Future Directions Workshop on Controlled Living Organisms

May 23–24, 2023

Adam Arkin, *Lawrence Berkeley National Laboratory*

Michael Desai, *Harvard University*

Ahmad S. Khalil, *Boston University*

Domitilla del Vecchio, *Massachusetts Institute of Technology*

Prepared by:
Kate Klemic, *VT-ARC*
Matthew Peters, *VT-ARC*

Future Directions Workshop series
Workshop sponsored by the Basic Research Office, Office of the Under Secretary of Defense for Research & Engineering

VT-ARC
Virginia Tech
Applied Research Corporation

Contents

Innovation is the key to the future, but basic research is the key to future innovation.

–Jerome Isaac Friedman,
Nobel Prize Recipient (1990)

Preface

Over the past century, science and technology has brought remarkable new capabilities to all sectors of the economy; from telecommunications, energy, and electronics to medicine, transportation and defense. Technologies that were fantasy decades ago, such as the internet and mobile devices, now inform the way we live, work, and interact with our environment. Key to this technological progress is the capacity of the global basic research community to create new knowledge and to develop new insights in science, technology, and engineering. Understanding the trajectories of this fundamental research, within the context of global challenges, empowers stakeholders to identify and seize potential opportunities.

The Future Directions Workshop series, sponsored by the Basic Research Directorate of the Office of the Under Secretary of Defense for Research and Engineering, seeks to examine emerging research and engineering areas that are most likely to transform future technology capabilities. These workshops gather distinguished academic researchers from around the globe to engage in an interactive dialogue about the promises and challenges of each emerging basic research area and how they could impact future capabilities. Chaired by leaders in the field, these workshops encourage unfettered considerations of the prospects of fundamental science areas from the most talented minds in the research community.

Reports from the Future Direction Workshop series capture these discussions and therefore play a vital role in the discussion of basic research priorities. In each report, participants are challenged to address the following important questions:

- How will the research impact science and technology capabilities of the future?
- What is the trajectory of scientific achievement over the next few decades?
- What are the most fundamental challenges to progress?

This report is the product of a workshop held May 23–24, 2023, at the Basic Research Innovation Collaboration Center in Arlington, VA on the future of research in the field of controlled living organisms. It is intended as a resource for the S&T community including the broader federal funding community, federal laboratories, domestic industrial base, and academia.

Executive Summary

The field of synthetic biology aims to create biologically-based systems that display complex functions like those seen in nature. In recent years, we have seen enormous growth in the field, from the creation of basic cellular circuits utilizing transcriptional, post-transcriptional and translational mechanisms to the identification of minimal viable genomes leading towards synthetic cells. In particular, significant work has been accomplished utilizing simplified signaling and regulatory circuits within single cell systems. This research may inform the eventual creation of user-controlled, programmable, multifunctional single (or multi) cell-based systems. These could be applied across a wide range of future capabilities, including but not limited to, wound healing, manufacture of fully functional and complex organs, smart and adaptive materials, and sensing. These advances may also lead to the design of new multi-cellular organisms capable of operating in diverse environments.

The *Future Directions Workshop on Controlled Living Organisms* was held on 23-24 May 2023 at the Basic Research Innovative Collaboration Center located at 4100 N. Fairfax Road, Suite 450, Arlington, VA 22203. It gathered 19 researchers from a variety of fields, including synthetic biology, molecular biology/biochemistry, control theory, systems biology, physics, mathematics, computer science, and bioethics to examine the prospects for applying new approaches, theories, and tools in basic research to enable these capabilities over the next 10-20 years.

The workshop was organized for highly interactive small group discussions with whole-group synthesis across three layers of development: characterization of systems, control of systems and design of systems. Participants determined the challenges, opportunities, and trajectory of research for each:

Characterization of Systems

To control a system, one must understand how to interface control systems to it and what internal and external (environmental) factors are likely to affect performance. The high priority research challenges and opportunities in the characterization of CLO systems include:

Transformation and Manipulation Technologies

Challenge: the ability to efficiently manipulate organisms beyond the widely used model systems

Opportunities

- Development of combinatorial transformation technologies
- Development of DNA protective design technologies
- Development of biomolecular "intervention" systems

Metrology

Challenge: collecting useful data for characterizing systems relevant to synthetic biology

Opportunities

- Increasing precision, accuracy and "resolution" of measurement
- Allow real-time non-invasive measurement of multiple factors *in situ*
- Tracking population variation and fitness
- Standardized protocols, experimental designs, analysis, and data reporting

Laboratory Twins and In Situ Measurement

Challenge: developing laboratory twins of deployment environments that allow us to vary critical parameters over the ranges they are likely to vary in the real world and be able to perform the measurements above in the *in situ* context

Opportunities

- Environmental/agricultural observatories with new hyperspectral imaging methods for tracking biological distribution that are fueling the development of laboratory twins at different scales
- Industrial bioreactors that are being outfitted with increasingly sophisticated online apparatus for monitoring operation and production alongside periodic sampling to exploit the advanced in omics technologies
- Development of laboratory scale devices to simulate *in vivo* environments (e.g. organ-on-a-chip)

Open, Scalable, Automated Laboratories

Challenge: enable multimodal measurement of laboratory twins over combinatorially large numbers of conditions

Opportunities

- Create automated measurement/characterization user-facilities using operational models that mirror at least in part sequencing centers, beamlines, and synthetic biology foundries

Data Collection, Representation, and Sharing

Challenge: make available a very large compendia of well-labeled and meaningfully structured data

Opportunities

- Metadata, unified identifiers, and ontologies
- Semantic structuring of biological and physical data
- Data repositories and unified data fabric
- Incentives for sharing and management of data, analyses, and models

Model-driven Design

Challenges

- Determine what to measure and in which conditions that will be most valuable and informative for the next experiments
- Determine how to infer from those measurements
- Determine how to use this information to better design the next iteration

- Reduce high data requirements and allow better data integration
- Harness the power of natural and artificial evolution to help us find novel solutions to engineering challenges involved in designing biological systems
- Integrating information across scales from models of molecular function to pathway/circuit dynamics, through to cellular behaviors, and population function *in situ*.

Opportunities

- Discovery and characterization of "endogenous" control systems
- Making evolution and ecology part of the equation

Workforce and Research Organization Limitations

Challenge: Composing multidisciplinary teams of geneticists, physiologists, application specialists, computational biologists, synthetic biologists, and engineers

Control of Systems

Based on the foundation of characterization of systems, participants made forward predictions of a system's behavior based on its current state and environmental inputs. The high priority research challenges and opportunities in the control of CLO systems include:

Learning the Nature of Endogenous Control

Challenge: Discover the "low"-dimensional control surfaces of endogenous regulatory systems that are obtained with respect to deployment environments.

Opportunity: Control objective-driven experiments to train/test models of controllability of the target plant in host/environment context.

Discover/derive new theories of control that address the special needs of biology

Challenge: Biological systems may require types of control beyond those developed for man-made systems.

Opportunity: Development of new theory and 'online architectures' for biological control.

Formal Specification of Control Objectives

Challenge: high uncertainty in the space of environmental perturbation and operation of system components, and (for some applications) the ambition of the control problem itself can make it difficult at the outset to define the core objectives and acceptable tradeoffs of the control circuit to be synthesized

Opportunity: If we can precisely state possible multi-objective performance functions linked to cost and benefit of the sub-elements, then pareto optimal solutions can be found, tradeoffs can be understood, and we can develop automated tooling for design of control algorithms.

Creation of Sensors, Controllers, and Actuators for Implementing Biological Control Systems

Challenge: There is not a standard, ready, scalable supply of compatible biological elements with which to build controllers in any target organism.

Opportunity: Creation of open-repositories of species-tested characterized elements from mined and generated 'parts' families.

Design of Systems

Design impacts every aspect of work in the field from the design of experiments for characterization of controllers and the controlled, to the generative design of new molecules for sensing novel signals and novel circuits for computing on these sensors and actuating responses. For true applications, it goes further because constraints on control performance—its cost, precision and accuracy and uncertainty/risk—are linked to other measures of impact that the system will have on its environment and the economic feasibility of the system. Therefore, the challenges and opportunities of design are baked into those described for characterization and control of systems.

Trajectory for Controlled Living Organism Research

The workshop participants developed a trajectory for the research opportunities identified for the field of controlled living organisms with a vision for the five-, ten-, and twenty-year horizons.

Five-year vision

The following research advances have a 5-year horizon:

- Non-invasive real-time measurements
- Systematic tools for domestication of biological diversity, creation of new model systems
- Curation methods for high quality datasets
- Quantitative models of evolution
- Improved approaches and incentive structures for data management and sharing
- Identify applications where control of systems will be critical
- Establish education, training, and research infrastructure enabling model-driven experiments and consequent data-driven models
- Incorporate AI/ML in the model-building process and make it appropriate for design (composable ML)
- Establish Lab automation infrastructure
- Establish infrastructure for data, model, and experimental protocol sharing
- Programmable insertion of large cassettes/precision genome modifications

Ten-year vision

The following research advances have a 10-year horizon:

- Non-invasive measurements that are simultaneously resolved in space and time
- Expanded genetic parts and tools for a range of domesticated organisms
- Extending evolutionary models to include complex interactions, including ecology
- Methods for characterizing and visualizing low-dimensional and nonlinear structure in high-dimensional data
- Development of AI/ML methods for description and prediction, and to mine existing knowledge.
- New (Input/Output) modeling frameworks that account for context dependence and can quantitatively predict behavior within 5% error
- New control design approaches and architectures, relying on the new modeling frameworks, that can handle unprecedented levels of uncertainty
- Enabling first applications with clear success metrics that without control of systems would not be satisfied
- Development of platforms that can mitigate context-effects (orthogonal molecular processes and artificial cells)

20-year vision

The following research advances have a 20-year horizon:

- Predictive and systematic abstractions of high-dimensional data
- Multiscale modeling that leverages improved abstractions to link dynamics across spatial and temporal scales and levels of biological organization
- Systematic methods to extract biologically meaningful interpretations of our representations of large-scale biological data sets
- Develop methods for accelerating the evolutionary process of variation-selection for the design of sophisticated control architectures
- (Bioinspired) control architectures that leverage evolutionary dynamics to achieve adaptation in changing environments

This workshop report outlines a path forward for the field of Controlled Living Organisms, with an architecture of development toward the creation of living, responsive, and possibly adaptive systems that can stably provide services over time and under uncertain and changeable conditions. We acknowledge that this path is challenging, with several fundamentally difficult questions that remain to be addressed. The workshop participants are optimistic that the field will overcome these challenges through development of new theories for describing the level of complexity needed to achieve reliable and predictive control, the ability to harness AI/ML methods for effective design, and ultimately when we can design biological systems that find their own solutions. We anticipate that these emergent technologies, when coupled with theory of control and evolution, will form the basis of an exciting field focused on applying controllable evolution to generate self-optimizing biological systems that drive toward optimum behaviors while, in turn, revealing the critical layers of design complexity for achieving such behaviors.

Introduction

The field of synthetic biology aims to create biologically-based systems that display complex functions like those seen in nature (Khalil, 2010; Andrianantoandro, 2006; Cameron, 2014). Inherent in this mission is the revolutionary idea that engineering approaches, which were developed to enable us to describe, design, and control systems but which have been largely unfamiliar and unexplored in biology, could be used to facilitate its manipulation toward productive ends.

In recent years, we have seen enormous growth in the field, from the creation of basic cellular circuits utilizing transcriptional, post-transcriptional and translational mechanisms to the identification of minimal viable genomes leading towards synthetic cells. In particular, significant work has been accomplished utilizing simplified signaling and regulatory circuits within single cell systems. This research may inform the eventual creation of user-controlled, programmable, multifunctional single (or multi) cell-based systems. These could be applied across a wide range of future capabilities, including but not limited to, wound healing, manufacture of fully functional and complex organs, smart and adaptive materials, and sensing. These advances may also lead to the design of new multi-cellular organisms capable of operating in diverse environments.

