researchers pieced together the 582,970 base pair (bp) genome of *Mycoplasma genitalium* by assembling smaller DNA segments into larger fragments comprising up to a fourth of the entire genome (~144,000 bp). This impressive achievement serves as a proof-of-principle for more ambitious synthetic genome projects that will be possible as the cost of DNA synthesis continues to decrease.

Several strategies for engineering genomes involve eliminating or reducing expanses of DNA that contain unnecessary regions (e.g., insertion sequence elements and associated effectors) while maintaining genes and regions essential to normal growth and division. Harnessing the natural process of recombination, in which nucleotide sequences are exchanged between similar regions of two DNA molecules, nonessential sequences can be deleted by replacing long segments of intervening sequence with truncated versions of the sequence. This technique was successfully used to reduce the genome of an *Escherichia coli* strain by roughly 15%. The new strain grew normally, had lower mutation rates, and was easier to transform.

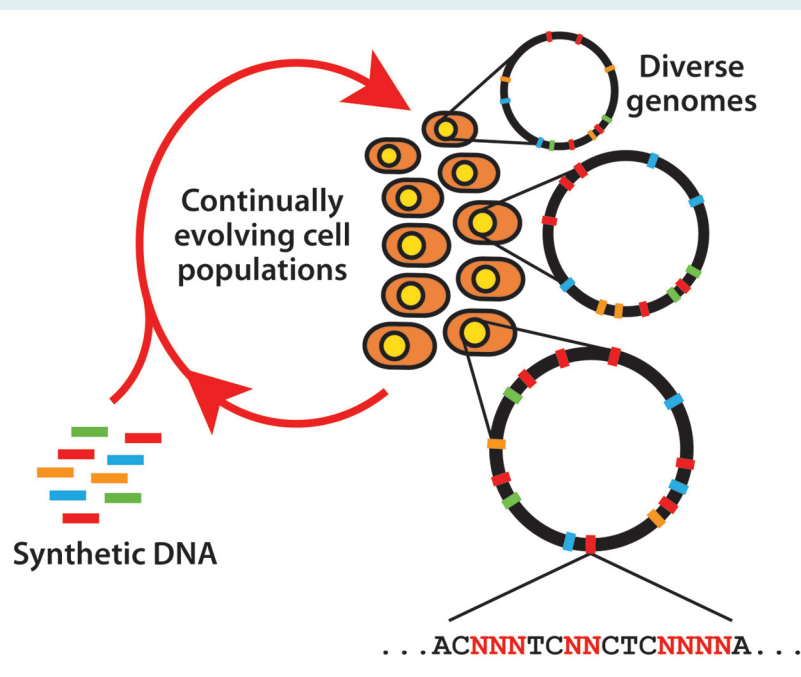
Recombineering—recombination-based genetic engineering—has been automated in a powerful new process called Multiplex Automated Genome Engineering (MAGE). Using single-stranded pieces of synthetic DNA that target many different genomic locations in live cells, MAGE can simultaneously introduce hundreds of specific

genetic changes across a population of cells (see figure). In just days, MAGE can generate billions of cells with various combinations of changes in multiple genes or regulatory regions, an effort that could take months or years using traditional genetic engineering methods. This large-scale approach to genomic reprogramming presents many opportunities for rapidly modifying dozens of genes related to synthesizing particular chemical products.

These select examples of innovation in microbial genome engineering provide a substantial foundation for further development and application to a wider range of organisms. The next few years hold tremendous potential for revealing new tools that use genomic engineering to expedite the design and analysis of synthetic biosystems.

## Reference

Gibson, D. G., et al. 2008. "Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome," *Science* **319**, 1215–20.



**Multiplex Automated Genome Engineering (MAGE).** This process enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations across a large population of cells through the repeated introduction of synthetic DNA. [Adapted by permission from Macmillan Publishers Ltd: From Wang, H. H., et al. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated Evolution," *Nature* **460**, 894–98.]

# Biological Design Principles

A still-unresolved question in biology involves the fundamental laws or principles governing the function of evolved biological systems. Extensive modification of existing networks or outright design of synthetic systems can lead to answers, both by serving as a proof of our understanding of biological systems and by providing novel tools for interrogating basic biological function. Just as early trial and error in airplane design refined the understanding of fundamental fluid dynamics laws and enabled today's fully automated designs, iterative design cycles in biology will reveal the complex interconnections among biological laws, experimental observations of biological systems, and design of new biological behaviors. This knowledge will then lead to a better understanding of biological functions (see Sidebar 3, Moving Toward Design-Based Biosystems Engineering, p. 6).

Every biological function, from the molecular to community level, is ultimately determined by the genetic information encoded in the genome. Genome sequence information—when integrated with other large-scale datasets such as comprehensive profiles of metabolites, proteins, transcripts, epigenetic modifications, and small RNAs—provides the foundation for understanding fundamental biological functions, as well as biological system behavior and response to internal or external cues. However, even *Escherichia coli* and *Saccharomyces cerevisiae*, probably the most-studied organisms at the genome level, are far from being completely understood. In fact, what qualifies as “complete” understanding of an organism is the subject of debate. Accepted standards are needed not only to define what constitutes complete knowledge of a biological system, but also to guide the collection of specific data. Nonetheless, genome-wide comparative studies and other integrative analyses of genome annotations and other “omics” datasets allow researchers to infer the functions of genes, molecules, and pathways and the relationships among them. Although the accuracy of conclusions reached in such studies strongly depends on the annotation quality of each genomic element, the conclusions can be used to identify or predict networks of related biological elements or “parts” that can be considered functional modules, in relative isolation from their cellular context.

## Understanding Characteristics and Functional Attributes of Biological Modules

Modularity is a major principle that enables the design of complex systems. Since biological modules are context dependent, general modularity must be approximated in artificial designs and implementation of new biological systems. Contextual parameters may strongly influence design. Indeed, a challenge to collecting the critical information needed for biological design stems from the contextual information embedded in biological parts. For example, although a particular promoter may be effective for most genes in a specific bacterial isolate, considerable variability may exist among other strains and species. The same promoter thus would not be expected to work similarly in every system. Consequently, there is value in examining many diverse but carefully chosen biological systems.

A deeper understanding of the characteristics of biological modules and their functional attributes will significantly aid the rational redesign of biological systems. In experimental observations, these principles emerge at many scales. At the level of proteins and enzymes, design principles, for example, can focus on the modularity of protein domains that can be recombined to introduce new hybrid functionality. Many natural enzymes are built from protein domain combination through horizontal gene transfer and homologous recombination. These same principles of protein structure and recombination can aid the design of new multifunctional enzymes.

Remarkably, existing knowledge of the physicochemical aspects of amino acids and protein structure can be used to design new proteins with desired biochemical properties starting from an appropriate scaffold. Proteins designed using advanced algorithms and computational tools can generate variants with new functions, and their biochemical behavior can be validated experimentally. Although the thermodynamic principles governing protein structure form the underlying basis of living systems, their level of complexity makes a comprehensive understanding of biological systems impractical from a thermodynamic point of view. For example, designing new enzymatic pathways is constrained by the

enzymes' fundamental properties, such as thermodynamic favorability and kinetic directionality of enzymatic reactions. Such dynamic enzyme function information can be difficult to collect at a whole-cell scale under different conditions, and modeling of many dynamic parameters at different scales rapidly becomes computationally intractable. Dynamic metabolic models have been developed for central carbon metabolism and other isolated pathways, but whole-cell metabolism modeling remains challenging, except at the stoichiometric level, which does not account for kinetics. Optimistically, genomics and systems biology approaches will be advanced analogously to protein design, achieving the same level of predictive understanding of complex biological systems.

