Subsystem Integration for Design and Implementation of Synthetic Cells

Richard M. Murray
California Institute of Technology

DRAFT, February 15, 2021— Please do not distribute

Project Summary

Advances in synthetic biology and molecular sciences have substantially advanced our ability to produce genetically-programmed synthetic cells from molecular components. These efforts provide techniques for the bottom-up construction of cell-like systems that can provide scientists with new insights into how natural cells work and harness the power of biology to create nanoscale, biomolecular machines. Work in the US through the Build-A-Cell consortium and similar efforts in other countries have established communities of researchers interested in pursuing the construction of synthetic cells, and these activities are an exciting pathway for exploration of the rules of life.

The long term goal of this research is to create genetically-programmed synthetic cells consisting of multiple subsystems operating in an integrated fashion. Unlike more traditional synthetic biology approaches, synthetic cells are non-living: they make use of genetic elements provided by biology, but they do not replicate, mutate, or evolve. Applications range from synthesis of bio-compatible materials, to environmental monitoring and remediation, to self-assembly of complex multi-cellular machines. Pursuing this goal requires fundamental research in biological engineering, aimed at moving from creation of clever biomolecular devices to systematic specification, design, integration, and testing of circuits, subsystems, cells, and multi-component systems.

Intellectual Merit: The specific emphasis in this proposal is on establishing the design methodology, subsystem technologies, and system architectures that can enable the development of genetically-programmed, synthetic cells via subsystem integration. The proposed effort emphasizes demonstration of fundamental principles that can enable construction of complex, genetically-programmed systems, leveraging work by other groups on various components and subsystems. Specifically, while some work on individual components or subsystems will be required, the emphasis is on *integrating* subsystems produced by multiple research groups into functioning synthetic cells.

The technical approach for this project includes a combination of theory, computation, and experiments, aimed at developing a scalable and modular framework for creation of biomolecular circuits and systems that implement complex and robust behaviors. Central elements of the approach include the use of control theory as a unifying mathematical basis, the use of feedback as a core mechanism for managing uncertainty and design of dynamics, and the use of cell-free methods for prototyping and implementation. Two synthetic cell examples—distributed event detection and flagellar-controlled locomotion—will be used to demonstrate the feasibility of synthetic cells; validate the proposed design framework, subsystem implementations; and assembly techniques, and identify areas of future research.

Broader Impact: The technical work in this proposal will be carried out in the context of two ongoing national efforts on development of synthetic cells (Build-A-Cell in the US and fabriCELL in the UK). Existing and expanded collaborations with researchers who are part of those projects will be used to establish mechanisms for dissemination of the results as well as a source of collaborations to expand the set of components and subsystems that will form the basis of our integration efforts. This project will also provide substantial opportunities for undergraduates to participate in research activities. Increased participation by underrepresented groups will be sought through targeted programs available at Caltech, as well as building on previous success in recruiting women and underrepresented minorities into graduate and postdoctoral research activities.

1 Motivation and Background

Synthetic biology has made significant strides over the past 20 years in demonstrating the ability to engineer biological systems by "programming" DNA to carry out specific operations both in cells [30, 39] and in cell-free systems [25, 61]. Currently demonstrated systems have been of modest complexity (typically less than a dozen programmed elements) and have focused on relatively simple operations (oscillators, logic operations, metabolic pathways). A major challenge in the field is learning how to systematically design and implement biomolecular circuits of much higher complexity (hundreds to thousands of programmed elements) that can carry out more complex operations, spread across multiple functional units (*a la* multi-cellular organisms).

One approach to moving the field forward is to shift the focus from engineering of living organisms to the creation of genetically-programmed *synthetic* cells. Unlike more traditional synthetic biology approaches, synthetic cells are non-living: they make use of genetic elements provided by biology, but they do not replicate, mutate, or evolve. Cell-free systems offer many intrinsic advantages, including portability (e.g., paper-based cell-free circuits [47]), safety (via true orthogonality and lack of self replication), and stability (due to lack of mutation and evolution). In this proposal we define a set of fundamental research challenges related to the design of synthetic cells that present a high-risk, high-reward approach to synthetic biology and its applications.

A schematic diagram of the type of system that we envision is shown in Fig. 1. A synthetic cell consists of a number of subsystems, described in more detail below, that provide the core functions required for operation. We propose to carry out the fundamental research required to demonstrate that systems of this complexity can be designed and implemented.