There are still many open and fundamental scientific questions that need answers to understand the complexity as one progresses from intra-cellular mechanisms to inter-cellular and systems scales. The current frontiers of controlled living organisms (CLO) span, for example, the ambition to engineer microbes for use in diverse sustainable feedstocks to produce high value products; to program the development and assembly of mammalian cells into functional tissues outside of the body that serve as valid models of disease or transplantable function; to intervene and program microbial communities *in situ*; and to improve nutrient mobilization in plants while sequestering carbon in the soil in a fashion resilient to the changeable, open

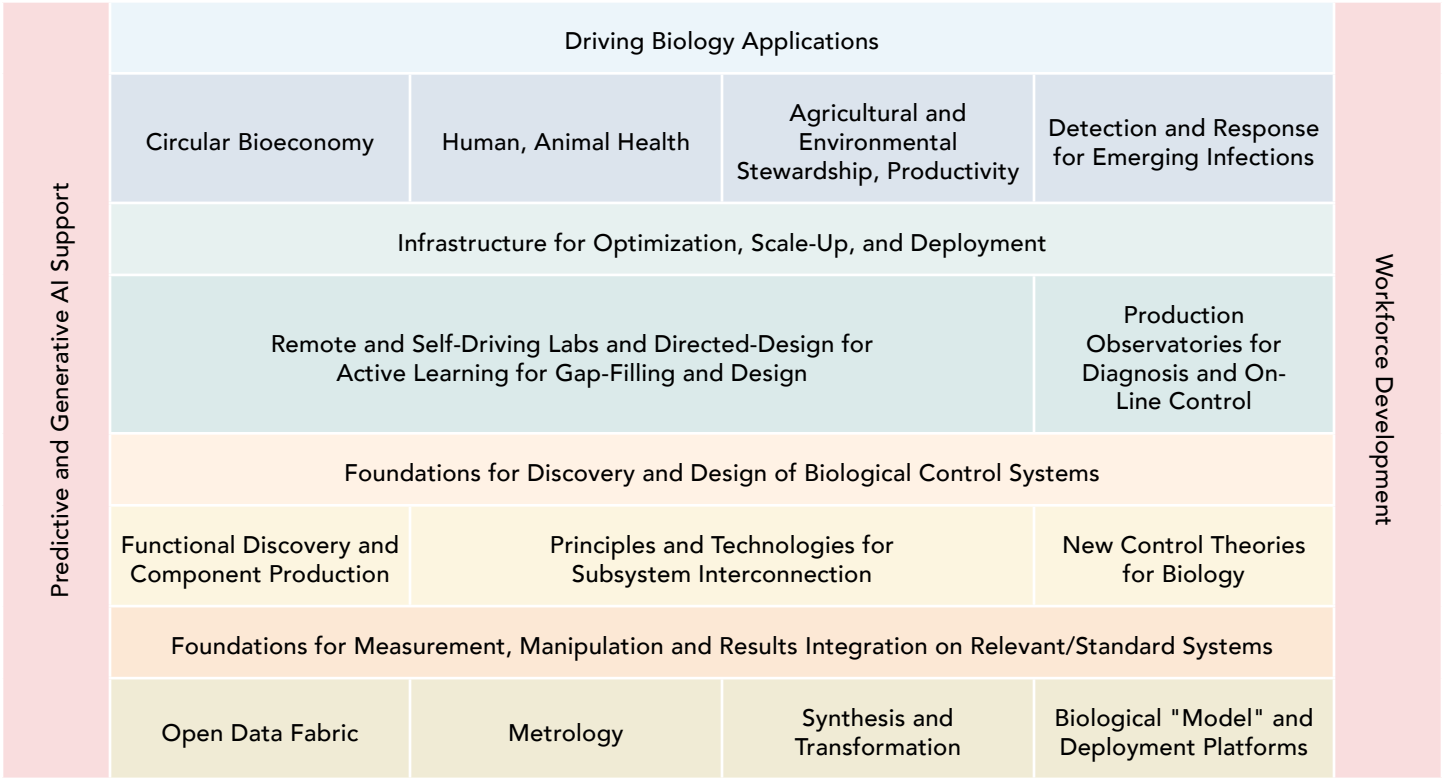


Figure 1 An "architecture" for the characterization, design, and control of living systems for applications within and beyond the bioreactor. Efforts build on a foundation which focuses on how to measure and manipulate (i.e., genetically change) organisms, their components, and their environments for (environmentally responsive) function in "validated" models of the target environment (bottom layer). This creates the context in which the biological elements that can accomplish control and connect to cellular "actuators" are discovered, and new control theories based on the nature of the biological system and its environment can be formulated (second layer). With this componentry in hand, infrastructure for assembling designs into organisms, testing their operation, and optimizing based on these operations is critical. Given the complexity and high dimensionality of the space, self-driving labs with active learning "controllers" to efficiently learn and explore the parameter space will become increasingly important. Similarly, once deployed, many of the biological designs and controllers will be augmented with abiotic controllers reading state and feeding back information to the organisms for online control need to be in place. These are also involved in quality assessment and control (third Layer). These are all scoped and driven by the different application areas. Across all layers we expect a deep integration of data representation, basic analysis, and advanced AI to be developed and integral to success. Further, the strong collaborative, multidisciplinary and multiscale nature of these efforts will require new workforce development across the field.

environments around them. These strike at critical national needs to accelerate the circular economy, solve critical problems in human health, and address the scalable biological components of climate change in a planet with diminishing resources.

The "holy grail" for CLO is the creation of living, responsive, and possibly adaptive systems that can stably provide services over time and under uncertain and changeable conditions. Figure 1 describes an architecture of development towards this vision. Scoped conceptually by application needs, the architecture layers characterization, controller design, and control implementation efforts in ways that would be synergetic and most effective. At the top, is the services provided (Driving Biological Applications) setting the requirements of the system across four big challenge areas. These might include: the scalable and economic biosynthesis of chemicals and materials in controlled bioreactors, but with variable feedstocks; the ability to provide persistent diagnostic, preventive or therapeutic functions in controlled and contained tissues of human and animal hosts or be 'programmed' to develop into such tissue; increased agricultural resilience and productivity by engineering plants to use variable resources in their soil, respond to stressors such as drought, brackish water, heat and pathogens while reporting on their status in remote detectable ways; increased persistent sequestration and storage of carbon and nitrogen in variable wetland, forest and agricultural soils to reduce greenhouse gas emission, improve nutrient status of soils and prevent erosion; and provide rapidly designable and deployable live therapeutics (e.g. bacteriophage and microbial competitors) for emerging infections.

This, and other properties of biological systems, requires the development of new theories and designs and opens the door for new approaches to control not available in non-living physical systems (currently).

Control, in this context, requires stable and predictable delivery of designed function given uncertain spatial and temporal variation of the environment in which the system is to operate. Specific to biology, is that these variations might not just affect the direct function of the CLO, but also its reproduction. Living systems can be outcompeted or evolve in the face of environmental challenges in ways that are difficult to control and predict. This, and other properties of biological systems, requires the development of new theories and designs and opens the door for new approaches to control not available in non-living physical systems (currently). For all control approaches we must start with the specification of a desired nominal behavior and an allied specification of the sources of uncertainty to which this behavior needs to be robust. The classical control engineering approach to this is founded on feedback and feedforward control architectures that "measure" some portion of the system, surrounds and processes the information, and then actuates some parts of the system to "steer" it towards the desired behavior under the uncertainty in the environment and the system itself. This basic approach has been an indispensable enabler of telecommunications and other signal processing, flight, and robotic motion. The benefits are increased higher precision of operation, robustness and resilience to perturbation and other uncertainties, both of which enable better modularity of function. However, the addition of control systems increases complexity and resource utilization and, when poorly designed or driven past theoretical limits, they can introduce instability and amplify noise. For more modern and complex systems, the advent of digital and even full computational controllers allows for more sophisticated control programs that integrate many signals and can adapt and learn about the perturbations against which they are stabilizing the system performance.

In biology, many of the properties of manufactured abiotic systems do not hold, so the well-known principles of design, manufacture, and operation of abiotic controllers do not apply. In biological control there is less partitioning between the process and its controller. The controller elements usually suffer the same extent of uncertainty and environmental sensitivity as the process elements, sensing itself can be a perturbative process as it may consume the very signal it measures. They often compete for resources to power their operation and have an impact on the overall cellular fitness leading to inadvertent feedback on their combined operation. This also leads to different "selection" on cells as controllers operate. Since biological systems are often reproducing under this selection, this can lead to population level divergence from expectations, and even amplification of "broken" mutated systems that lose control and take over the population. On the other hand, biological mutation does open the door to novel classes of adaptive and learning control if the rate, location, and nature of mutation itself can be under a controller design. These factors suggest that biological control laws will be complex—at a level that would be difficult even for a classical abiotic system- and yet implementing such controllers with biological parts is an even greater challenge. This is made more challenging by the paucity of biological parts to implement controllers, and the fact that a single type of "manufactured" part cannot be reused in same "system" easily since "wiring" is most often implemented by chemical specificity rather than spatial separation and interconnection by homogenous carrier materials, such as wires with electrons. There are other challenges as well: manufacturing biological control systems is still hard and expensive at the lowest level. Cells of different sorts differ in our capability to get DNA into. DNA synthesis and assembly remains expensive and somewhat error prone (even though this has improved greatly). There are not good, validated testbeds that allow scalable testing of biological prototypes in laboratory environments validated to simulate real deployment environments that allow rapid cycles of prototyping and testing that translate well to real deployment environments.

There have been advances towards solving some of these problems. New designs for low-burden biomolecular processes that use RNA as a key tool for sensing and actuating molecular reactions and orthogonalize the use of common resources; the use of spatial separation to partition processes and reduce unwanted contact and burden; the development of mathematical modeling tools that account for intra-cellular context to achieve quantitative prediction ability (at least in highly controlled conditions); methods for reducing and controlling mutation rates; improved measurement of genome and metagenome scale biomolecular abundances including at the single cell level with improved spatial and temporal resolution; and new AI-enabled design and optimization of circuit components. The development of organoid and organ-on-a-chip systems for laboratory simulation of deployment into human/animal systems and EcoPods for simulation of deployment in more environmental contexts are examples of progress in this direction.

These are indicative of, perhaps, a convergence that could propel the design of biological control systems to the next level and meet the challenge applications noted above. However, this will require new and foundational dynamical and control theories that account for the factors above; new technologies for measurement and manipulation of these systems; concerted efforts to design componentry and validated testbeds for critical applications; and infrastructure for supporting data science and scalable automated testing.

The *Future Directions Workshop on Controlled Living Organisms* workshop gathered researchers with a broad expertise spanning synthetic biology, molecular/cell biology, control theory, systems biology, physics, mathematics, and computer science to examine the prospects for applying these new approaches, theories, and tools in basic research to enable these capabilities. Workshop participants were divided into small groups working across three perspectives:

Characterization of Systems: We must identify the critical features of the biological system to be modeled, as well as its environment and the representation of their interaction and activity. From a physical perspective this would mean tracing the causal chemical and physical interaction of the cellular chemistry from the regulated transcription and translation of individual genomes, through metabolic and signaling systems, through cell growth and cell-to-cell interactions, to the active and passive dispersal of these cells and their aggregates up to organism level and beyond. While this accounting would be complete, it is difficult to achieve, and computation would be hard to scale. Thus, other levels of abstraction, physical and otherwise, need to be deployed and used together in reasonable ways. Here reasonable means that it is possible to experimentally observe and estimate the critical elements and the parameters of their interaction. How do we create multi-model representations of multiscale biological systems that can be effectively 'parameterized' by well-designed experiments to enable predictive power needed for control and design?

Control of Systems: Based on the foundation of Layer 1, we can make forward predictions of a system's behavior based on its current state and environmental inputs. Control theory allows us to determine what inputs we can make to the environment to move the state of the system to a desired outcome. Because of the high uncertainty, nonlinearity, and noise the environment and operation of most biological systems, and because of ignorance of much of their composite mechanisms of operation, there are new challenges in developing effective theories of control and building real control systems (cell external or even within the cell itself) that can observe the right variables and produce the right inputs at the appropriate time/space scales to achieve the outcome.

Design of Systems: Once we have a theory of control based on validated models of the target system, we can begin to design systems within and across cell (and organismal) populations to autonomously achieve outcomes through design of their endogenous systems and the environments in which they operate. In some ways, this is a natural extension of control. However, this opens a new fundamental avenue which is how to create a reliable engineering 'infrastructure' and 'supply-chain'. We need to be able to design novel functions (e.g., new regulators or metabolic activities) based on the needs of the problem, and we need to be able to add these elements into the biological systems and account for their 'loads' and off-target effects, as well as their defined function. This leads to new challenges in the design of biological systems that are currently very different in other disciplines.