## Achieving Robust Biological Design Under Unstable Conditions

Reaching a holistic understanding of an organism's attributes is still a distant goal. For even the simplest prokaryotes on which extensive genomic information is available, a substantial fraction of gene functions is not known. Thus, a “bottom up,” *de novo* approach to designing a free-living organism is a tremendous challenge. Nevertheless, engineering new biological functions in living systems can be achieved using a “top down” approach even without a full understanding of all the system's components. A key issue will be focusing on the relevant aspects of biological systems required for a specific design purpose. This approach may enable the redesign of complex systems such as plants or bacterial communities for which the tools and knowledge available for the simplest organisms or extensively studied model systems do not yet exist.

Regardless of the strategy, an important challenge is integrating biological design approaches across temporal and spatial scales in a way that allows modules and components to be interchanged between different biosystems of variable complexity. Another difficulty is that biological design must be achieved under significant uncertainty, making robustness and insensitivity to environmental disturbance major considerations. Here, engineering design and quantitative modeling may offer valuable insight. Systems biology analyses of biological networks have identified many ways in which biological systems remain robust under unstable conditions. However, the precision of biological models is still far from that of physical models, so qualitative, conceptual advances are more common than precise quantitative understanding. The difficulty arises when qualitative behavioral changes, such as the dynamic regime in nonlinear systems, depend on quantitative parameters. Without precise parameters and with the centrality of noise in evolved biological systems,

biological design needs to avoid a dependency on a narrow range of parameters for success. Consequently, iterative design cycles will connect theoretical understanding of biological robustness with experimental data to develop mathematical models and design principles that can guide robust synthetic biological design.

## Improving Data Collection, Integration, and Analysis Technologies

Modeling and designing new biological entities require large amounts of information and an improved capacity to distinguish biological variability from noise. Existing datasets are invaluable to this end, but certain data types are more easily accessible than others, and data quality varies widely. For genes on which little more than structural information is available, metabolic flux measurements, genotype-phenotype relationships, and high-throughput phenotype quantifications are examples of important knowledge and capability gaps. These gaps become wider as our ability to integrate this disparate data from a broad variety of organisms and communities grows more slowly than our ability to collect it. Improved cross-disciplinary technologies are greatly needed to gather, integrate, and analyze such data at the community, organismal, cellular, and subcellular levels.

Clearly, the range of redesign possibilities is correlated to the degree of system complexity. Bidesign of microbial communities and plants will achieve its potential when a comprehensive understanding is reached regarding important concepts such as compartmentalization, intercellular transport, cellular specialization, and cell-cell communication. Other common features observed in eukaryotes—such as centromeres, telomeres, introns, absence of operons, and other genome architecture characteristics—pose additional challenges for engineering more complex organisms.

By capitalizing on an increased understanding of fundamental principles, as well as a greater ability to derive useful approaches from data-driven knowledge or statistical information, significant advances in biological design have been possible. With increased information across scales and systems, more ambitious designs can be undertaken for increasingly complicated systems and applications. Furthermore, newly designed biological parts and modules with specific functions can be shared within the scientific community for use in similar or different model systems, catalyzing further innovation. This exchange of biological parts among laboratories and systems will require standards and repositories to facilitate usage and accessibility.

## Sidebar 3

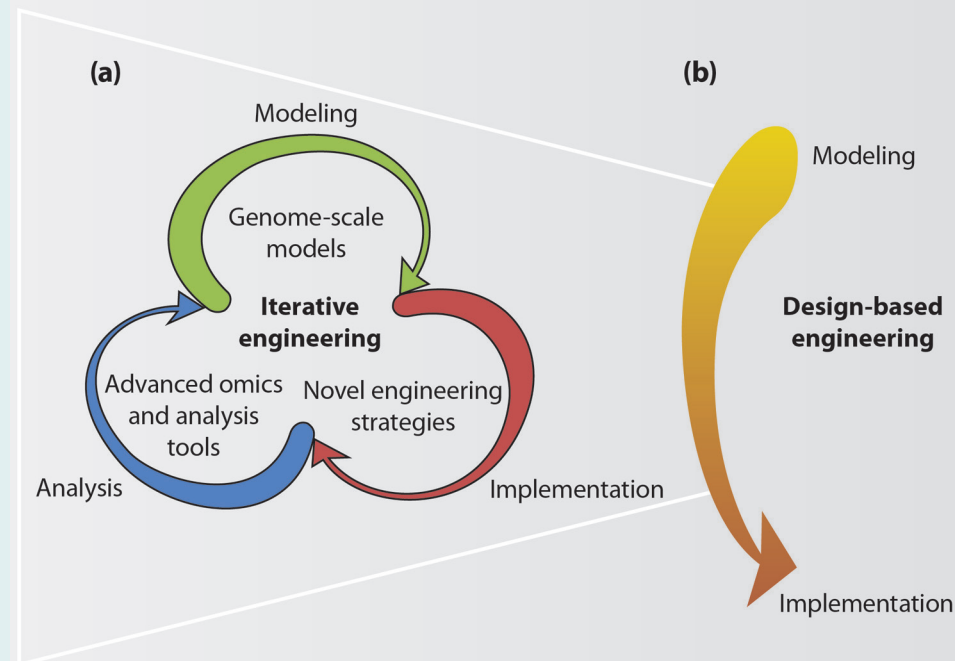
## Moving Toward Design-Based Biosystems Engineering

For traditional engineering fields, such as mechanical, electrical, or chemical engineering, the design-based construction of a device (e.g., electric circuit or engine) directly from a plan is possible because this engineering is based on a complete mathematical understanding of the physics controlling the device components. Engineering biological systems is more challenging because our knowledge of the principles that govern these living systems is largely incomplete.

To design a microbial system that produces a desired product, for example, researchers typically modify existing organisms based on models that are limited in their predictive capabilities. As a result, the engineering of microbial systems follows an iterative cycling of modeling, implementation, and analysis (see figure). Researchers design a strategy (e.g., introducing new enzymes or knocking out competing pathways) for engineering a microbial system based on genome-scale knowledge and models. Available models of microbial systems do not encompass all activities of a cell, so insights from the “omics” analyses of the engineered microbial sys-

tem are used to revise the genome-scale models and improve predictability.

Transitioning biosystems engineering from an iterative process to a linear, design-based approach requires a synergy between biodesign and systems biology. Biosystems design can help reveal how the addition or elimination of different pathways (or other biological components) can influence the dynamic behavior of the complete system. By providing a comprehensive view of how all biological components work together, systems biology enables the discovery of organizing principles needed to assemble different components into a functional system. Building synthetic biosystems based on these principles provides an important mechanism for testing and validating biological design strategies. Other key requirements for making biosystems engineering faster and more predictable include increasing the number of well-characterized biological modules available to the research community, enhancing interoperability of those modules, and standardizing the tools and strategies for assembling them.



## References

- Tyo, K., K. Kocharin, and J. Nielsen. 2010. "Toward Design-Based Engineering of Industrial Microbes," *Current Opinion in Microbiology* **13**, 255–62.
- Smolke, C. D., and P. A. Silver. 2011. "Informing Biological Design by Integration of Systems and Synthetic Biology," *Cell* **144**, 855–59.

## Engineering Microbial Systems.

As the tools for engineering microbial systems improve and the predictive capabilities of genome-scale models become more reliable, engineering microbes will transition from (a) an iterative cycle to (b) a linear, design-based process. [Adapted from Tyo, Kocharin, and Nielsen 2010 © with permission from Elsevier.]



# Biological Design Strategies, Methodologies, Approaches, and Knowledge Gaps

Despite dramatic progress over the last two decades in our understanding of biological systems and their components, several formidable obstacles remain in applying biosystems design to increase our understanding of biological systems and in fully capitalizing on its potential to address energy, climate, and environmental challenges. Bridging key gaps in knowledge-, technical-, and computational-based approaches would enable broader biosystems design applications.