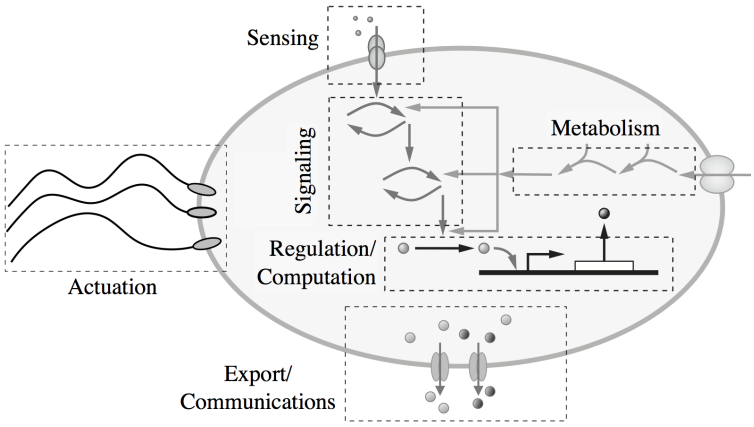


Fig. 1: Conceptual diagrams of a synthetic cell (adapted from Del Vecchio and Murray [10, Fig. 1.5])

Related Work The work that we propose builds on a set of tools for cell-free synthetic biology that my group and others have developed over the last 5–10 years. These efforts involve the development of systematic frameworks for implementing circuits and pathways [15, 18, 20, 23, 40, 42, 58, 70, 74], methods for interconnecting components and isolating unwanted interactions [14, 20], methods for compartmentalizing circuit operations [1, 7, 46], and methods for spatially localizing molecules using programmable scaffolds [27, 69].

The concept of synthetic cells is one that many groups around the world are pursuing. As evidence of the interest in this area, multiple national consortia and interest groups have been organized with the goal of building synthetic cells, including Build-A-Cell in the US [6], fabriCELL in the UK [13], and the European Synthetic Cell Initiative to name a few. What differentiates our activities in this proposal is the focus on subsystem definition and integration as a required element for building biomolecular machines with complexity approaching that of living cells.

Project Goals and Objectives We propose to develop and demonstrate the key design tools, molecular components, and system-level architecture for synthetic cells. While we do not anticipate that a single project will lead to a functioning synthetic cell in a 3–5 year time, we believe that it will be possible implement and integrate a variety of subsystems that establish the feasibility of synthetic cells and serve as a starting point for a larger community effort.

Toward that end, the specific objectives for this project are to:

1. Develop a mathematically rigorous, design-oriented, computational framework enabling modeling, analysis, and design of multi-subsystem synthetic cells, incorporating core biomolecular processes, component-level dynamics, subsystem behaviors, and multi-subsystem interactions.

2. Implement and characterize a collection of at least three biomolecular subsystems capable of providing robust and modular operations in spatially-isolated, cell-free environments.

3. Create and optimize a method of assembling multiple biomolecular subsystems into a multi-element ensemble within an encapsulated environment, with structured interactions between the individual subsystems.

4. Demonstrate feasibility of synthetic cells by implementing at least two experimental demonstrations consisting of interacting, multi-subsystem behaviors that validate the proposed design and implementation frameworks and help identify future of areas research.

5. Disseminate the results of the project via community-driven, open-source software and wetware repositories, as well as collaborations with US and international researchers in synthetic biology.

2 Technical Approach

By leveraging work in the synthetic biology and molecular programming communities over the past decade, we are plausibly within 10–15 years of being able to produce genetically-programmed synthetic cells and multi-cellular machines that can carry out useful engineering operations. Pursuing this vision will require new approaches to biomolecular systems engineering, focused on moving from creation and characterization of devices and simple circuits to systematic specification, design, and integration of circuits, subsystems, cells, and multi-component systems.

A high-level diagram of the type of system we have in mind for a single cell is shown in Fig. 1. The architecture consists of a collection of subsystems, most of which we believe can be implemented using available technology. At this stage, it is not clear whether the synthetic cell should be "booted up" using cell lysate with genetically-programmed DNA sequences or whether subsystems should be created separately and integrated in fully expressed form. In either case, we imagine that the synthetic cells would not replicate themselves per se, though some self-assembly would be required. They would only function as long as an external (probably chemical) energy source is present.