These areas were chosen to focus initial mapping efforts that could identify the critical paths and approaches that would have the most impact in the next 10-15 years. In the course of writing this summary report, we discovered that the findings of Design of Systems are baked into the opportunities and challenges described for Characterization and Control of Systems, since design impacts every aspect of work in the field from the design of experiments for characterization of controllers and the controlled, to the generative design of new molecules for sensing novel signals and novel circuits for computing on these sensors and actuating responses. Therefore, the report organizes the opportunities, challenges, and trajectory primarily along the Characterization and Control of Systems with a brief description of how they tie into Design of Systems. At the end, there is a fundamental challenge in making the inventories of characterized systems, their data and design tools available in an open and standardized way to enable a broad technological community to engage in reproducible successful production of CLO for effective applications.

Characterization of Systems

To control a system, one must understand how to interface control systems to it and what internal and external (environmental) factors are likely to affect performance (see Figure 2). For example, consider a gut probiotic microbe to be engineered to biosynthesize a consistent dose of a therapeutic under variation in host conditions. There is a biosynthetic pathway to be controlled that has been engineered into a bacterial chassis. The expression of the genes in this pathway or the activity of its constituent enzymes and transporters draw from the chassis' resources. Both the pathway elements themselves and the chassis' physiology are impacted by changes in the nutrient and other physical conditions of the gut environment in which the chassis is growing. There may be tradeoffs between pathway activity and host fitness - its ability to grow and persist in the gut community. Now, consider an endogenous biological controller: a genetically encoded system to ensure a relatively constant production of the therapeutic over time and host condition. It is also using host resources, and its components are similarly subject to the environmental variation of the host. To achieve control, the designer needs to be able to: 1) transform the chassis to efficiently insert the DNA encoding the pathway and controller; 2) understand what intermediates are present on which the pathway can draw to create its products; 3) quantitatively understand the possible burden on the cell induced by production of pathway and controller components; and, 4) exploit natural and engineered mutational processes to allow both neutral and adaptive evolution of the cellular population. Knowledge of the causal and mechanistic basis of the observations would be most effective in informing designers how to mitigate hurdles in manufacture of the system and its operation. However, in the absence of precise cause and mechanism, effective measurement of the system in relevant environmental conditions can allow more statistical understanding of the chassis, pathway, and controller element functions and their relationship to performance goals.

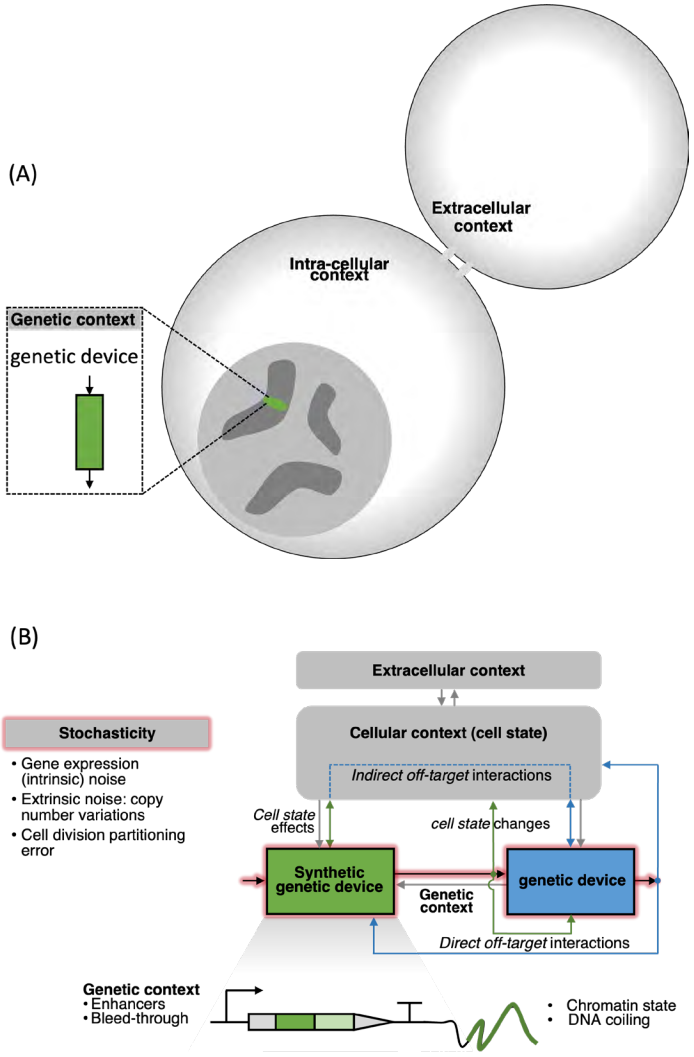


Figure 2 Dependence of synthetic genetic systems on their environment (context). (A) Genetic context includes the activity of surrounding genes, the direction of surrounding genes, and chromatin state. Cellular context includes off-target interactions, inter-module loading, resource sharing, growth rate feedback, cell state, and in general any hidden interaction of the engineered system with cellular and other synthetic systems. Extracellular context includes Competition for nutrients with other species/strains, temperature, pH, acidity, oxygen, to name a few, cell-cell signaling, and the cell niche. (B) Block diagram representation of how the environment affects any synthetic genetic circuit that we plug in the cell. [Adapted from Shakiba, 2021]

Research Challenges & Opportunities

Below we outline some of the high priority research challenges in the characterization of CLO systems, which suggest important directions in which research in the field should progress. Meeting these challenges and opportunities for characterization is beneficial not only to the field of CLO but for the field of biology at large because fundamental understanding of biological mechanisms and operation in diverse environments is a goal of all biological fields.

Transformation and Manipulation Technologies

Challenge: Most extant applications and studies in controlled living systems have been limited to a small number of model species or model cell lines (Fatma, 2020). There is a strong need to expand this set of organisms to meet the demands of the diverse applications and their attendant resources and environments. However, the ability to efficiently manipulate organisms beyond the widely used model systems has extremely variable success. First, it is difficult to get DNA into many cell types, and once inside keep it stable and replicating either as an autonomous replicon or by insertion into the genome. Further, a great deal of precise biological characterization requires the ability to manipulate the internal systems including gene knockdowns/outs, overexpression, and other such genetic manipulations. Finally, we need to be able to quantitatively and combinatorially perturb biomolecular functions in time and space if we are going to create predictive models of their functions. In some cases, it would be advantageous as well to perturb *in situ*.

Opportunities

- **Development of combinatorial transformation technologies**, including creation of new genetic elements, libraries of engineerable transducing phage and conjugation systems that facilitate transfer of DNA in the lab and *in situ*; and more basic science to understand the mechanistic basis for factors that impact transformation and recombination outcomes. Allied with this is the support of technologies that decrease the cost of long, reliable DNA synthesis.
- **Development of DNA protective design technologies** since most chassis have defense systems that protect themselves from intrusion of foreign nucleic acids. There are opportunities to determine properties like codon usage and methylation patterns so that DNA can be designed to be more easily translated and less easily degraded.
- **Development of biomolecular "intervention" systems**, including 1) new genetic editing methods that modify DNA "permanently"; 2) development of controllable extra-chromosomal replicons that allow for more rapid insertion and prototyping of circuitry; 3) controlled mutation systems to understand the functional landscape accessed by mutation around a "wild type" variant of the system (these systems may also have dual use as elements of adaptive controllers as noted below); and 4) activity modulation systems, including constitutive and inducible promoters, CRISPR-i/a type systems, and various modular regulatory elements for RNA and protein.

Metrology

Challenge: There have been astounding advances over the past twenty years in methods for measuring the structure and function of biomolecules, cells, and their superstructures, sometimes even *in situ*. However, many of these are not universally applicable to the systems under study, limiting our ability to collect useful data for characterizing systems relevant to synthetic biology. For example, many of these are highly perturbative making it difficult to track coherent behaviors over time and space and/or lack the temporal and/or spatial resolution necessary to capture the dynamics necessary for modeling control. Finally, describing the evolutionary landscape requires strain-level tracking to detect the birth and fate of new genetic variants over space and time in the relevant environments.

Opportunities

- **Increasing precision, accuracy and "resolution" of measurement** including through better measurements of dynamic structure/activity of biomolecules and OMICS techniques scaled to be sensitive enough to measure single cell molecular abundances ideally with spatial dimensions at the subcellular level.
- **Allow real-time non-invasive measurement of multiple factors *in situ*** to enable dynamic characterization of open-loop and controlled systems.
- **Tracking population variation and fitness**. Sequencing technologies are enabling highly accurate tracing of allelic variation and strain identity in evolving populations, and when paired with insertion of genetic barcodes, can track emergence of differentially fit mutations. There is opportunity to combine these techniques with model-driven efforts and high-throughput genetic techniques (CRISPRi or RB-TNSEQ) to map epistatic evolutionary surfaces to understand challenges and opportunities in performance optimization, control, and containment.
- **Standardized protocols, experimental designs, analysis, and data reporting**. This is critical both for understandable, open, and reproducible science and to ensure data can be reused and combined with other measurements. This is especially critical if biology is to take advantage of advances in artificial intelligence and machine learning (AI/ML) which is accelerated by well-labeled, high quality, structured data.

Laboratory Twins and In Situ Measurement

Challenge: The behavior of biological parts and systems are exquisitely context dependent. To ensure the maximum utility of measurements for design and deployment of systems in diverse environments, it is necessary to measure them in realistic settings. To that end, we need to both develop laboratory twins of deployment environments that allow us to vary critical parameters over the ranges they are likely to vary in the real world and be able to perform the measurements above in the *in situ* context. Current examples remain highly bespoke, unscalable, and unreliable. Likewise, most current *in situ* measurement systems lack sensitivity, spatial or temporal scale, specificity to the deployed system and its immediate surrounds, and/or scale.

Opportunities: There are great opportunities to integrate with, improve upon, and build on efforts currently being developed across different application spaces for detailed observation of biological entries and simulations of *in situ* environments. This includes: environmental/agricultural observatories with new hyperspectral imaging methods for tracking biological distribution that are fueling the development of laboratory twins at different scales; industrial bioreactors that are being outfitted with increasingly sophisticated online apparatus for monitoring operation and production alongside periodic sampling to exploit the advanced in omics technologies; and development of laboratory scale devices to simulate *in vivo* environments (e.g. organ-on-a-chip). Each of these comes with their own unique opportunities and challenges that need to be solved to support rapid design, build, test and learn cycles for CLOs that are predictive of operation at scale.

Open, Scalable, Automated Laboratories

Challenge: An extension to the laboratory twin is driven by the massively combinatorial nature of both the characterization and controller design process. In general, many different conditions for many different variations of cellular designs need to be explored. While there is opportunity to explore this space intelligently, there is still a need to enable multimodal measurement of laboratory twins over combinatorially large numbers of conditions. Attempts to create laboratories capable of carrying out sophisticated genetic experimentation are few and far between, tend to be highly specialized and

bespoke for the group running it, and are still only partially automated. Automation for this will be key, as will increasing our sophistication of biological data science and workforce training.

Opportunities: There is a great opportunity to create automated measurement/characterization user-facilities using operational models that mirror at least in part sequencing centers, beamlines, and synthetic biology foundries. This requires 1) standardized ecology of experimental controllers, driven by a common operating "language" for example, and adherence to standards for size, power, interface with integrating robotic systems, and for data management that would enable computer-control of input to and measurement of possibly multiple "copies" at a time of the laboratory twin systems above, 2) an operating system for configuration, control, and query of possibly distributed characterization systems, and 3) support for all by robust data transfer, storage, representation, and access framework. This has an amplifying effect of possibly decreasing cost over time by increasing the statistical power of data by integration with past measurements of related systems by other users.

Because of the high dimensional space for cellular function, for environmental parameters, and for the design space for controllers we need efficient ways to explore and model it to test key hypotheses or to accurately train predictive models for control and design.

Data Collection, Representation, and Sharing

Challenge: One of the true limitations to effective design is how to find, organize and collect the critical data necessary to both inform your designs and diagnose limitations in its performance. It is impossible for one lab or even one organization to collect all the requisite information. Thus, there is a premium on effective data sharing and data interoperation, harmonization, and integration. There has been a great deal of improvement in standards for data sharing, including the recent rise of standards for making data Findable, Accessible, Interoperable and Reusable (FAIR). However, there are no generally accessible open systems capable of sharing the data, protocols and models that are necessary to support a community engaged in the design of CLOs; and there are social challenges in encouraging people to build open infrastructure and make the data, analyses, and tools within as open as possible (Mante, 2023). If we are to take advantage of the rapidly moving advances in data science, we need to make available a very large compendia of well-labeled and meaningfully structured data for these tools to work on.