Scientists face a number of challenges related to the principles of genome organization, construction and development of improved phenotypes, and ability to harness evolutionary forces for implementing biosystems design in complex organisms. Multiple control mechanisms for gene stability, expression, and evolution obscure information relevant for design. Particularly challenging is the hierarchical nature of biological modules such as genes and the proteins they encode because they can function under certain or unique conditions but not necessarily in all environments. The prospects for redesigning organisms that possess specific, useful, and unambiguous attributes will increase with improved knowledge of the factors (genes, regulatory circuits, and signals) that give rise to complex phenotypes such as metal sequestration by bacteria or drought resistance in plants. Understanding the nature of evolutionary change and how its impact on an organism will influence the environment will be essential for realizing the significant potential of biological design.

## Focusing on Key Organisms

One approach for facilitating broader biological design implementation would be to focus research on key organisms to enhance understanding of photosynthesis, nitrogen fixation, and other biological processes important to DOE mission areas.

There are a variety of biological subsystems, organisms, and communities that would be useful for establishing general biosystems design principles. The choice of a reference

system depends on the design goals. These reference systems, whether new or existing, also could be used as a basis or foundation on which to further improve their function in addressing energy and environmental challenges. Good progress has been made with well-characterized reference systems in microbes, especially for *Escherichia coli* and *Bacillus subtilis*, and some plants, including *Arabidopsis thaliana*. More reference systems need to be characterized, however, to fully understand the functional context of biological components. Although our understanding has benefited tremendously from studies on a few model reference systems, there is great value in characterizing a wider range of systems identified based on a particular application. Some organisms may be more relevant for addressing a certain biological question (e.g., cellulose degradation by certain types of fungi or bacteria), but emphasis also could be placed on designing communities and subsystems, including, but not limited to, the construction of discrete pathways, gene circuits, and feedback loops. Subsystem examples include DNA or organelle replication, transcription (and its regulation), recombination, DNA repair, and metabolic pathways. An increase in the number of new and interesting biological components that can be used as biological parts in biosystems design would be an important advantage of having more reference systems and subsystems.

Biological design in plants poses some particularly challenging problems relative to similar work in microorganisms. For instance, knowledge is lacking on many of the mechanisms plants have for transmitting information across tissues or organs. Additionally, the vast diversity of specialized compounds that plants synthesize, many in response to stimuli or for cell-cell communication, is not fully understood. Although some common ideas have emerged on the mechanisms governing cellular interaction and communication, many aspects of this interaction and how homeostasis is achieved are simply unknown and, at this point, cannot be programmed into a synthetic design project.



## Accelerating Technical Developments

Building on a greater knowledgebase that provides the foundation for efficient biodesign strategies and applications, a number of technical developments also would greatly accelerate synthetic design in biology. Such technologies include homologous recombination techniques that increase the efficiency of introducing large DNA molecules for stable integration into host organisms. Additionally, the efficient transformation of cells across multiple species is required to better understand, evaluate, and engineer the contextual issues related to the function of biological modules. Such a technical advance will be especially important for large nucleic acid segments containing numerous genes and regulatory sequences. As the cost of DNA synthesis decreases, possible bottlenecks will arise from the availability of biosensors, assays, and screens to assess the introduction of synthetic genetic material into simple organisms. Groundwork in simple bacteria provides some clues, but additional, specialized work is needed for more complex species, especially plants. Reproducible and efficient transformation systems have not been developed for many plant systems, including many of interest for biofuel development. In these cases, new methods and approaches are needed. Similarly, understanding and manipulating epigenetics will be key for achieving large-scale stable plant transformation because epigenetic modifications can affect engineered gene expression, and these modifications, in turn, can be affected by environmental factors.

## Developing New Genetic and Molecular Tools

Along with more comprehensive information and advanced technologies, robust tools for manipulating biological properties will be essential assets in successful biodesign. Advancing the development of genetic and molecular tools is crucial as capabilities improve for constructing new, large, and inexpensive segments of genetic material and introducing it stably, efficiently, and reproducibly into a wide spectrum of cell types and organisms. These tools are needed for elevating more organisms (especially those of relevance to energy and the environment) to “model” organism status. Important species would be those involved in elemental carbon, nitrogen, sulfur, and phosphorus cycles.

One consequence of incorporating new genes into cells is that the enzymes they encode can redirect synthesis of a wide range of cellular components, particularly small molecule metabolites. Metabolomics, the study of small molecule metabolite complements in an organism or tissue, has benefited greatly from mass spectrometry advances in mass

accuracy measurements and molecular identification and specialized sample preparation approaches. However, even with these developments, only the most abundant or most stable molecules present in these systems can be detected in many cases. Instrumentation advances and new approaches that preserve unstable species and increase our ability to detect smaller concentrations with even less sample material will enable a more accurate implementation of biological design principles for programmed metabolic control and engineering. Applying new mass spectrometry technologies and capitalizing on cell separation by flow cytometry could potentially address these shortcomings. Moreover, the ability to probe small molecule metabolites, as well as proteins and their complexes, inside the cell where they function would be particularly useful in successful design applications.

## Controlling Gene Expression in Plants

Plants pose some specific challenges to large-scale biosystems design. One critical technical advance would be the development of a broad spectrum of components that can control gene expression. Our present level of understanding suggests that many plant promoter systems show different tissue specificity in different species. Consequently, generating interchangeable parts for design experiments in plants that require tightly controlled, tissue-specific gene expression creates additional complications relative to similar approaches for microorganisms.

A well-developed and promising area for implementing synthetic approaches for biosystems design involves building on decades of progress in protein engineering to establish a more robust and predictable correlation between protein structure and function, particularly for less well resolved systems such as membrane proteins, transporters, and photosynthetic components. A deeper understanding of the structure and stability principles for these systems, as well as how they compare with more extensive studies in soluble enzymes and proteins, will significantly impact the design and engineering of specific properties into complex biological systems.

## Developing Uncertainty Prediction Tools

Another key issue in facilitating the engineering of biological systems is the ability to integrate new data from technological advances with new knowledge gleaned from comparisons of a variety of systems using genomics and other biological databases and biodesign computational tools. Envisioned are computational advances that will capitalize on greater interaction between biological experimentalists and computer scientists to generate powerful tools that quickly and accurately

visualize data and information on biological systems. These new computational analyses would be able to predict uncertainty in experimental outcomes, a capability important for validating calculated results against experimental biological data. Community-wide, objective assessments of algorithm performance have had some success and have driven progress in protein structure prediction and interactions. They could be extended to systems analysis of biological organisms. A move toward standardized data deposition for high-throughput biological measurements such as gene expression, protein, and metabolite levels could facilitate these developments. Indeed, increasingly sophisticated modeling approaches could enable researchers to account for structural and biological changes in protein properties that lead to system-wide cellular changes in an organism. This level of data and knowledge integration, accompanied with user-friendly tools to evaluate design and emergent properties, will truly transform the science of biosystems design, allowing unprecedented use of engineered biological systems.

## Evolutionary and Ecological Processes and Dynamics for Engineering Biological Systems

One attribute of biological organisms not shared by engineered systems is their ability to self-replicate through an error-prone process, a mechanism that leads to mutational changes and is a cornerstone of evolution. Thus, a major challenge in biosystems design is to develop appropriate practical understanding of the replicative nature of biological organisms and their ability to evolve. This understanding is needed to preserve the intent of the original rational design and incorporate the biologically unique feature of evolution into the design process. The way evolutionary processes tend to modify an engineered system is an important consideration for attaining robustness in newly designed biological systems and controlling their potential implications and outcomes. This issue is relevant not only to the sustainability of redesigned systems, but also to environmental and societal aspects, which were addressed in a previous DOE–Alfred P. Sloan Foundation workshop report titled *Societal Issues Arising from Synthetic Biology: What Lies Ahead* (see [www.synbioproject.org/events/archive/what\\_lies\\_ahead/](http://www.synbioproject.org/events/archive/what_lies_ahead/)).