2.1 Cell-Free Synthetic Biology

Prior work in cell-free synthetic biology establishes a basis for technologies and methodologies that will be utilized in this project. In this section we provide a brief introduction to those tools that we have developed under prior funding, as examples of some of the existing results that are available (as well as to establish some credibility and highlight our prior work).

Genelet circuits "Genelet" circuits are DNA- and RNA-based systems that rely only on transcription and binding of complementary sequences of DNA and RNA to create genetically-programmed functions [31, 34, 35]. The primary mechanism of action in a genelet circuit is the use of a partially double-stranded sequence of DNA with an incomplete promoter region that can be completed by using a single-stranded DNA activator. Upon activation, RNA polymerase transcribes downstream region into RNA. Activators can be displaced from the template if they present an overhang, or toehold: this exposed area allows an inhibitor strand (RNA or DNA) to initiate binding and eventually strip off the activator from the template to reach a more favorable thermodynamic state. In addition to this basic mechanism, sequestration and degradation reactions can also be utilized.

Fig. 2 shows an example of a oscillator designed using genelets [14, 35]. Other circuits that have been constructed include a "rate regulator" circuit in which the production two RNA species are modulated to maintain a constant ratio of production rates [15] and an "insulator" that is used to minimize the coupling between an RNA-based oscillator and a set of DNA-based "tweezers" that opens and closes based on the presence of a complementary RNA-strand [14]. Genelets provide a highly programmable approach to implementing transcription, repression, sequestration, and degradation reactions. They can operate in some transcription/translation systems (e.g., the NEB PURExpress system) and we are exploring their use in cell-free extracts.

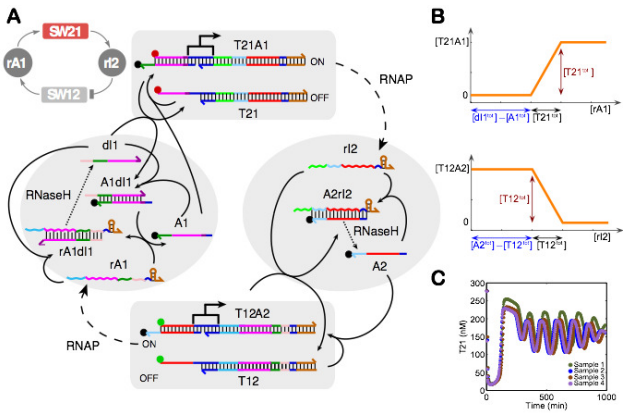


Fig. 2: RNA-based "genelet" oscillator [14, 35].

Cell-free prototyping From roughly 2010–2015, we helped improve on a transcription-translation (TX-TL) platform—originally developed by Vincent Noireaux [56, 62]—to construct and characterize synthetic gene circuits in a cell-free environment. The TX-TL system uses *E. coli* extract containing the cell's protein synthesis machinery and can express genetically-encoded circuits by simply adding DNA encoding the desired circuits into a test tube. Importantly, TX-TL can use linear DNA from a PCR machine, enabling rapid and inexpensive prototyping.

An overview of the current biomolecular breadboard technology is shown in Fig. 3. Column A represents examples of circuits that we have designed and implemented, including a biomolecular "event detector" [27] and a pathway for producing 1,4 butanediol [70]. The main component of the breadboard is the TX-TL extract system, represented in Column B. We typically run 10 μl reactions and are able to obtain 6–10 hours of gene expression and circuit/pathway operation. A detailed modeling toolbox is also available for simulation and analysis of circuits and pathways [66]. More sophisticated test environments are shown in columns C and D, which illustrate the use of a commercially-available droplet-based microfluidic system and a continuous flow reactor (developed by EPFL [41, 42]). We are able to transform circuits into *E. coli* (column E) and are currently developing extracts for other micro-organisms.