Opportunities

- **Metadata, unified identifiers, and ontologies.** CLO requires a level of metadata not generally necessary for less quantitative and dynamic fields.
- **Semantic structuring of biological and physical data.** We need to arrange our data in ways that express their organization and physical relationships so that models can be more easily built or inferred from the compendium of available information.
- **Data repositories and unified data fabric.** This should also include support for curation that ensures high quality labeling of data, development of quality assessment, and organization of high value data sets.
- **Incentives for sharing and management of data, analyses, and models.** Social challenges remain the largest challenge to data sharing. To overcome these, journals should provide or leverage effective means for both sharing and curating data using open standards and repositories. At the same time, journals and funding agencies should increase the consequences for not sharing well-curated data.

Model-driven Design

Challenges: Because of the high dimensional space for cellular function, for environmental parameters, and for the design space for controllers we need efficient ways to explore and model it to test key hypotheses or to accurately train predictive models for control and design. Significant challenges exist for determining what to measure and in which conditions that will be most valuable and informative for the next experiments; how to infer from those measurements—which often provide exceptionally high dimensional information—the most significant features; and how to use this information to better design the next iteration.

Currently, we use a disorganized collection of different modeling approaches that range from biophysical to statistical, almost none of which are widely adopted or generalizable. There are several commonly used purely statistical methods that attempt to reduce the "dimensionality" of the data and infer critical features predictive of some phenotype. Clustering techniques (e.g., PCA, tSNE, UMAP) attempt to discover some "grouping" of related observations that represent some of the larger covariation structure in the datasets. While useful, this does not make specific assertions about feature importance for prediction or control. More or less causally structured statistical models (e.g., deep learning neural networks) can, with generally high data requirements, take both raw or dimension-reduced data across a measurement set and link specific features (e.g., biological elements) to phenotypes/behaviors. One proposed way to reduce high data requirements and to allow better data integration is to introduce more causal/biophysical information into these modeling cycles. An impressive recent example is protein structure prediction and design using fully biophysical, mixed physical and statistical and ML based methods (tools like Rosetta and Alpha Fold) (Leman, 2020; Jumper, 2021). However, even these are hampered by

lack of actual data about protein function/activity. Finally, an important research challenge for synthetic biology is to harness the power of natural and artificial evolution to help us find novel solutions to engineering challenges involved in designing biological systems. On the flip side, evolution has the potential to disrupt the design and function of systems, as spontaneous mutations, and natural selection acts to degrade intended functions. These challenges demand a predictive quantitative understanding of evolutionary dynamics, which is a prerequisite to any attempts to design systems and processes either to harness evolution or to limit its effects.

Finally, there is a challenge in integrating all this information across scales from models of molecular function to pathway/circuit dynamics, through to cellular behaviors, and population function *in situ*. At the center of the challenges, of course, are the complexity of the systems in that they are high dimensional, nonlinear, stochastic, and operate across scales in some cases from single molecules to ecologies. To develop unified multiscale models and predictions that can correctly bridge these scales and levels of organization, we must understand what can be ignored and what details are essential: we must find the correct abstractions. There are a few examples of machine-learning linked designs of experiments for designing CLOs; but these too tend to be bespoke and specialized. These challenges are deeply intertwined with the challenges described above related to data collection, organization, and sharing.

Opportunities

- **Discovery and characterization of "endogenous" control systems.** The characterization of endogenous (evolved) control systems both may provide principles of design of synthetic controllers and may provide insight in where to place sensors and actuators within the system to achieve the desired effects. There is a fundamental goal to determine what we hypothesize will be relatively low-dimensional control surfaces (compared to the high dimensionality of the molecular state space of the system) which map specific cellular and environmental parameters to performance outcomes. Methods that characterize these surfaces effectively will help to formally define levels of abstraction and allow more efficient design and implementation of controllers. There is an opportunity to leverage the complementary advantages of statistical and more causal methods together with formal model-driven design of experiments to create efficient characterization of both controllers and controlled systems.
- **Making evolution and ecology part of the equation.** One key area for model-driven design is in evolutionary dynamics, where development of stochastic dynamical models of the interactions between mutations, natural selection, and genetic drift offers the opportunity to direct and control evolution to achieve design and control aims. Extending these evolutionary models to address more complex interactions (e.g., metabolic feedbacks and ecological interactions) is an emerging frontier. Opportunities exist here to both integrate evolutionary dynamics with existing dynamical models of genetic and metabolic networks, and to combine these with ecological theory. These models can then help us leverage evolution to solve problems as a mechanism-independent approach that is complementary to traditional synthetic approaches.

Workforce and Research Organization Limitations

Finally, the activities noted above generally require people from diverse backgrounds to work together—geneticists, physiologists, application specialists, computational biologists, synthetic biologists, and engineers generally do not work together, even on relatively small, specialized projects. Composing such multidisciplinary teams has been challenging, and it is difficult to create and maintain intellectual and technical centers (e.g., Foundries) where the principles and practice of this work are accelerated and deployed for use by the larger community effectively. We need to: 1) design better training routes for workers in this area including curricula designed for team science, 2) develop opportunities to build dense local intellectual communities to solve problems, and 3) create infrastructure to support this scalable and open access measurement, including support of multidisciplinary centers that integrate research and outreach.

Control of Systems

Concepts of control endogenous to biological systems have been entwined with the observations of robust and repeatable development, homeostasis of metabolic and thermoregulation, the ability to track and move towards sources of food or away from toxins, and to detect and respond to immunological threats. In natural systems, there are clear examples of control architectures like what a control engineer might expect: there are evolutionarily convergent examples of feedback and feedforward control that, for example, stabilize a concentration level of a protein or reject noise in an input to a signal transduction pathway respectively; biological integral feedback controllers have been found in chemotactic systems; and layered control systems that show pareto optimal expression of heat shock proteins to stabilize system function.

While these motifs and their functions have been verified, it is not clear if there are principles of control design local to a given cellular subsystem or across cells and tissues that are like what we find in artificial systems. Strong functional modularity with high impedance interconnection, strict isolation of information paths through "orthogonal" wiring systems, and "design" to prevent state-dependent fluctuations in the power system for the plant and controllers do not seem to be as evident in evolved systems. Natural biological systems show only modest evidence of subsystem modularity, hierarchical control, signal path modularity through use of different chemical specificity of system constituents, and minimization of 'load' on systems resources. And, while these systems may have arrived at recognizable control 'laws' for parts of their function, they were not constrained by design and manufacturing standards, which means that any means necessary for the system to respond to their complex environments sufficient to ensure reproduction could be used.

There are three other fundamental aspects which make designing controllers for living systems different than those for most other abiotic systems. First, in general, designers of cellular-level controllers, for example, may at first be thinking of the cell as the unit to be controlled. However, it is likely in this application that there will be from hundreds to billions of copies of that cell in the deployment each of these experiencing a slightly different microenvironment that in turn affects each cell's available power, the signals entering the plant and controller, and (very distinct from abiotic systems) its growth rate and competitive advantage compared to other cells. Second, when such controllers are operating at the biomolecular/cellular scale the low numbers of molecules and wide range of time scale of molecular activity can lead to highly idiosyncratic stochastic dynamics. Third, one famous stochastic effect is the mechanism of mutation that occurs in all biological systems that leads to continuous "irreversible" diversification, which can lead to changes in the host background that violate the assumptions of the controller, change or break the controller, or change/break the plant.

There may be new principles of control we can infer from these involved multilayered and novel control systems through the characterizations above though it may be that some of these are undesirable for the designed control systems we want to implement. These integrated control systems have evolved to respond to multiphysical, multivariate, time and spatially resolved signals "expected" in their environments; inferring from the system itself what the critical environmental parameters are that need to be sensed and operating on could be useful for future synthetic control system designers. We also hypothesize that many of the critical behaviors of cells have evolved to have relatively simple, low-dimensional control surfaces with respect to key parameters. For example, the temperature and pH dependence of microbial growth rates both show roughly Arrhenius curve-like behaviors despite the underlying complexity of the systems' response mechanisms. These hold out promise that for practical applications such low dimensional abstracted representations could be discovered.

It is critical, therefore, at the outset of any project, for the CLO designer to formally specify the controller. Controllers are meant to constrain the envelope in which plant dynamics are allowed to operate. Thus, one of the challenges for CLO designers is definition of the formal problem objective incorporating performance metrics that address the types of disturbances expected in the context of the controller/plant host, its population, and under mutation. As in all control systems, there may be tradeoffs in state accuracy, response speed, power efficiency, robustness, and risk of failure. To address all of these given the novelty of the biological medium and environment will require new theories of control and methods for system and environmental abstraction. We can expect that the complexity and novelty will differ with application from simpler controllers for biosynthetic production of high value chemicals in stable biochemical reactors at one end, operation of signaling circuits in single autonomous immune cells (such as with CAR-T immunotherapies) in less controlled but still homeostatic human environment in the middle, and control of beneficial microbial communities at agricultural plant roots in open fields at the other extreme.

Of course, to achieve any controller, the CLO engineer must have elements to "implement" the design. Even if part of the controller is abiotic, there needs to be a set of readily available and/or designable biomolecular sensor, controller, and actuator parts. Parts, such as light and chemical sensitive proteins, must be able to have the sensitivity and specificity necessary to respond to the environment. Similarly, the temporal response must be fast enough relative to the signal and plant dynamics to exert the appropriate control. Further, these elements need to be compatible with each other. In many cases there need to be sets that are orthogonal to one another with no overlap in the input signals or the specificity for their output targets (e.g., binding sites for gene expression, phosphorylation sites on target proteins). Also, ideally, they will not together overuse a

resource necessary for powering them or their host (e.g., overuse the ribosome budget, or consume too much ATP). These need to be sufficiently characterized for a given host and environment such that computational tools can 'synthesize' a biological implementation of an "abstract" control program specified by the designer.

Research Challenges & Opportunities

The above considerations then specify a set of research challenges and opportunities:

Learning the Nature of Endogenous Control

Challenge: Discover the "low"-dimensional control surfaces of endogenous regulatory systems that are obtained with respect to deployment environments.

In essence, find the critical locations in the system where variation of its parameters is effective in modifying its performance. The key hypothesis is that these represent a much smaller set of variables than the entire space of possibilities and that these drive system outputs on relatively constrained dynamical manifolds determined by the underlying physics.

Opportunities: Control objective-driven experiments to train/test models of controllability of the target plant in host/environment context.

Given the control objectives defined, validated Laboratory Twins (defined in Characterization of Systems), and perhaps constraints on intervention types that a conceptual controller could have, design the minimum time/cost experiments to train and test models mapping environmental and system parametric variation to variation in performance measures. Depending on the complexity of the system to be controlled and the knowledge about it, models could range from mechanistic to purely statistical and places in between. There is still a great deal of work to do to learn how to define reasonable and compatible hybrid mechanistic and statistical models of complex systems like this.

Discover/Derive New Theories of Control that Address the Special Needs of Biology

Challenge: Biological systems may require types of control beyond those developed for man-made systems.

Biological systems are multiphysical, nonlinear, stochastic, and discrete (meaning non-differentiable in some ways). Differentiating controllers that operate on single cells, coherent (meaning collaborating) or independent populations is novel and important when considering biological control design. Systems will be far less modular, hierarchical, and insulated from each other in terms of resource utilization. The environments can be unpredictably changeable and exert differential selective pressure on the biological system which, unlike abiotic systems, is growing and reproducing at different rates in these environments. Finally, the biological systems may be mutating in more or less directed ways, leading to neutrally and adaptively diverse populations containing the designed controller. Nearly all these features violate standard assumptions of classical control and others have only barely been addressed whereas they will be central in biological control.

Opportunities: Development of new theory and 'online architectures' for biological control.

Results from work to discover the architectural principles of natural biological system operation above provide an opportunity to learn the evolved control architectures that achieve their specific types of goals which couple fitness to activities in their native environments. One of the most interesting of these is, of course, the process of heritable epigenetics and mutation; given newly harnessed mechanisms that allow us to control (epi)mutation rates/types to targeted genomic locations we have an opportunity to exploit these for new types of control algorithms if we can feedback the results effectively and control performance and relative fitness of "selected" mutants. Indeed, controlled mutation is one class of method for dealing with the evolving environments encountered during system operation. In traditional engineering applications, such as robotics and autonomous vehicles, control systems achieve these through sophisticated control architectures, which can include recurrent and deep neural networks, thereby allowing the control systems to learn about the environment "as we go" and change control policies accordingly. This is possible because we can implement these sophisticated control laws on a computer. It would be rather difficult to implement them through electro-mechanical circuits and even more so in biological circuits. While some progress might be made by coupling biological control systems to abiotic systems which can compute on measured signals, in the long run we need to find a way to implement these sophisticated functions through biomolecular processes, cell-cell interactions, and more. This may entail new types of architecture that look less like classical digital logic or feedback control loops and more like neuromorphic architectures.