In laboratory settings, specific conditions typically are used to select and fix particular genotypes of organisms responsible for improved fitness or a reproductive advantage under unique environmental parameters. A more comprehensive understanding of how evolution influences biological components to create new phenotypes is essential for designed

biosystems to meet specific, predetermined attributes and avoid undesired functional drift. Successful large-scale bio-design therefore will rely on a better understanding of how fitness is achieved in natural environments and on a greater appreciation of biodiversity and its role in evolution. Such knowledge will enable more robustly designed biosystems that maintain key functional attributes, even as their environmental niche changes. The ability to link experimentally all of the traits responsible for fitness in a particular environment, with the aid of new computational approaches, may provide useful clues in highlighting environmental conditions that favor selection of one genotype relative to another.

Better understanding also is needed of the impact of evolution at many levels within a functional description of a designed biosystem. For instance, simple evolutionary changes seen in individual genes or simple replicons in many model organisms can be explained. However, enhanced understanding is needed of the impact these evolutionary changes have on larger networks and circuits and their development, as well as on the effect of species loss and invasion in natural communities.

Despite the challenges of incorporating a practical working theory of evolutionary processes into biosystems design, this fundamental biological property already is providing a powerful tool for engineering biological systems. Directed evolution approaches have been widely employed to improve rational and computational designs of proteins and gene circuits. They also have been applied more comprehensively to engineer complex pathways and even bacterial strains. This laboratory-based approach for accelerating the evolutionary process requires judicious application of appropriate environmental conditions to select desired phenotypes. Doing so has frequently demonstrated that a particular function can be incrementally improved over a reasonable time frame.

A possible alternative to incorporating evolutionary theory into biosystems design might consist of bypassing evolution entirely, developing biosystems and organisms that are disposable rather than replicative. This approach could involve components that would be essential under one condition but would prevent replication under another. Consequently, once the designed system's intended purpose is accomplished, the system no longer would replicate and evolve but die off instead and not be constrained by our limited ability to incorporate evolutionary theory into design.

In the pursuit of designing complex biosystems, our understanding of evolution will likely increase as the trajectory of genotypic changes is evaluated for biological components

that change under environmental selection. Processes can then be developed to prevent designed systems from mutational corruption.

## Compartmentalization and Cellular Communication at Multiple Levels in Designing Biological Systems

Natural biological systems utilize compartmentalization to spatially separate biological processes. Such compartmentalization may be critical for biological function. For example, *Anabaena*, a photosynthetic bacterium, forms specialized cells called heterocysts that fix nitrogen gas by converting it into ammonia using the enzyme nitrogenase. Nitrogenase is sensitive to oxygen, a product of photosynthesis. Spatially separating these two biological processes (photosynthesis and nitrogen fixation) allows both to occur in the same cell. Compartmentalization can be useful at either a subcellular or cellular level.

Using compartmentalization in biological design will be beneficial for a variety of reasons. First, toxic or oxygen-sensitive chemistries can be segregated through the use of compartments (e.g., nitrogenase activity). Second, substrates could be isolated and concentrated, driving higher reaction rates. For example, carboxysomes are bacterial microcompartments used to concentrate carbon dioxide, thereby improving carbon fixation efficiency. Third, organisms and communities could be more easily controlled. Lastly, compartmentalization would facilitate the specialized functions that community members need to survive in extreme environments, such as those with low pH or high temperatures.

The ability to analyze and act on signals is another key concept that engineered and biological systems have in common. Even simple bacterial organisms exhibit the ability to detect and analyze signals from one another within a community, for example, modulating growth and proliferation. In essence, they communicate with one another through the action of small molecule effectors in an exquisitely regulated process known as quorum sensing (see Sidebar 4, Molecular Communication Systems for Biodesign, p. 12).

In its simplest form, quorum sensing in Gram-negative bacteria relies on small molecule acyl-homoserine lactone autoinducers. These diffusible chemicals bind to intracellular receptors within cells to promote transcription activation of operons encoding cognate synthases that produce even more autoinducers. This amplification cycle—used predominantly for intraspecies communication because of the receptor proteins' high specificity level for their cognate autoinducer—

results in a cell density increase and, in effect, enables bacteria to display population-wide behaviors analogous to multicellular organisms. Small molecule signals can be quite diverse among bacteria. In addition to acyl-homoserine lactones, quinolines, fatty acids, and cyclic dipeptides also have been shown to enable communication within species. Alternatively, the reasonable conservation demonstrated among sensing mechanisms of prokaryotes and eukaryotes suggests that these components could be incorporated into facile designs that target interkingdom communication. The similarity in signaling mechanisms observed among many systems is useful, but more information is needed on how signals are regulated to maintain stable output. Not surprisingly, simple communication systems have been studied with interest within the biodesign research community to develop model systems for evaluating signal processing and control circuits in biology. However, adapting or developing more elaborate systems for biological design would be one of the challenges for interspecies communication, especially via negative feedback.

In contrast to the relatively limited repertoire of chemical entities known to be used in bacterial signaling, the range of small molecules found for communication in plants is extremely complex. Extensive studies have identified many thousands of secondary metabolites that plants use to convey information about stress, disease, and pathogens between not only plants, but also plants and pollinators as well as predators. Despite major efforts to identify the natural plant products that function as signals, especially among the most abundantly synthesized products, the many mechanisms whereby they act are less well resolved than for bacteria, especially for metabolites made at much lower levels. Consequently, intense investigation is being focused on plant signaling systems with the goal of applying relevant components to new designs (see Sidebar 5, Plant Signaling Systems for Biodesign, p. 13).

A better understanding of intra- and interspecies communication in microorganisms and plants would give the biosystems design research community valuable tools to adapt and control complex interactions. For instance, modified viruses or bacteriophages could be used to introduce complementary features for new behaviors, better discriminate between self and nonself, decouple biosynthesis from growth, and more precisely regulate stimuli-dependent actions through deregulation by inhibition.



## Integrating Systems Biology Data and Approaches to Inform Biological Design Principles

Redesign of biological systems will strongly synergize with ongoing programs, such as the DOE Joint Genome Institute, that pursue genome sequencing and functional characterization of new organisms (see Sidebar 6, DOE Joint Genome Institute, p. 14). Not only will more extensive genome sequencing identify new biological components, but variants of genes and regulatory elements will become available, increasing the breadth and depth of the repertoire of biological parts and modules that can be modified and utilized for redesigning biological systems.

Additional metabolic data also are needed for enabling biosynthetic system design. These data include new enzyme substrates, products, and cofactors, as well as quantitative information about the stoichiometry of individual reactions, particularly those carried out by transmembrane transporters and electron transport chain components. Currently, prediction of transporter substrates remains a challenge. Transporters can be identified and classified by type (e.g., MFS or ABC transporters) based on sequence, but substrate specificity and energetic requirements for moving these substrates are difficult to predict.

As interest in the design and engineering of biological systems grows, the significant gap between our current understanding of biological systems, their components, and how they fit and function together, along with our ability to achieve highly precise design requirements, is increasingly being recognized. Systems biology investigations—with their focus on high-throughput, quantitative measurements of genes, proteins, and metabolites at the cell, organism, and community level—linked with computational approaches to integrate comprehensive datasets will provide formalisms and mathematical models to account for biological processes and inform biological design principles. However, two key issues limit biological design from systems biology approaches. The first involves technological challenges related to quantitative measurement and the degree to which exact data are needed to enable biological design. Second, more seamless connections between high-throughput and bioinformatics tools need to be developed to link system information and provide the information accuracy necessary for redesigning systems to validate our understanding and for moving confidently into novel biosystems design.

Although these limitations in our understanding of biological systems have not been a fundamental deterrent to biosystems

design, they may be a more significant issue as goals become more ambitious and designs more complex. For instance, a number of designed biosystems have been constructed with an “either/or” sensor to assess the mere presence of a specific compound in the environment. It is much more challenging to not only detect, but also modify and remove a compound by converting it into something more useful or less hazardous. This additional complexity requires a more robust understanding of the linkages between biological components, as well as how these components give rise to system behavior.

## Advancing Computational Biology Tools to Facilitate the Design of Novel Molecules, Networks, Cells, and Communities

Computational approaches that can be used to predict protein structure and function, metabolic fluxes, and biological control mechanisms (e.g., transcriptional or post-translational regulation) will continue to play a vital role in biological design. With these capabilities, computational biology enables the *in silico* design of biological components and systems (including unicellular and multicellular organisms) that have desirable properties. Nevertheless, additional advances are needed for computational biology to enable improved *in silico* designs and reduce the experimental effort needed to achieve stated design criteria.

The DOE Systems Biology Knowledgebase presents a significant opportunity to link some of the proposed Knowledgebase activities to better tools for biosystems design (see Sidebar 7, DOE Systems Biology Knowledgebase, p. 15). These computational approaches have applications ranging from the design of macromolecules, pathways, large-scale networks, microbes, plants, and complex communities. Although some of these computational biodesign needs may fall outside the Knowledgebase initiative’s scope, all computational approaches should enable new hypothesis generation and refinement based on observed experimental behaviors. At the macromolecule level, improved computational tools are needed for increasing the relative ease and accuracy of *de novo* enzyme design. Such tools would enable the design of enzymes with increased and new activities (e.g., improved cellulases). At the pathway level, use of chemi-informatics approaches would enable novel metabolic pathway designs (and novel metabolite synthesis and degradation). By leveraging the Knowledgebase and other informatics capabilities, comparative genomics also would allow a search of biological sequence space for the best gene candidates to build such pathways. At the network

level, computational approaches are needed to elucidate gene networks for plants, microbes, and communities. Also needed are improved constraint-based models of metabolic and regulatory networks with greater accuracy and scope (e.g., by including thermodynamic and kinetic constraints). At the cellular, plant, and community levels, the ability to develop and utilize compartment-specific models (where a compartment could be an organelle, cell type, or tissue)

and model interactions between compartments is needed. Finally, modular, multiscale computational models are needed for integrating different computational tools spanning different biodesign scales (from macromolecules to communities to environmental scales). These models could be used to predict how a change in enzyme activity, for example, would affect cellular, organismal, and community behavior.

#### Sidebar 4

### Molecular Communication Systems for Biodesign

Communication among organisms is crucial for developing synchronized behaviors that benefit the community (see figure). Molecular communication systems, which have been found in even the simplest bacteria, are used to sense dense populations and trigger responses at the genetic level to adapt to new environmental conditions.

The small molecule chemical signals (called autoinducers) used in bacterial communication are synthesized in cells and secreted into the surrounding environment where their concentration increases with increasing cell density. They can be internalized by neighboring organisms and recognized by intracellular receptors to regulate the transcription of genes that stimulate more autoinducer production and initiate a biological response to the changing environment. One of the earliest and most-studied examples of bacterial communication is quorum sensing in the Gram-negative marine bacterium *Vibrio fischeri*. In high population density environments where nutrient concentrations are typically greatest, these bacteria use acyl-homoserine lactone autoinducers to activate genes that control bioluminescence. Interestingly, bacterial communication systems are widespread, though autoinducers differ between species. Some bacteria, such as *Pseudomonas aeruginosa* use multiple quorum sensing systems to regulate many complex cellular functions related to virulence.



**Bacterial Communication.** Bacteria, such as these *Shewanella putrefaciens* cells growing on iron oxide particles, use chemical signals to coordinate biofilm formation and other community-level behaviors. [Credit: DOE Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory]

Molecular communication systems have been an attractive target for bio-systems design given their important role in bacterial growth and physiology and the relative ease of building and manipulating genetic circuits consisting of autoinducer synthases and receptors. A number of interesting bacterial systems have been designed to have controlled and predictable behavior. One example is the construction of biological “logic” circuits, in which one organism synthesizes an autoinducer and another organism receives the signal and executes a functional response. These systems have been adapted for bacterial population density control, controlled pattern formation, and synthetic predator-prey systems. Importantly, the communication principles developed in these bacterial designs have been extended to engineer communication systems in other higher organisms and even between different organisms such as yeast and plants.

Early results using bacterial communication systems have helped develop computational approaches that account for simple behaviors and predict outcomes of more complex designs. However, continued advances in genome engineering and computational modeling of increasingly complex behaviors will be needed to reach the long-term goal of *de novo* design of synthetic biosystems for diverse biotech applications.



## Sidebar 5

## Plant Signaling Systems for Biodesign

Redesigning gene products, pathways, and systems in plants represents a significant opportunity for developing novel, important applications. An interesting multidisciplinary approach combining computational protein design, synthetic biology, and plant biology has recently led to the generation of plants that could detect trace levels of virtually any small molecule.

The design of these synthetic phytodetectors is based on a two-component signaling system derived from bacterial and plant proteins. In its simplest form, this system consists of a membrane-spanning receptor and a response regulator. Upon effector binding, the receptor is activated, resulting in dimerization and autophosphorylation. After accepting the phosphate from the receptor, the response regulator activates a latent DNA-binding domain to regulate transcription. Once the modular system components were initially constructed and evaluated in bacteria, the signaling system was successfully transferred to plants for *in situ* functional analyses.

Three key attributes were required for constructing a synthetic plant signaling system: (1) an appropriate receptor to initiate the signal, (2) a mechanism to transmit an external signal into the cell, and (3) a suitable output result to indicate the presence of the effector of interest. The receptor component of the signaling system was a bacterial periplasmic binding protein that was computationally redesigned to recognize a new chemical instead of its normal effector, ribose. One part of the transmission component is a hybrid transmembrane receptor consisting of extracellular and intracellular domains from two different bacterial systems. This hybrid receptor also was modified to enable its expression at the plant plasma

**Chemical-Detecting Plants.**

A synthetic signal transduction pathway activates degreening genes, causing a plant to change color over 48 hours after detecting minute amounts of a chemical.

[Credit: Colorado State University]



membrane. The other part of the transmission component, the response regulator, was a fusion between bacterial and viral proteins designed to accept phosphorylation and activate transcription in eukaryotic systems. Once phosphorylated, the response regulator is translocated to the nucleus where it initiates the transcription of specific genes. By linking this synthetic signal recognition system to the activation of genes that rapidly produce an observable phenotype, plants can be used to detect minute quantities of particular chemicals relevant to many different application areas. For example, a hybrid signaling system has been used in plants to turn on degreening genes, which trigger a pale color in leaves that is observable within hours after exposure to a selected effector.

This research illustrates the need for multidisciplinary approaches that can combine complementary components from plants and bacteria to construct new designs, test functional attributes of new hybrid systems in simpler bacterial organisms, and then transfer these constructs to plants for *in vivo* analysis. Although a quantitative model has not yet been developed to describe the system's performance, these achievements mark a breakthrough in the development of complex, synthetic signaling systems in higher organisms.

## Sidebar 6

## DOE Joint Genome Institute

The U.S. Department of Energy (DOE) Joint Genome Institute (JGI) is the only federally funded high-throughput genome sequencing and analysis facility dedicated to genomes of nonmedical microbes, microbial communities, plants, fungi, and other targets for DOE missions in energy, climate, and the environment. Located in Walnut Creek, California, JGI annually provides more than 1,800 users worldwide with access to massive-scale DNA sequencing capabilities that underpin biosystems design and modern systems biology research. High-quality sequencing data and annotations from JGI are needed to continue to improve comparative genomics approaches, transcriptional regulatory network inferences, and predictions of how gene expression is controlled at the transcriptional level.

As a leader in analyzing the genomes of organisms with novel capabilities discovered through bioprospecting in diverse environments, JGI is expanding the range of biological components and metabolic functions that can be used to design new DOE-relevant biological systems. By enabling the publication of project results in more than 150 peer-reviewed journal articles each year and depositing data in public databases, JGI ensures that the larger scientific community can access and benefit from its contributions to genome science. Supported by the DOE Office of Biological and Environmental Research, JGI is managed by Lawrence Berkeley National Laboratory.