Formal Specification of Control Objectives

Challenge: Formal specification of the control objectives in biology are complicated.

The complex interconnected nature of biological systems, high uncertainty in the space of environmental perturbation and operation of system components, and (for some applications) the ambition of the control problem itself can make it difficult at the outset to define the core objectives and acceptable tradeoffs of the control circuit to be synthesized.

Opportunities: If we can precisely state possible multi-objective performance functions linked to cost and benefit of the sub-elements, then pareto optimal solutions can be found, tradeoffs can be understood, and we can develop automated tooling for design of control algorithms.

Creation of Sensors, Controllers, and Actuators for Implementing Biological Control Systems

Challenge: There is not a standard, ready, scalable supply of compatible biological elements with which to build controllers in any target organism.

The biological elements of controllers are difficult to standardize in both function and manufacture, there are not generally common "protocols" for their interconnection, and they often cannot transfer across organisms and environments and maintain their properties. Further, many can be resource intensive or have non-beneficial secondary effects There is often not a common core "class" of each controller element like there might be in electro-mechanical controllers which can reuse copies of the same element in different parts of their circuit and can manufacture variants with known formal changes in function. This leads to high overhead in characterization of the large numbers of diverse parts for each type of "function" in the controller and large modeling error of composite systems. Overall, these issues prevent rapid prototyping and scaling of designs and the development of automated design and 'manufacturing' software.

Opportunities: Creation of open-repositories of species-tested characterized elements from mined and generated 'parts' families.

Most controllers built to date have exploited a relatively limited set of modestly modified natural promoters, transcription and translational control elements, protein-protein interaction systems, sensors, and signal producers (e.g., quorum sensing molecules). However, there have been a few key innovations in recent years that could be exploited. First, the discovery and engineering of RNA-guided DNA and RNA modifying proteins has opened the door to scalable, and possibly cross-species, design of orthogonal transcriptional and translational regulators, focused mutation systems, and copy-number control systems. High-throughput characterization and machine learning based methods could be used to derive improved/generalized models for target specificity and activity. There has been evidence as well of families of portable recombinases and other programmable sequence specific proteins that could be used to implement complex control circuits. However, while small case studies exist for these and other types of elements spanning optogenetic controllers to phospho-relay componentry and electrical signal sensing systems few of these show the ability scale within a system or across systems. The most programmable of these elements are also "slow" compared to the control response needed in many cases, so the need to move beyond transcription and translation in the core controller is urgent. Finally, there may be a place for "designed" deployable artificial environments that contain the system and control the variation in the environment. For example, encapsulation of deployed microbes into the soil encapsulated in protective hydrogels containing critical nutrients and buffers could be a novel component for controller design. There is perhaps a thread to pull using generative AI methods that can use large libraries of measurements mapping sequence (and other measurables) to function to generate new members of a class. While early in its development this points a way to create compatible families of parts given sequence/environment function data measured in the automated labs above. It is necessary however to design the performance measures carefully including specificity, sensitivity, temporal response, load/toxicity to the cell/organism, etc. It is an open question how to do this in a scalable generalizable manner. The other essential aspect is sociological - the data, models, generated molecules, and resultant performance measures must be made findable, accessible, interoperable, and reusable (FAIR) as possible so that the community can learn from each other and advantage themselves of this network effect.

Design of Systems

Design is baked into the opportunities and challenges above since it impacts every aspect of work in the field from the design of experiments for characterization of controllers and the controlled, to the generative design of new molecules for sensing novel signals and novel circuits for computing on these sensors and actuating responses. For true applications, it goes further because constraints on control performance—its cost, precision and accuracy and uncertainty/risk—are linked to other measures of impact that the system will have on its environment and the economic feasibility of the system.

Organization of the resources discussed for characterization and control of systems into a multi-scale, multimodal design system—one backed by a computational infrastructure allowing designers to computationally accurately synthesize and compare different designs for performance, safety/risk, and cost — is rife with challenges. Currently there are few commonly used design tools, and these are generally fairly low level, such as ribosome binding site designers, some basic cloning route designers, and a few generally used tools for macromolecular design (Salis, 2009). There have been demonstrations of more sophisticated biological pathways and circuit design tools, but these have proven to be limited and bespoke and have not propagated deeply into the broader community. Industry has begun to have more success in building largely internal and private suites of design-support systems that include computation tools for biomolecular and circuit design, cloning route mapping, control of characterization experiments, bioprocess engineering, and techno economic analysis.

There is also a good argument to include metrics of sustainability and socioeconomic equity and impact as many applications of CLO are in spaces where these are front and center such as in food production/protection, land stewardship, or health protection. These are generally deeply supported by proprietary reagents- application specific biological strains, transformation systems, genetic parts, etc. and standardize characterization and scale-up experimental designs that make their bespoke computational tools more effective since they are trained with these resources. However, all of these are largely unavailable to the academic community. There is significant opportunity to organize the community towards creating an open ecology of computational tools, training data and core reagents on which common understanding can be built and which are compatible with the open foundry laboratories with validated laboratory twins discussed above.

Research Trajectory

The workshop participants developed a trajectory for the research opportunities identified for the field of controlled living organisms with a vision for the five-, ten-, and twenty-year horizons.

Five-year vision

The following research advances have a 5-year horizon:

- Non-invasive real-time measurements
- Systematic tools for domestication of biological diversity, creation of new model systems
- Curation methods for high quality datasets
- Quantitative models of evolution
- Improved approaches and incentive structures for data management and sharing
- Identify applications where control of systems will be critical
- Establish education, training, and research infrastructure enabling model-driven experiments and consequent data-driven models
- Incorporate AI/ML in the model-building process and make it appropriate for design (composable ML)
- Establish Lab automation infrastructure
- Establish infrastructure for data, model, and experimental protocol sharing
- Programmable insertion of large cassettes/precision genome modifications

Ten-year vision

The following research advances have a 10-year horizon:

- Non-invasive measurements that are simultaneously resolved in space and time
- Expanded genetic parts and tools for a range of domesticated organisms
- Extending evolutionary models to include complex interactions, including ecology
- Methods for characterizing and visualizing low-dimensional and nonlinear structure in high-dimensional data
- Development of AI/ML methods for description and prediction, and to mine existing knowledge.
- New (Input/Output) modeling frameworks that account for context dependence and can quantitatively predict behavior within 5% error
- New control design approaches and architectures, relying on the new modeling frameworks, that can handle unprecedented levels of uncertainty
- Enabling first applications with clear success metrics that without control of systems would not be satisfied
- Development of platforms that can mitigate context-effects (orthogonal molecular processes and artificial cells)

20-year vision

The following research advances have a 20-year horizon:

- Predictive and systematic abstractions of high-dimensional data
- Multiscale modeling that leverages improved abstractions to link dynamics across spatial and temporal scales and levels of biological organization
- Systematic methods to extract biologically meaningful interpretations of our representations of large-scale biological data sets
- Develop methods for accelerating the evolutionary process of variation-selection for the design of sophisticated control architectures
- (Bioinspired) control architectures that leverage evolutionary dynamics to achieve adaptation in changing environments

Conclusion

While we have outlined a path forward for the field of Controlled Living Organisms, we also acknowledge that it is a challenging one, with several fundamentally difficult questions that remain to be addressed.

First and foremost, it is unclear how complex each of the layers has to be. Said another way, what level of complexity is sufficient to achieve reliable and predictive control? The answer to this question is inherently governed by the practicalities of the problem we are trying to solve. The goal of control is to meet an objective (i.e., steer toward a desired behavior) while compensating for uncertainties in the environment. Our intuition is that for some applications—for example involving simple objectives in closed, highly controlled, laboratory environments—relatively "simple control" will be sufficient. However, our ambitions for the field are to be able to implement multi-objective control of living systems in a highly robust manner in complex and open environments (see Figure 3). This means that the level of complexity of what we are trying to design is driven by our articulation and understanding of the objectives and environments. On the latter, examples of staging engineering living systems in open environments (*in situ*) have been limited. Here, we still don't have a strong quantitative understanding of many critical aspects, such as competition for resources in the environment. These types of uncertainties about the environment force us to expand the complexity of control systems. On the former, a unique and exciting, but inherently, challenging aspect of biological systems (relative to traditional engineered systems) is that they come with their own objectives and complex control mechanisms, written and refined over evolutionary time, and of which we still don't have complete understanding. As synthetic

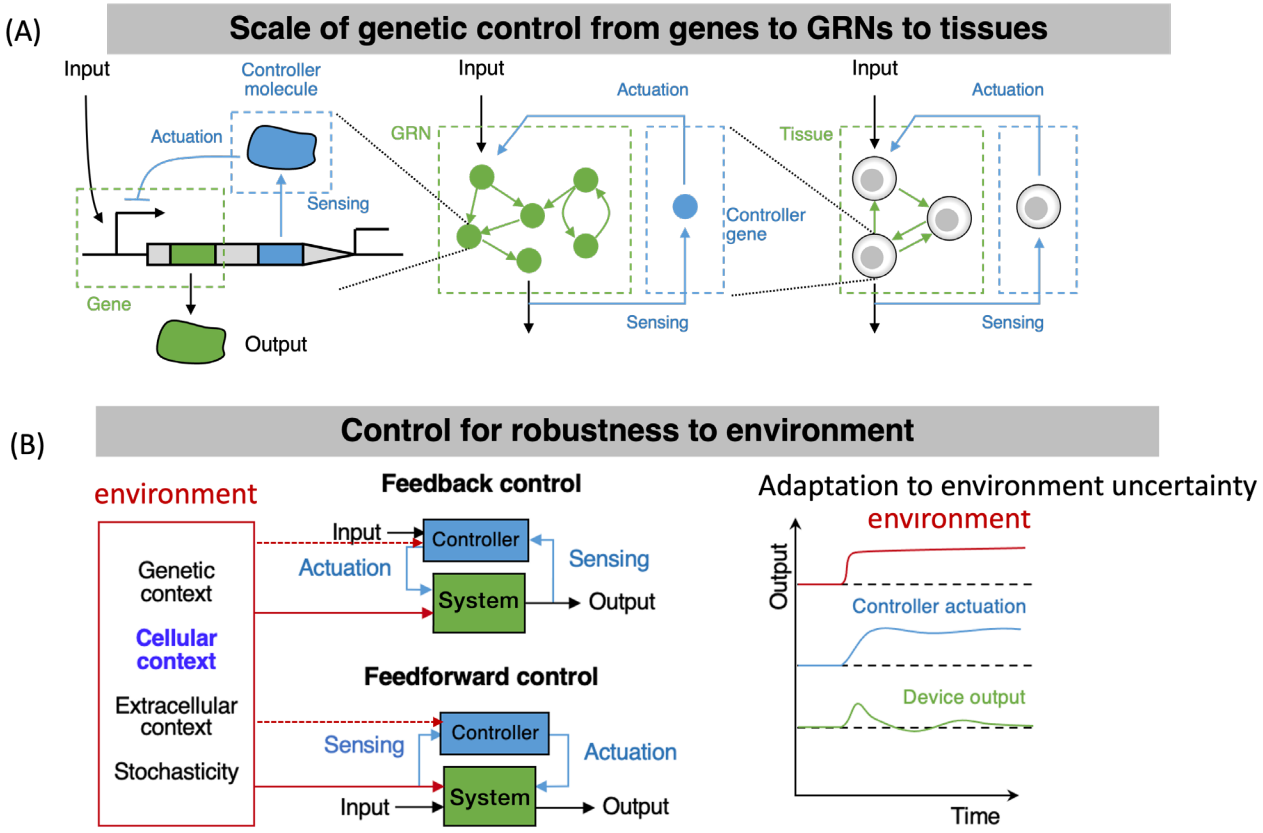


Figure 3 Control at different scales of complexity for robustness to environment. (A) Control of biomolecular systems can occur at different scales of complexity: at the single gene level to control the output protein's level robustly to environmental changes, at the regulatory network level to make a specific node of the network achieve a desired output despite environmental changes, and at the cell population level to control the composition of a specific consortium despite discrepancies on growth rate and environmental changes. (B) The essence of feedback and feedforward control architectures. In feedback control, we measure one or multiple outputs of a system, compare them to what we would like them to be, and accordingly adjust the actuation on the system to compensate for discrepancies. In feedforward control, we typically have a way to indirectly measure some of the environmental perturbations hitting the system such that we can compensate for their action on the system directly on the output of interest. In either (A) or (B), the controller boxes can implement arbitrarily complex control laws, including those that perform learning of the environment and adaptation to it. How to physically implement them through biomolecular processes is still a major challenge, as so far we only reached high-gain negative feedback (a form of proportional controller), derivative control, and integral control. [Adapted from Shakiba, 2021]

biologists, much of our efforts center on developing what are effectively small variations of these complex extant programs. How to design these variations that operate alongside—and when necessary, overwrite—extant programs to achieve user-defined objectives is still unclear and serves as a key driver of the level of complexity needed for design.

The question of complexity and the difficulty of identifying what levels of complexity are sufficient for effective design naturally raises two corollary questions. The first is what will enable us to harness AI/ML methods for effective design? There has been an extraordinary increase in the technological capability of AI/ML methods, which have shown stunning recent success in solving problems in other fields, ranging from enabling mathematical discoveries to computer vision and protein structure prediction. Not surprisingly, a common theme that emerged in our workshop focused on the challenges and opportunities of harnessing AI/ML methods in the field of CLO. These powerful tools, which operate on a substrate of data/information, offer intriguing opportunities to help identify levels of complexity to accomplish a goal. However, this will only be successful when there is structured data on which to make inferences and domain experts to help develop AI algorithms (e.g., AI-assisted biophysics models). We foresee critical challenges in linking together metrology and information representation, as well as model-building to make AI maximally powerful to aid in control for design. We predict that addressing these issues will be a major theme going forward, which will be necessary to understand how to iteratively develop and improve these powerful data-driven frameworks for engineering biology.

The second question is can we design biological systems that find their own solutions? Conceptually, this represents an alternative way of performing self-optimization, wherein instead of implementing an AI/ML algorithm *in silico* to navigate biological search space, here the focus is on creating biological systems that perform their own search and self-optimization *in vivo* based on a set of conditions. If successful, this would mean that it would not be necessary for the user to rationally and directly design a biological system for a desired solution. To realize this vision, we need to consider what are designs for controllable evolution to specified optima and how do we implement such controllers? Exciting technologies are emerging that may provide components to design such systems. These include directed evolution systems that enable continuous hypermutation of genes *in vivo* (i.e., the engine) (Molina, 2022), methods to create circuits that couple the desired biological activity or function to cell viability or other selectable cellular phenotypes (i.e., the guidance system), and even flexible and scalable automation technologies to run these evolution experiments in a self-driving manner (i.e., the plant) (Zhong, 2020; Wong, 2018; Debenedictis, 2022). We anticipate that these emergent technologies, when coupled with theory of control and evolution, will form the basis of an exciting field focused on applying controllable evolution to generate self-optimizing biological systems that drive toward optimum behaviors while, in turn, revealing the critical layers of design complexity for achieving such behaviors.

Glossary

abiotic controllers – non-living factors that affect biological functions (ex: temperature).

allelic variation – the presence or number of different allele forms at a particular location on a chromosome.

Alpha Fold – an AI system that predicts a protein's 3D structure from its amino acid sequence.

Arrhenius curve – results from an equation used to determine the effect of a change of temperature on the rate constant, and consequently on the rate of the reaction.

bacterial chassis – the cellular host used as a recipient of engineered biological systems in synthetic biology.

beamline – a beam of particles (such as photons or electrons) emitted from a particle accelerator.

bioreactors – any manufactured device or system that supports a biologically active environment.

biosynthesis – the production of complex molecules within living organisms or cells.

CAR-T immunotherapies – a cross of immunotherapy, gene therapy, and cellular therapy.

codon usage – regulates the speed of translation elongation, resulting in non-uniform ribosome decoding rates on mRNAs during translation that is adapted to co-translational protein folding.

constitutive and inducible promoters – An *inducible promoter* is a strong regulatory promoter that helps in the effective expression of the desired gene. A *constitutive promoter* carries a continuous transcription process in the desired gene.

CRISPRi – Clustered Regularly Interspaced Short Palindromic Repeats interference is an RNA-based method for highly specific silencing of the transcription in prokaryotic or eukaryotic cells.

dynamical manifolds – the multiscale-multifaceted features of bodily processes at the molecular, cellular, tissue, organ, and systemic levels that comprise the human body.

EcoPods – A commercially available container of an ecologically balanced consortium of copepods.

endogenous – growing or originating from within an organism.

extra-chromosomal replicons – a region of an organism's genome that is independently replicated from a single origin of replication outside of the genome.

heritable epigenetics – epigenetic modification (e.g., DNA Methylation) which is passed down to offspring.

hyperspectral imaging – collecting images from across the electromagnetic spectrum.

in situ – Latin for "on site" or "in position."

in vivo – performed or taking place in a living organism

methylation patterns – Addition of Methyl (CH3) groups to DNA increases stability during transcription processes.

metrology – the scientific study of measurement.

OMICS – any of several areas of biological study defined by the investigation of the entire complement of a specific type of biomolecule or the totality of a molecular process within an organism.

ontology – a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

organismal – of or relating to an organism or organisms.

organ-on-a-chip – systems containing engineered or natural miniature tissues grown inside microfluidic chips.

orthogonal – of or involving right angles; at right angles.

phenotypes – the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment.

pareto – a theory maintaining that 80 percent of the output from a given situation or system is determined by 20 percent of the input.

phospho-relay componentry – The phosphor-relay is a complex variation of a two-component regulatory system. It includes phosphotransferases that transfer the phosphoryl group from the sensor kinases to the ultimate target. The sporulation initiation phosphorelay is the paradigm of this class of signal transduction systems.

phosphorylation – the addition of a phosphoryl (PO3) group to a molecule.

RB-TNSEQ – random bar code transposon-site sequencing.

Rosetta – a software suite includes algorithms for computational modeling and analysis of protein structures.

stochastic – randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely.

unified data fabric – replicates data as needed and without breaking security and governance controls.

Bibliography

Cameron DE, Bashor CJ, Collins JJ. A brief history of synthetic biology. Nat Rev Microbiol. 2014 May;12(5):381-90. doi: 10.1038/nrmicro3239. Epub 2014 Apr 1. PMID: 24686414.

DeBenedictis, E.A., Chory, E.J., Gretton, D.W. et al. Systematic molecular evolution enables robust biomolecule discovery. Nat Methods 19, 55–64 (2022). https://doi.org/10.1038/s41592-021-01348-4

Fatma Z, Schultz JC, Zhao H. Recent advances in domesticating non-model microorganisms. Biotechnol Prog. 2020 Sep;36(5):e3008. doi: 10.1002/btpr.3008. Epub 2020 May 12. PMID: 32329213.

Jumper J, Evans R, Pritzel A, Green T. et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15. PMID: 34265844; PMCID: PMC8371605.

Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nat Rev Genet. 2010 May;11(5):367-79. doi: 10.1038/nrg2775. PMID: 20395970; PMCID: PMC2896386.

Leman JK, Weitzner BD, Lewis SM. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat Methods. 2020 Jul;17(7):665-680. doi: 10.1038/s41592-020-0848-2. Epub 2020 Jun 1. PMID: 32483333; PMCID: PMC7603796.

Mante J, Myers CJ. Advancing reuse of genetic parts: progress and remaining challenges. Nat Commun. 2023 May 23;14(1):2953. doi: 10.1038/s41467-023-38791-0. PMID: 37221178; PMCID: PMC10206149.

Molina, R.S., Rix, G., Mengiste, A.A. et al. *In vivo* hypermutation and continuous evolution. Nat Rev Methods Primers 2, 36 (2022). https://doi.org/10.1038/s43586-022-00119-5

Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009 Oct;27(10):946-50. doi: 10.1038/nbt.1568. Epub 2009 Oct 4. PMID: 19801975; PMCID: PMC2782888.

Shakiba N, Jones RD, Weiss R, Del Vecchio D. Context-aware synthetic biology by controller design: Engineering the mammalian cell. Cell Systems, Volume 12, Issue 6, 2021, Pages 561-592, ISSN 2405-4712, https://doi.org/10.1016/j.cels.2021.05.011.

Wong, B., Mancuso, C., Kiriakov, S. et al. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat Biotechnol 36, 614–623 (2018). https://doi.org/10.1038/nbt.4151

Zhong, Ziwei, Brandon G. Wong, Arjun Ravikumar, Garri A. Arzumanyan, Ahmad S. Khalil, and Chang C. Liu. "Automated Continuous Evolution of Proteins *in Vivo*." ACS Synthetic Biology 9, no. 6 (2020): 1270–76. https://doi.org/10.1021/acssynbio.0c00135.

Appendix 1 – Workshop Attendees

Workshop Co-chairs

Adam Arkin, *University of California, Berkley*
Michael Desai, *Harvard*
Mo Khalil, *Boston University*
Domitilla del Vecchio, *Massachusetts Institute of Technology*

Workshop Participants

Ahmed Badran, *Scripps Research Institute*
Caleb Bashor, *Rice University*
William Bentley, *University of Maryland*
Mary Dunlop, *Boston University*
Elisa Franco, *University of California, Los Angeles*
Katie Galloway, *Massachusetts Institute of Technology*
Jeff Gore, *Massachusetts Institute of Technology*
Neha Kamat, *Northwestern University*
Chang Liu, *University of California, Irvine*
Leonardo Morsut, *University of Southern California*
Grant Murphy, *Merck*
Chris Myers, *University of Colorado*
Michael Springer, *Harvard University*
Jeff Tabor, *Rice University*
Huimin Zhao, *University of Illinois at Urbana-Champaign*

Government Observers

Bindu Nair, *OUSD(R&E) Basic Research Office*
Jean Luc Cambrier, *OUSD(R&E) Basic Research Office*
Ben Wolfson, *OUSD(R&E) Basic Research Office*
Cindy Achat-Mendes, *OUSD(R&E) Basic Research Office*
Jennifer Becker, *OUSD(R&E) Basic Research Office*
Melissa Edwards, *OUSD(R&E) Basic Research Office*
Sithira Ratnayaka, *OUSD(R&E) Basic Research Office*
Ben Epstein, *DARPA*
Aura Gimm, *Department of Energy*
David Rockliffe, *National Science Foundation*

VT-ARC Staff

Kate Klemic, *Division Director*
Kellie Perry, *Program Manager*
Lynne Ostrer, *Rapporteur*
Matthew Bigman, *Rapporteur*
Matthew Peters, *Rapporteur*
Cyndie Ramboyong, *Facilitator*
Jordan Brown, *Event Coordinator*

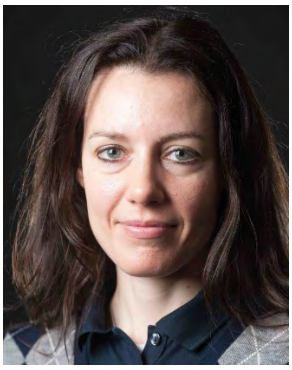
Workshop Participant Short Biography

Adam Arkin, University of California Berkeley
Dean A. Richard Newton Memorial Professor; Senior Faculty Scientist, Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory
aparkin@lbl.gov | https://arkinlab.bio/
Adam Arkin is the Dean A. Richard Newton Memorial Professor in the Department of Bioengineering at the University of California, Berkeley and Senior Faculty Scientist at the Lawrence Berkeley National Laboratory. He and his laboratory specialize in the systems and synthetic biological approaches for discovery, prediction, control and design of microbial and viral functions and behaviors in environmental contexts.

Ahmed Badran, The Scripps Research Institute
Assistant Professor of Chemistry
ahmedhbadran@gmail.com | https://badranlab.com/
Ahmed H. Badran is an Assistant Professor in the Department of Chemistry at The Scripps Research Institute. His works aims to probe and engineer the most fundamental biomolecules and genetic circuits in living cells, and to develop next generation solutions to long-standing global issues in healthcare and climate change. Dr. Badran earned his B.Sc. in Biochemistry & Molecular Biophysics, as well as Molecular & Cellular Biology, from the University of Arizona. Subsequently, he earned his Ph.D. in Chemical Biology from Harvard University under the guidance of Prof. David R. Liu, leading the development and application of rapid methods for continuous directed evolution. He later became Fellow of the Broad Institute of MIT and Harvard, he developed new technologies to reprogram protein translation. Badran has earned several distinctions for his undergraduate and graduate research, including the Arnold and Mabel Beckman Scholarship, the National Science Foundation Graduate Research Fellowship, the Harvard Graduate School of Arts and Sciences Merit Fellowship, and the National Institutes of Health Director's Early Independence Award.