**DOE Joint Genome Institute (JGI).** With advanced capabilities for sequencing and analyzing genomes, JGI supports projects around the world investigating the genomes of microbes, microbial communities, and plants relevant to U.S. Department of Energy missions. [Credit: DOE JGI]

## Sidebar 7

kbase.us

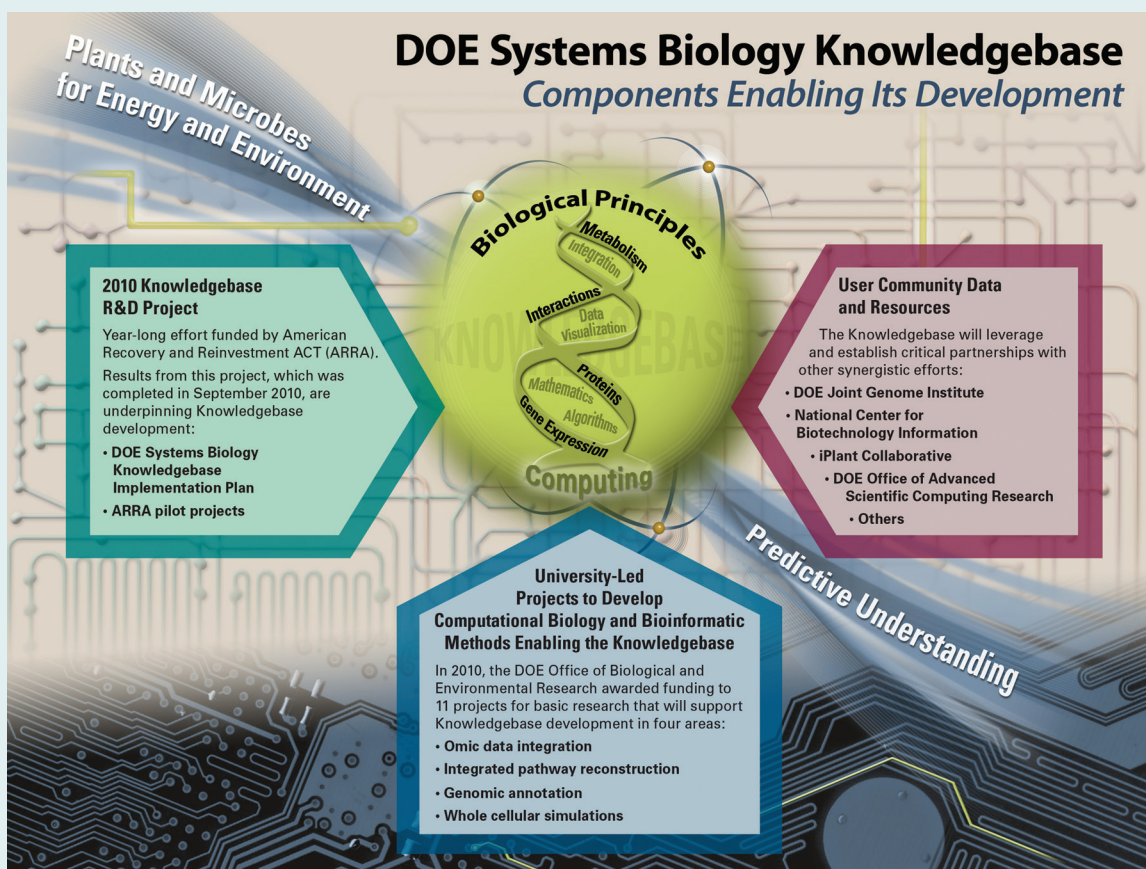
## DOE Systems Biology Knowledgebase

Driven by the ever-increasing wealth of data resulting from new generations of genomics-based technologies, research on biological systems is demanding a computational environment for comparing and integrating large, heterogeneous datasets and using this information to develop predictive models. To provide the research community with such a resource, the Genomic Science program is developing the DOE Systems Biology Knowledgebase (Kbase), led by Lawrence Berkeley National Laboratory, in partnership with Argonne, Brookhaven, and Oak Ridge national laboratories. A knowledgebase is a cyberinfrastructure consisting of a collection of data, organizational methods, standards, analysis tools, and interfaces representing a dynamic body of knowledge. The fully functional Kbase will support free and open access to data, analysis tools, resources for modeling and simulation, and information for the research community.

Kbase differs from current informatics efforts by bringing together research products from many different projects and laboratories to create a comprehensive computational environment focused on DOE scientific objectives in microbial, plant, and metacommunity (complex communities of

organisms) research. By democratizing access to data and computational resources, Kbase will enable any laboratory or project, regardless of size, to participate in a transformative community-wide effort for advancing systems biology and accelerating the pace toward predictive biology.

Some Kbase computational development efforts relevant to biosystems design should be leveraged, including the integration of heterogeneous data (e.g., genome sequencing, gene expression, protein expression, and metabolite concentration data) and the development of improved automated tools for annotation using multiple lines of evidence. Such approaches could be used to generate hypotheses about the mechanisms underlying complex traits (e.g., tolerance to biofuels or pH) or to identify candidate enzymes with desired metabolic activities. Advancing computational tools for biosystems design requires open source capabilities (so that tools can be maintained or modified by the community as needed), accessibility to both the biological and computational communities in a timely and verifiable manner, and interoperability with other computational tools and databases. These accessibility and usage criteria are consistent with Kbase goals and implementation requirements.





## Moving Toward Systems-Level Predictive Biology

The anticipated advances in biological design will lead to a number of significant improvements that will substantially impact bioenergy, environmental, and climate change research. Short-term costs for synthesizing large DNA fragments will continue to decrease, enabling the construction of synthetic DNA fragments with desired properties (e.g., optimized for gene expression and translation efficiency) that are more accessible to the research and industrial communities. Currently, DNA synthesis can be readily carried out for DNA molecules the size of a bacterial gene. Construction and assembly of large DNA fragments (several kilobases to megabases) likely will be possible soon. The ability to easily synthesize large DNA pieces and introduce them into living organisms will enable more rapid construction and testing of engineered microbes for improved bioenergy production and lead to better characterization and prediction of sequence-function relationships (e.g., mRNA stability and expression).

Progress in genome engineering to reduce codon redundancy and enable the routine introduction of altered amino acids using chemically or biochemically charged tRNAs will facilitate a wide range of biodesign investigations. These technologies could advance the development of symbiotic relationships with new functional attributes under specific environmental conditions.

Workshop participants could foresee progress toward a more complete list of biological components from DOE-relevant organisms that will be evaluated and characterized. Some questions that still need to be addressed include:

- From which reference organisms should parts or modules be derived?
- What constitutes a biological module (e.g., gene, promoter, gene fragment, operon, and metabolic pathway)?
- How can standard modules be accessed and shared by the research community?
- Where could information on the availability of biological modules be gathered?

Biosystems design has already been successful in improving biofuel production in plants and microbes. Additional research efforts will lead to improved yields, productivity, and metabolite titers, making biofuels more economical. Microbes capable of converting sugars into advanced biofuels with commercial potential are expected to be developed in coming years. To access these sugars from lignocellulosic biomass, new stable and economically viable enzymes will be developed for efficient biomass processing. Biofuel-producing microbial strains might require a complete redesign of certain metabolic pathways (e.g., glycolysis) for improved biofuel production. Additional improvements might involve water-use efficiency by algae and cyanobacteria for photosynthesis-driven biofuel production. Development of high energy-density plants that make fatty acids and hydrocarbons in their vegetative tissues also may become a reality in the near future.