Caleb Bashor, Rice University
Assistant Professor of Bioengineering & BiosSciences
caleb.bashor@rice.edu | http://bashorlab.rice.edu/index_small.html
Caleb Bashor is a synthetic biologist who's research focuses on using synthetic regulatory circuitry to control the behavior of living cells. His group engineers artificial gene and signaling circuits in diverse human cell types, including immune and stem cells, with an eye on developing transformational cell-based therapeutics. In addition to using theory and modeling to guide circuit design, his lab has developed novel techniques that incorporate pooled DNA assembly and next-generation sequencing to build and characterize libraries of circuits in ultra high-throughput. Design-to-function datasets generated by this process can be used to train machine learning models that dramatically accelerate the engineering of synthetic biological systems.

William Bentley, Universtiy of Maryland
Director, Robert E. Fischell Institute for Biomedical Devices
bentley@umd.edu | https://bentley.umd.edu/
William E. Bentley is the Robert E. Fischell Distinguished Chair of Engineering and the Inaugural Director of the Robert E. Fischell Institute for Biomedical Devices. He is also appointed to the Department of Chemical and Biomolecular Engineering at the University of Maryland, College Park and the Institute for Bioscience and Biotechnology Research. At Maryland since 1989, Dr. Bentley has focused his research on the development of molecular tools that facilitate the expression of biologically active proteins, having authored over 300 related archival publications.



Domitilla Del Vecchio, MIT

Associate Professor of Mechanical Engineering; Member, Synthetic Biology Center

ddv@mit.edu | https://scripts.mit.edu/~ddv/index.php

Domitilla Del Vecchio received the Ph.D. degree in Control and Dynamical Systems from the California Institute of Technology, Pasadena, and the Laurea degree in Electrical Engineering (Automation) from the University of Rome at Tor Vergata in 2005 and 1999, respectively. From 2006 to 2010, she was an Assistant Professor in the Department of Electrical Engineering and Computer Science and in the Center for Computational Medicine and Bioinformatics at the University of Michigan, Ann Arbor. In 2010, she joined the Department of Mechanical Engineering at the Massachusetts Institute of Technology (MIT), where she is currently Professor and member of the Synthetic Biology Center. She is a Fellow of the International Federation of Automatic Control (2022), an IEEE Fellow (2021), and a recipient of the Newton Award for Transformative Ideas during the Covid-19 Pandemic (2020), the 2016 Bose Research Award (MIT), the Donald P. Eckman Award from the American Control Council (2010), the NSF Career Award (2007), the American Control Conference Best Student Paper Award (2004), and the Bank of Italy Fellowship (2000). Her research focuses on developing techniques to make synthetic genetic circuits robust to context and on applying these to biosensing and cell fate control for regenerative medicine applications.

Michael Desai, Harvard

Fisher Professr of Natural History; Professor of Organismic and Evolutionary Biology

mdesai@oeb.harvard.edu | https://www.desai-lab.com/

Natural selection and other evolutionary forces leave characteristic signatures in the genetic variation within populations. My group uses a combination of theoretical and experimental approaches to study how this genetic variation is created and maintained, and to develop methods to infer the evolutionary history of populations from the variation observed in sequence data. Our focus is primarily on natural selection in asexual populations such as microbes and viruses. We are developing new approaches to population genetic theory to better understand the structure of genetic variation in these populations. We complement this with high-throughput experimental evolution in budding yeast, evolving thousands of lines simultaneously to explore the distributions of phenotypic changes and their correlations with the evolution of genetic variation within and between populations.

Mary Dunlop, Boston University

Associate Professor of Biomedical Engineering

mjdunlop@gmail.com | http://www.dunloplab.com/

Mary Dunlop is an Associate Professor of Biomedical Engineering at Boston University with additional appointments in Bioinformatics and in the Molecular Biology, Cell Biology & Biochemistry program. She graduated from Princeton University with a B.S.E. in Mechanical and Aerospace Engineering and a minor in Computer Science. She then received her Ph.D. from the California Institute of Technology, where she studied synthetic biology with a focus on dynamics and feedback in gene regulation. As a postdoctoral scholar, she conducted research on biofuel production at the Department of Energy's Joint BioEnergy Institute. Her lab engineers novel synthetic feedback control systems and also studies naturally occurring examples of feedback in gene regulation. In recognition of her outstanding research and service contributions, she has received many honors including a DOE Early Career Award, an NSF CAREER Award, the ACS Synthetic Biology Young Investigator Award, and an NSF Transitions Award. She is also the recipient of several teaching awards, including Boston University's Biomedical Engineering Professor of the Year Award (2019) and the College of Engineering Teaching Excellence Award (2020).

Elisa Franco, UCLA

Associate Professor of Mechanical and Aerospace Engineering and Bioengineering

efranco@seas.ucla.edu | http://www.seas.ucla.edu/francolab/people.html

I am a Professor of Mechanical and Aerospace Engineering and Bioengineering at UCLA, where I lead a research group dedicated to advancing the fields of synthetic biology, feedback control, and nucleic acids nanotechnology. My academic background includes a Master's and Ph.D. from the University of Trieste, Italy, as well as a Ph.D. from Caltech, where I specialized in automation and dynamical systems theory with a focus on biochemistry and nanotechnology.

Katie Galloway, MIT

W. M. Keck Career Development Professor in Biomedical Engineering

katiegal@mit.edu | https://gallowaylab.mit.edu/

Katie Galloway is the W. M. Keck Career Development Professor in Biomedical Engineering and Chemical Engineering at Massachusetts Institute of Technology (MIT). Her research focuses on elucidating the fundamental principles of integrating synthetic circuitry to drive cellular behaviors. Her lab focuses on developing integrated gene circuits and elucidating the systems-level principles that govern complex cellular behaviors. Her team leverages synthetic biology to transform how we understand cellular transitions and engineer cellular therapies. Galloway earned a PhD and an MS in Chemical Engineering from the California Institute of Technology (Caltech), and a BS in Chemical Engineering from University of California at Berkeley. She completed her postdoctoral work at the University of Southern California. Her research has been featured in Science, Cell Stem Cell, Cell Systems, and Development. She has won multiple fellowships and awards including the NIH Maximizing Investigators' Research Award (MIRA) R35, the NIH F32, and Caltech's Everhart Award.

Jeff Gore, MIT

Associate Professor of Physics

gore@mit.edu | http://www.gorelab.org/

Jeff joined the MIT Physics Department as an Assistant Professor in January 2010 after spending the previous three years in the Department as a Pappalardo Fellow working with Alexander van Oudenaarden. With the support of a Hertz Graduate Fellowship, in 2005 he received his PhD from the Physics Department at the University of California, Berkeley. His graduate research in single-molecule biophysics was done in the laboratory of Carlos Bustamante, focusing on the study of twist and torque in single molecules of DNA. Jeff is excited to be in the Physics Department here at MIT, particularly since this is where he studied as an undergraduate in the late '90s.

Neha Kamat, Northwestern University

Assistant Professor of Biomedical Engineering

nkamat@northwestern.edu | https://www.nehakamat.com/

Neha is currently an Assistant Professor in the Biomedical Engineering Department at Northwestern University. The Kamat Lab's interests lie in constructing minimal systems, or artificial cells, as a tool to understand and recreate certain cellular behaviors. They use emerging engineering methods in material science and synthetic biology to construct in vitro models of cellular membranes to be used for fundamental studies on membrane mechanobiology and for the design of new therapeutic tools.

Ahmad Khalil, Boston University

Associate Professor of Biomedical Engineering

khalil@bu.edu | https://www.bu.edu/khalillab/

Ahmad (Mo) Khalil is Professor of Biomedical Engineering, Dorf-Ebner Distinguished Faculty Fellow, and Founding Associate Director of the Biological Design Center at Boston University. He is also a Visiting Scholar at the Wyss Institute for Biologically Inspired Engineering at Harvard University, and Co-Director of a NIH/NIGMS T32 PhD Training Program in synthetic biology. His laboratory develops synthetic biology tools to investigate fundamental principles of cellular regulatory networks, and in turn uses these insights to program therapeutically-useful cellular functions for next-generation gene- and cell-based therapies. His team also develops novel continuous evolution technologies that are automated and scalable, which they are using to generate biomolecules with radically altered or new functions to address unmet needs in biology, medicine, and biotechnology. He is recipient of numerous awards, including a Schmidt Science Polymath Award, Presidential Early Career Award for Scientists and Engineers (PECASE), DoD Vannevar Bush Faculty Fellowship, W.M. Keck Medical Research Award, NIH New Innovator Award, NSF CAREER Award, DARPA Young Faculty Award, and Hartwell Foundation Biomedical Research Award, and he has received numerous awards for teaching excellence at both the Department and College levels.

Chang Liu, University of California Irvine

Professor and Chancellor's Fellow of Biomedical Engineering, Chemistry, and Molecular Biology & Biochemistry; Director, Center for Synthetic Biology

ccl@uci.edu | https://liulab.com/ccl/

Chang Liu is Professor and Chancellor's Fellow of Biomedical Engineering, Chemistry, and Molecular Biology & Biochemistry and the Director of the Center for Synthetic Biology at UC Irvine. Liu graduated from Harvard in 2005 with a bachelor's degree in chemistry and carried out his PhD at the Scripps Research Institute. His PhD work, done in the laboratory of Peter Schultz, focused on expanding bacterial genetic codes for the co-translational incorporation of post-translational modifications and using expanded genetic codes in the evolution of novel protein function. From 2009-2012, Liu was a Miller Fellow at UC Berkeley where he worked with Adam Arkin on the predictable design of complex regulatory systems using the special properties of RNA switches. In 2013, Liu started his lab at UC Irvine. Liu's research is in the fields of synthetic biology, protein engineering, chemical biology, and directed evolution. His group engineers orthogonal genetic systems that continuously and rapidly mutate chosen genes *in vivo*. These systems surpass the mutational speed limits of host genomes to enable gene evolution at unprecedented speed, scale, and depth in order to engineer new protein functions, probe the rules of evolution, and understand the fundamental sequence-function relationships governing proteins and other macromolecules. These systems also allow researchers to record transient information as heritable mutations in order to track animal and cancer development at high cellular resolution.

Leonardo Morsut, University of Southern California

Assistant Professor

morsut@usc.edu | https://morsutlab.usc.edu/

Leonardo Morsut is an assistant professor of stem cell biology and regenerative medicine at the Keck School of Medicine of USC. He holds a joint appointment with the Biomedical Engineering department at the USC Viterbi School of Engineering, where he is also director of a new Center for Integrated Electronics and Biological Organisms (CIEBOrg). Morsut was born and raised in Italy with an early fascination with science in general and embryology in particular; he acquired a broad spectrum of degrees in medical biotechnologies, mathematics, and developmental biology. His doctorate work at the University of Padova (Italy) involved the discovery of the role of YAP/TAZ in mechanotransduction and the signaling of early mouse embryogenesis. In his postdoctoral work in the Wendell Lim's at the UC - San Francisco, he developed a new class of synthetic receptors called Synthetic Notch that have been successfully licensed for cancer cell therapies. Morsut started his lab in 2017 at USC with the mission of developing the discipline of synthetic developmental biology, i.e., the design and implementation of genetic circuits to control developmental transitions. Leonardo is also a yogi, father, and husband. Between 1999 and 2006, Morsut was a professional volleyball player in Italy.

Grant Murphy, Merck

Executive Director, Discovery Biologics

grantm@princeton.edu | https://www.linkedin.com/in/grant-murphy-syn-bio/

Prior to Merck, Grant was a NIH Kirschstein fellow at Princeton University with Professor Michael Hecht. As a postdoc, Grant developed computational methods to design de novo protein libraries and developed screens and selections to identify de novo enzymes. This work led to the creation of a completely de novo enzyme. Grant earned his Ph.D. at the University of North Carolina working with Professor Brian Kuhlman. As a graduate student, Grant wrote flexible backbone design algorithms as part of the Rosetta protein design software. Using these methods, Grant designed and experimentally confirmed some of the first computationally designed proteins.