Advances in computational modeling will facilitate these goals. For example, automated reconstruction of genome-scale models will become possible for Archaea (single-celled microbes morphologically similar to bacteria but actually constituting a distinct domain of life) and eukaryotes. These reconstructions—supported in part by efforts like the DOE Systems Biology Knowledgebase—will have equal or better quality than those existing for reference organisms such as *Escherichia coli*. Such reconstructions will enable *in silico* evaluation of potential genetic and environmental changes and their impact on bioenergy and the environment. *De novo* enzyme designs for improved activity, altered specificity, and stability also will be possible, impacting a number of biosystems design research areas.

Newly developed organisms will be able to better manage or overcome a variety of stresses, including those derived from climate change (e.g., high temperature and drought), high cellular-concentration fermentation systems, and toxic metabolic products (e.g., alcohols and other biofuels or their precursors).

A variety of genetic sensors also will be needed, combining attributes from signaling receptors that activate gene expression under defined conditions with regulons that would

produce outputs in response to stimuli. These systems, such as sentinel plants, would have broad applications in environmental research and lay the groundwork for developing more complex circuits.

Another near-term advance involves ongoing efforts to establish more reliable genotype-phenotype linkages for better annotation of biological components, combining details of metabolism, enzymology, and even chemistry into a gene's phenotypic description.

Longer-term advances rely on improvements in technology and informatics, biomimetics, and synthesis of large DNA fragments and their introduction into a variety of organisms, especially plants, which are currently transformation-limited. These, like many other revolutionary advances, will undoubtedly require synergy at the interfaces of several fields of study.

Although algae require basic nutrients for growth and have been engineered for highly efficient biodiesel production, additional redesign and engineering could allow algae to thrive in environments where combination technology could significantly help contamination mitigation efforts. Additional improvements concerning increased efficiencies in transport and energy cycles are foreseen, as well as in enzymology and microbiology, that could advance single-cell photosynthesis to rival land plants' ability to sequester carbon.

Applying the increased information from high-throughput approaches to genome sequencing and functional characterization, combined with computational advances and algorithm developments, could contribute to a comprehensive understanding of all the enzymes in a set of DOE-relevant reference organisms. Eventually, a complete *in silico*, genome-wide annotation of newly sequenced organisms could emerge as a result. These developments would significantly advance biosystems design by vastly increasing our knowledge and ability to select reagents for defined needs from sets of components with specific properties and characteristics. Currently unknown is whether every gene in a genome, even for DOE-relevant organisms, needs to be annotated to enable biosystems design efforts. Also uncertain is the level of accuracy needed for informatics-based annotation. Progress in this field will enable an assessment of these issues, as will more widespread investigations using biological redesign approaches.

Melding of biology and materials science for advancing biomimetics would have a significant impact on the DOE mission. Combining designed biological components with micro- and nanoengineered materials could transform field sensing and assessment outside the laboratory. Designing and developing stand-alone, nonreplicating tools is a compelling idea because they could be used to address specific challenges without attendant complications arising from the release of genetically modified organisms. For example, a stimulus-driven system for cell-free protein synthesis of enzymes *in vitro* could be developed for DOE-relevant applications.

Further out on the horizon, extraordinary capabilities for manipulating biological systems are possible, and there is tremendous excitement for newly designed functional biosystems. For instance, moving biosynthetic pathways from one organism to another could aid the development of new bacterial membranes composed of lipids from Archaea, making them more resilient for biofuel production. Enabling the inducible degradation of plant material by accelerating the deconstruction of lignin from cellulose for fermentation to ethanol also could enhance biofuel production. Other new designs are envisioned to increase plant efficiencies for nitrogen utilization and water uptake (see sidebar 8, Visions for the Future of Biodesign, p. 18).

Introducing large amounts of genetic material, which will be routine in microbes, also will be possible in plants, probably as entire chromosomes or, more likely, as multiple separate but simultaneous insertions, deletions, and replacements of genes or other DNA elements. This capability will occur in parallel with the increasing ability to manipulate plant regulatory elements to obtain tightly controlled and tunable functions. The resulting phenotypes of such newly designed plants would be predictable with a high level of confidence.

Crop performance, particularly for biomass, is determined not only by a plant's genetic makeup, but also by the microbial communities with which it interacts. Bacteria and fungi, as well as complex microbial communities, could be engineered to increase crop biomass, protect crops from disease, and reduce their fertilizer requirements. Such resources may bring us closer to the ambitious goal of sustainable biofuels production without increasing the environmental impact of agriculture.



## Sidebar 8

## Visions for the Future of Biodesign

One strategy to expand biomass production while potentially benefiting the environment is to create energy crops that grow on marginal or degraded land not currently useful for agriculture. Such crops can be engineered so that they not only yield biomass for biofuel production, but also help reclaim degraded environments. For example, expected progress in plant engineering could lead to new crops that express their own nitrogen-fixing enzymes without the need for bacterial symbionts. Such advances in biosystems engineering that enhance nitrogen utilization or water-use efficiency in plants would enable the development of new, sustainable energy crops. In addition to making plant cell walls more amenable to sugar extraction, lignin can be engineered to create useful products, such as high-value carbon fiber. In this approach, the fuel would become a byproduct, and the enhanced value of plant-based materials would facilitate economic sustainability.

Other biodesign strategies could improve energy crops by increasing carbon fixation efficiency or transferring the beneficial properties of perennial plants to annual crops. To enhance carbon fixation and sugar formation in crops that use the C3 pathway of photosynthesis, components from the more efficient C4 pathway of some crops such as maize and sugarcane could be engineered into C3 plants. To increase the sustainability of energy crops, annual crops could be re-engineered to acquire the advantages of perennial crops. Perennials are less harmful to the environment because they do not need tillage. These crops also use nutrients more efficiently by storing them in underground tissues after the growing season, leaving mostly the carbohydrates above ground to harvest for biofuel production.

With continued progress in structural biology, enzymology, membrane biology, and genome engineering, components of photosynthesis could be bundled into functional modules and introduced into nonphotosynthetic organisms. Successfully reconstructing and incorporating photosynthesis as a modular process into an organism would enhance our fundamental understanding of this process, define the minimum set of components needed to carry out photosynthesis, and reveal new combinations of components that can maximize photosynthetic efficiency.

Understanding stress in microbes is another critical issue. Almost every fermentation process is currently limited by the tolerance of the microbe to the final product. One way to overcome this problem is to modify cell membrane composition by introducing genes to synthesize new membrane lipids. For example, the lipids making up Archaea membranes are very different from those of bacteria. Thus, moving the pathways for lipid biosynthesis from Archaea to other microbes could create engineered organisms more tolerant to alcohol and thus more resilient to the stress of biofuel production.

Harnessing the wealth of information obtained from metagenomic sequencing of microbial communities in diverse environments is limited by the inability to culture and study many of the identified organisms in the laboratory. Advances in computer modeling are helping researchers study the metabolism of novel organisms *in silico* and identify nutrients, cofactors, and physicochemical conditions needed to grow and characterize these microbes *in vitro*.

## Conclusions

As we move further into the 21st century, dubbed the “century of biology,” the biodesign field is gaining momentum and is poised to play a central role not only in advancing the biological sciences, but also in progressing toward a sustainable bioeconomy. The forward-thinking group of experts gathered during the Biosystems Design Workshop foresees major technological breakthroughs in genomic engineering, facilitated by dramatic improvements in computational modeling and DNA transformation techniques for microbes and plants. Such breakthroughs will present virtually endless possibilities for redesigning plants, microbes, and their communities.

The scientific advances enabled by biodesign technologies will serve crucial needs in the production of advanced liquid fuels from plant biomass. Likewise, redesigned plant or plant-microbial systems will contribute to urgent needs, such as improving carbon sequestration and nutrient- and water-use efficiency in plants.

Identifying relevant functional modules that can be manipulated in different biological contexts and developing computational tools for modeling virtual biological systems will be critical for designing new biological functions that can achieve these ambitious goals. The extensive modification of existing regulatory and metabolic networks and the *de novo* design of new synthetic biosystems are powerful approaches for generating not only new useful biological functions, but also tools for conducting more informed interrogation of fundamental biological processes. Such advances will lead to a better understanding of the principles that rule these processes.

Complex features of biological systems—such as compartmentalization at all levels of biological complexity, cell-cell communication within both organisms and communities, and the evolutionary processes they are subjected to—will have direct biodesign implications. Such complexity poses major challenges for understanding biological systems in a

holistic way. Nonetheless, “top down” approaches are within reach for re-engineering relevant functions and modules that can deliver the expected outcomes without complete understanding of all the elements in the system under study. Furthermore, powerful new technologies, such as directed evolution of biological systems used for genome-wide redesign of bacterial systems, have great potential for applications in more complex model organisms.

An unshakable workshop conclusion is that society can expect stunning advances not only in our fundamental understanding of the nature, function, and evolution of biological systems and organisms as a consequence of increased biodesign research, but also significant potential for using this knowledge to solve outstanding problems of national and global interest. Next-generation biodesign scientists will be able to meet the challenge of using renewable biological systems to synthesize any chemical currently made from petroleum-based products or develop fast-growing, high-yield crops that thrive on marginal land with little water and no fertilizer.

These advances will rely on foundational research in molecular, structural, computational, and systems biology, as well as studies to increase the reliability and usefulness of synthetically designed biosystems. They will require increased integration of biological data—such as comprehensive and more accurate annotation of all the genes in a host of relevant organisms—with theory incorporating not only engineering principles but also the unique attributes of biological systems. For these advances to be generally useful, computational systems will be instrumental in increasing the relative ease of examining, analyzing, and assembling biological components for design. These developments should make biodesign affordable for widespread investigation and application and allow society to reap the economic and environmental benefits from what promises to become an “era of biosystems design.”



## Appendix 1: Workshop Agenda

### Sunday, July 17, 2011

6:00–8:00 p.m. *Informal mixer (Glen Echo Room)*

### Monday, July 18, 2011

7:30–8:15 a.m. *Breakfast (Salon A hallway)*

8:15–8:30 a.m. Welcome, introduction, and overview (DOE BER program representatives and co-chairs)

8:30–8:45 a.m. Around the table

**Plenary Session:** Three short introductory talks focus on visionary perspectives for the role of biological design in understanding nature and adapting it for unprecedented use (*Salon A*).

8:45–9:15 a.m. Plenary talk: Sarah Assmann, Pennsylvania State University

9:15–9:45 a.m. Plenary talk: Adam Arkin, Lawrence Berkeley National Laboratory

9:45–10:15 a.m. Plenary talk: George Church, Harvard University

10:15–10:35 a.m. *Refreshments*

10:35 a.m.–12:15 p.m. Breakout Session I

12:15–1:20 p.m. Lunch (*Salon A hallway*)

1:20–3:00 p.m. Breakout Session II

3:00–3:20 p.m. *Coffee*

3:20–5:00 p.m. Breakout Session III

5:00–5:30 p.m. Informal summary and general discussion  
(Are we asking the right questions? Are we getting good ideas?)

6:00 p.m. *Dinner (Salon A hallway)*  
Chairs and breakout leaders prepare for Tuesday

### Tuesday, July 19, 2011

7:30–8:00 a.m. *Breakfast (Salon A hallway)*

8:00–9:30 a.m. Additional questions on biological design principles, approaches, and applications

9:30–10:30 a.m. Presentations on breakout sessions

10:30 a.m.–12:00 p.m. Discussion

12:00–1:00 p.m. Working lunch, continue discussions (*Salon A hallway*)

1:00–1:30 p.m. Closing remarks (Co-chairs, DOE BER program representative)

1:30 p.m. Participants adjourn

1:30–3:00 p.m. Panel chairs, breakout session leaders, DOE BER staff

3:00–5:30 p.m. Writing session (Co-chairs, DOE BER staff)

### Wednesday, July 20, 2011

7:30–8:30 a.m. *Breakfast (Salon A hallway)*

8:30 a.m.–12:00 p.m. Writing session (Co-chairs, DOE BER staff)

## Appendix 2: Breakout Sessions and Charge Questions

### ***Breakout Session I: Biological Design Principles***

1. What are the fundamental principles and approaches needed to enable the facile design of molecules, modules, organisms, and communities?
2. What critical information is needed for biological systems (e.g., interactions among biological modules and processes, structures, and mechanisms) to provide a robust foundation for biological design?
3. What are the biological concepts that must be understood to integrate hierarchical modules to generate more realistic, complex, and useful functions of engineered biological systems?

### ***Breakout Session II: Strategies, Methodologies, and Approaches for the Design of Biological Systems***

1. What useful models could help establish general approaches for the design of complex biological systems at multiple levels of complexity [microbes, unicellular plants (e.g., algae), cell cultures, bioenergy crops, microbial communities, and microbial-plant communities]?
2. What are the key gaps in our knowledge, and what technological approaches are needed to design novel biomolecules, microbes, plants, and communities for specific purposes?
3. How can evolutionary processes be harnessed to accomplish forward engineering of biological systems from the gene to the community level?

4. What considerations of evolutionary or ecological dynamics are needed to effectively engineer systems with sustainable functionality?
5. How can cellular communication modules (intra- and/or inter-species) be used to coordinate and control synthetic systems?
6. What is needed to more fully integrate systems biology approaches to inform biological design principles?
7. What is the role of computational biology in the design of novel biological systems?
8. What computational biology tools are needed to facilitate design of molecules, networks, cells, and communities?
9. To what degree is compartmentalization (cellular, organismal, and community) important for developing designed biological systems?
10. What critical data are needed to advance the design of biological systems?

### ***Breakout Session III: Predictive Biology at the System Level***

1. What is on the short-, mid-, and long-term horizon for biological design that would advance bioenergy, bioremediation, and climate change research?
2. What are the basic and applied goals that biosystems research should pursue?
3. What major outcomes could be expected from research on biological design over the next decade?



## Appendix 3: Participants

Workshop Participants			
Group	Name	Institution	Leaders
1 (Salon A)	Peter Greenberg	University of Washington	Breakout Lead
	David Gang	Washington State University	Rapporteur
	Shota Atsumi	University of California, Davis	
	Mattheos Koffas	State University of New York, Buffalo	
	William Cannon	Pacific Northwest National Laboratory	
	June Medford	Colorado State University	
	Christina Smolke	Stanford University	
	Vipula Shukla	Dow Agrosiences	
	Adam Arkin	Lawrence Berkeley National Laboratory	Speaker
	John Glass	J. Craig Venter Institute	
2 (Timberland)	Costas Maranas	Pennsylvania State University	Breakout Lead
	Kechun Zhang	University of Minnesota	Rapporteur
	Christoph Benning	Michigan State University	
	Jay Keasling	Lawrence Berkeley National Laboratory	
	Gary Stacey	University of Missouri	
	Christopher Henry	Argonne National Laboratory	
	Ramon Gonzalez	Rice University	
	Scott Banta	Columbia University	
	Erich Grotewold	Ohio State University	
3 (Linden Oak)	Basil Nikolau	Iowa State University	Breakout Lead
	Maciek Antoniewicz	University of Delaware	Rapporteur
	Kristala Jones-Prather	Massachusetts Institute of Technology	
	George Church	Harvard University	Speaker
	Nitin Baliga	Institute for Systems Biology	
	Gary Saylor	University of Tennessee	
	Adam Guss	Oak Ridge National Laboratory	
	Kelly Dawe	University of Georgia	
	Sarah Assmann	Pennsylvania State University	Speaker

### Observers from the Office of Biological and Environmental Research

**Sharlene Weatherwax**, Associate Director

**Todd Anderson**, Biological Systems Science Division Director

**David Thomassen**, Chief Scientist

#### Program Managers

Dean Cole

Susan Gregurick

Cathy Ronning

Daniel Drell

Roland Hirsch

Prem Srivastava

Joseph Graber

John Houghton

Libby White

Arthur Katz

Shireen Yousef, Program Specialist