Chris Myers, University of Colorado Boulder

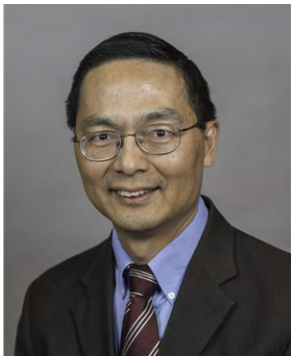
Professor and Palmer Leadership Chair in Electrical, Computer & Energy Engineering

chris.myers@colorado.edu | https://geneticlogiclab.org/#about

Chris J. Myers is Professor and Chair of Electrical, Computer, and Energy Engineering at the University of Colorado Boulder. Before arriving at CU Boulder, he was a professor and associate chair in the Department of Electrical and Computer Engineering at the University of Utah in Salt Lake City. Myers is the author of over 200 technical papers and the textbooks Asynchronous Circuit Design and Engineering Genetic Circuits. He is also a co-inventor on four patents. His research interests include asynchronous circuit design, formal verification of analog/mixed signal circuits and cyber-physical systems, and modeling, analysis and design of genetic circuits. He is a fellow of the IEEE and a member of the editorial boards for Engineering Biology and Synthetic Biology. He is also a leader in the development of standards for systems and synthetic biology. In particular, he has served as an editor for the Systems Biology Markup Language (SBML) standard, is the chair of the steering committee for the Synthetic Biology Open Language (SBOL) standard and is the past chair of the coordination board for the Computational Modeling and Biology Network (COMBINE).

Michael Springer, Harvard Medical School
Associate Professor of Systems Biology
michael_springer@hms.harvard.edu | https://springerlab.org/
Dr. Michael Springer is a professor at Harvard Medical School and co-director of the HIVE (a synthetic biology institute at HMS). Dr. Springer is also an associate member of the Broad Institute and Wyss Institute. The Springer lab is currently focused on developing cheap scalable diagnostic and engineering microbes for sustainability.

Jeffrey Tabor, Rice University
Professor of Bioengineering and BioSciences
jeff.tabor@rice.edu | http://www.taborlab.rice.edu/
Jeff Tabor is a Professor of Bioengineering at Rice University. He earned his Ph.D. in molecular biology under Professor Andy Ellington at the University of Texas at Austin in 2006. There, he led the team that invented bacterial photography and engineered a set of differentially translated mRNAs to demonstrate that ribosome competition increases gene expression noise. Jeff went on to train as an NIH postdoctoral fellow with Professor Christopher Voigt at UCSF from 2006-2010. As a postdoc, he engineered bacteria to solve the challenging image processing problem of edge detection and developed the first system for multiplexed optogenetic control of gene expression. He started his independent research group at Rice in 2010. His group focuses on discovering and repurposing bacterial two-component sensors for applications in optogenetics, diagnostic and therapeutic bacteria, environmental sensing, and other areas. He has received the ONR Young Investigator, NSF CAREER awards and is a fellow of the American Institute for Medical and Biological Engineering.



Huimin Zhao, University of Illinois at Urbana-Champaign
Professor of Chemistry, Biochemistry, Biophysics, and Bioengineering
zhao5@illinois.edu
Dr. Huimin Zhao is the Steven L. Miller Chair of chemical and biomolecular engineering at the University of Illinois at Urbana-Champaign (UIUC), director of NSF AI Institute for Molecule Synthesis (moleculemaker.org), and Editor in Chief of ACS Synthetic Biology. He received his B.S. degree in Biology from the University of Science and Technology of China in 1992 and his Ph.D. degree in Chemistry from the California Institute of Technology in 1998 under the guidance of Nobel Laureate Dr. Frances Arnold. Dr. Zhao has authored and co-authored over 410 research articles and over 30 issued and pending patent applications. In addition, he has given over 460 plenary, keynote, or invited lectures. Dr. Zhao received numerous research and teaching awards and honors. His primary research interests are in the development and applications of synthetic biology, machine learning, and laboratory automation tools to address societyâ€™s most daunting challenges in health, energy, and sustainability.

Appendix 2—Workshop Agenda and Prospectus

Basic Research Innovation Collaboration Center
4100 N. Fairfax Rd. | Fourth Floor| Suite 450
Arlington, VA 22203

DAY 1—TUESDAY, MAY 23, 2023

Time	Title	Speaker
8:00—8:15	Check-in and Continental Breakfast	
8:15 - 8:20	Welcome and Introductions and Expectations	Mo Khalil, Boston University
8:20 -8:45	Workshop Framing Talk	Co-chairs
8:45—9:00	Breakout Instructions and Morning Break	
9:00—10:45	Working Group I: Define the Problem *Small group discussions to frame a vision for controlled living organisms research and identify the greatest hurdles to achieving it.* Group A—Description of Systems Group B—Control of Systems Group C—Design of Systems	
10:45—11:00	BREAK - *Transition to main conference room and leads prepare outbriefing*	
11:00 –12:00	Working Group 1: Outbriefing	
12:00—1:00	LUNCH (provided for participants)	

DAY 1—TUESDAY, MAY 23, 2023

Time	Title	Speaker
1:00—3:45	Working Group II: Technical Capabilities and Opportunities *What are the promising research directions? What are the potential capabilities in the 10- to 20-year horizon?* Group A—Description of Systems Group B—Control of Systems Group C—Design of Systems	
3:45—4:00	BREAK - *Transition to main room and leads prepare outbriefing*	
4:00—4:45	Working Group II: Outbriefing	
4:45—5:00	Summary of Day	Co-chairs
5:00	MEETING ADJOURNED FOR THE DAY	

DAY 2—WEDNESDAY, MAY 24, 2023

Time	Title	Speaker
8:00—8:15	Check-in and Continental Breakfast	
8:15—8:30	Welcome and Day 1 Recap	Co-chairs
8:30—9:30	**What's Missing?** Discussion of topics which did not fit into the framework of day 1 but need to be discussed.	
9:30—10:00	**Big Question 1:** What will enable us to harness AI/ML for effective Design?	
10:00—10:15	BREAK	
10:15—10:45	**Big Question 2:** Complexity - what is sufficient? - to understand/describe - to control/design	
10:45—11:15	**Big Question 3:** Can we design systems that can find their own solutions/ find their own networks, without designing the solution directly?	
11:15—11:45	Discussion of Key Ideas / Components for Report	
11:45—12:00	Closing Remarks Co-chairs	
12:00	DEPARTURE	

Future Directions Workshop: Controlled Living Organisms

Basic Research Office, Office of the Under Secretary of Defense (R&E)

23-24 May 2023
Basic Research Innovative Collaboration Center
4100 N. Fairfax Road, Suite 450 Arlington, VA 22203

Co-chairs: Ahmad S. Khalil (Boston University), Domitilla del Vecchio (MIT), Adam Arkin (Lawrence Berkeley National Laboratory), Michael Desai (Harvard University)

Recent years have seen enormous growth in the field of synthetic biology, from the creation of basic cellular circuits utilizing transcriptional, post-transcriptional and translational mechanisms to the identification of minimal viable genomes leading towards synthetic cells. Significant work has been accomplished utilizing simplified signaling circuits within single cell systems. This research may inform the eventual creation of user-controlled, programmable, multifunctional single (or multi) cell-based systems that could be applied across a wide range of future capabilities including, but not limited to, wound healing, manufacture of fully functional and complex organs, smart and adaptive materials, and sensing. These advances may also lead to the design of new multi-cellular organisms capable of operating in diverse environments. The *Future Directions Workshop on Controlled Living Organisms* will gather researchers to examine the prospects for applying new approaches, theories, and tools in basic research to enable these capabilities.

There are still many open and fundamental scientific questions in understanding the complexity as one progresses from intra-cellular mechanisms to inter-cellular and systems scales. The current frontiers of prediction, control, and design of biological systems span, for example, the ambition to engineer microbes to use diverse sustainable feedstocks to produce high value products; to program the development and assembly of mammalian cells into functional tissues outside of the body that serve as valid models of disease or transplantable function; to intervene and program microbial communities *in situ*; to improve nutrient mobilization in plants while sequestering carbon in the soil in a fashion resilient to the changeable, open environments around them. These strike at critical national needs to accelerate the circular economy, solve critical problems in human health, and address the scalable biological components of climate change in a planet with diminishing resources. The physical inputs and scales to these biological dynamics problems vary greatly with the system and question being addressed. It remains an open challenge how to best represent a biological system and its relevant environment such that appropriate prediction, control, and design problems can be solved through model-driven experimental design and cellular engineering. Modeling and creating networks that effectively incorporate these intra- and extra-cellular dynamics, within and between multiple nodes, appears essential for the creation and utilization of controllable and complex cellular circuits. This will require new families of tools, as well as new and foundational dynamical and control theories that account for chemical and physical interactions, spatial dynamics, timescale separation, uncertainty, and other factors unique to biological networks, up to and including open systems and the likelihood of emergent behavior.

The *Future Directions Workshop on Controlled Living Organisms* workshop will gather researchers from a variety of fields, including synthetic biology, molecular biology/biochemistry, control theory, systems biology, physics, mathematics, computer science, and bioethics to work across three amplifying layers, representing the three technical areas of this workshop:

1. **Description of Systems:** We must identify the critical features of the biological system to be modeled and the representation of their interaction and activity. From a physical perspective this would mean tracing the causal chemical and physical interaction of the cellular chemistry from the regulated transcription and translation of individual genomes, through metabolic and signaling systems, through cell growth and cell-cell interactions, to the active and passive dispersal of these cells and their aggregates up to organism level and beyond. While this accounting would be complete, it is difficult to achieve, and computation would be hard to scale. Thus other levels of abstraction, physical and otherwise, need to be deployed and used together in reasonable ways. Here reasonable means that it is possible to experimentally observe and estimate the critical elements and the parameters of their interaction. How do we create multi-model representations of multiscale biological systems that can be effectively 'parameterized' by well-designed experiments to enable predictive power needed for control and design?

2. **Control of Systems:** Based on the foundation of Layer one, we can make forward predictions of a system's' behavior based on its current state and environmental inputs. Control theory allows us to determine what inputs we can make to the environment to move the state of the system to a desired outcome. Because of the high uncertainty, nonlinearity, and noise in most models of biological systems, there are new challenges in developing effective theories of control and building real control systems (cell external or even within the cell itself) that can observe the right variables and produce the right inputs at the appropriate time/space scales to achieve the outcome.

3. **Design of systems:** Once we have a theory of control based on validated models of the target system, we can begin to design systems within and across cell (and organismal) populations to autonomously achieve outcomes through design of their endogenous systems and the environments in which they operate. In some ways, this is a natural extension of control. However, this opens a new fundamental avenue which is how to create a reliable engineering 'infrastructure' and 'supply-chain'. We need to be able to design novel function (e.g. new regulators or metabolic activities) based on the needs of the problem and we need to be able to add these elements into the biological systems and account for their 'loads' and off-target effects, as well as their defined function. This leads to new challenges in the design of biological systems that are currently very different in other disciplines.

While these three layers are intertwined, each represents a separable extension of the preceding layer.

As work in these fields progresses, ethical ramifications become increasingly relevant. While not the primary focus of this workshop, it is essential that bioethics be considered and discussed within all technical areas. Future research in these realms must be conducted within appropriate ethical boundaries in multiple domains, which should be identified. These include biosafety, ecology, sociology, and others.

Participants will discuss opportunities and challenges in these fields, primarily in small-group breakout sessions and whole-group discussions. The workshop aims to focus discussion around three overarching questions:

1. How will this research impact future science and technology capabilities in the future?
2. What is the trajectory of scientific research in this area over the next 10-20 years?
3. What are the primary challenges to progress, and how can they be addressed?

A key outcome of this Workshop will be a roadmap of key basic science research needs that, if addressed in the next 10-20 years, can substantially advance this transformational vision. The discussions and ensuing distributed report will provide valuable long-term guidance to the DoD community, as well as the broader federal funding community, federal labs, and other stakeholders. Workshop attendees will emerge with a better ability to identify and seize potential opportunities in the different fields addressed. This workshop is sponsored by the Basic Research Office within the Office of Secretary of Defense, along with input and interest from the Services and other DoD components.

Agenda

Rather than a standard conference format, the workshop design emphasizes interactive dialogue with primarily small-group breakout sessions followed by whole-group synthesis of ideas.

Day One: The majority of the first day will be spent in small-group breakout sessions on fundamental challenges to progress and research opportunities for the three technical areas described above.

Day Two: The second day of the workshop is a half-day consisting of white-space, whole group discussions on topics that did not fall into the Day 1 framework or were especially ambitious and/or high-risk. Participants will also discuss cross cutting themes and the trajectory of the field over the next 10-20 years. At the end of the day, the whole group will discuss the overarching themes of the workshop that should be included in the final workshop report.