The cell-free toolbox is a significant advance over commercially available systems, both because of cost (10-100X cheaper) and because if provides a more realistic prototyping environment. We

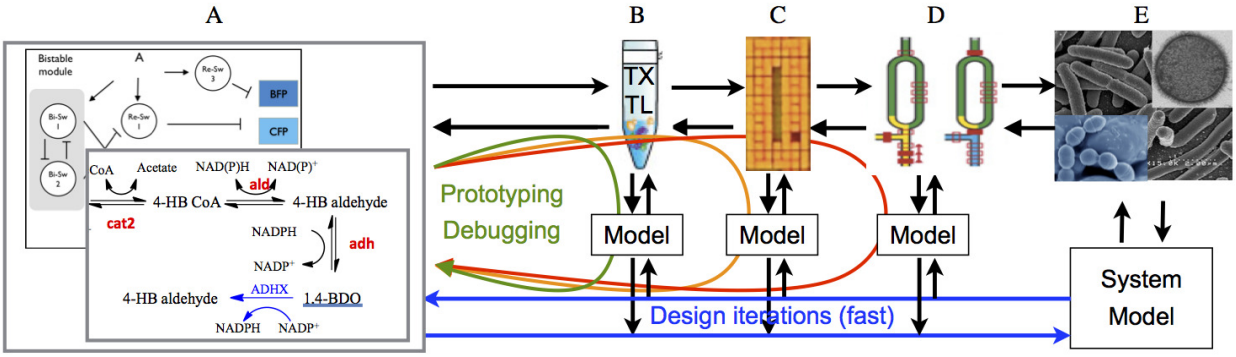


Fig. 3: Biomolecular breadboards framework for rapid prototyping of biological circuits.

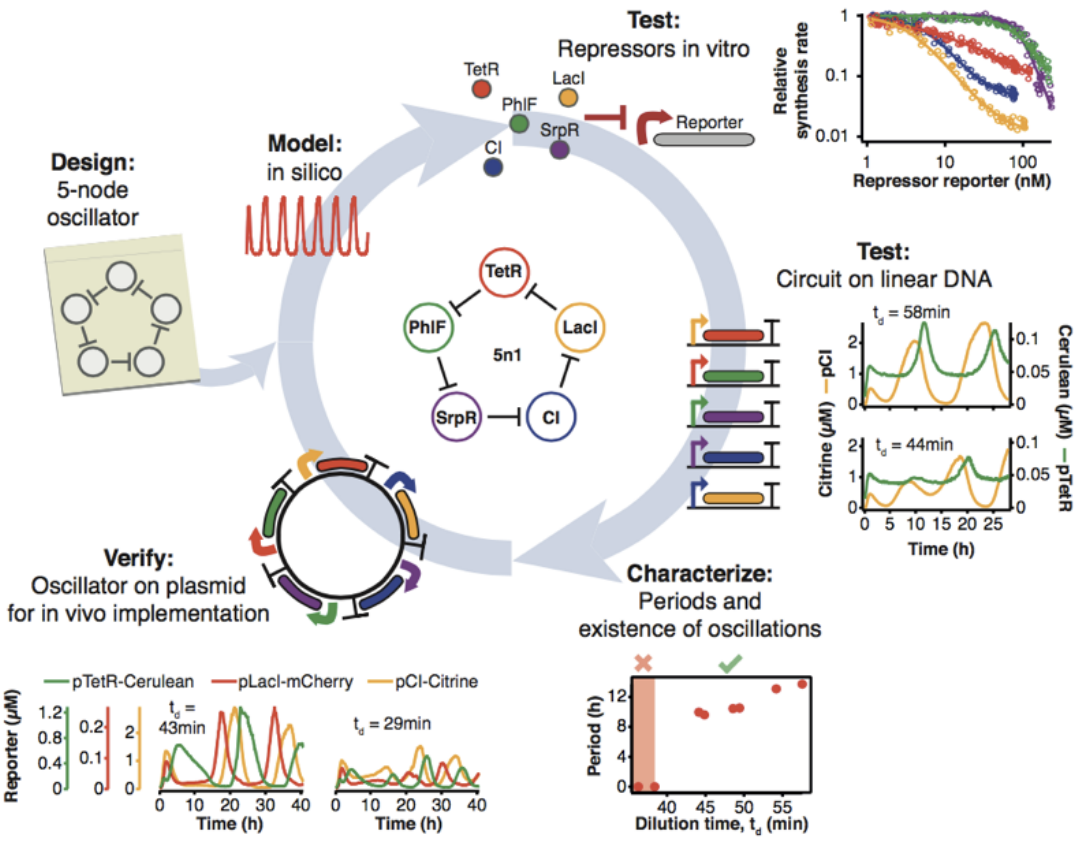


Fig. 4: Synthetic biology workflow using cell-free prototyping [42].

are able to prototype circuits with a cycle time of less than 8 hours from design to data [42, 63]. We have successfully used TX-TL prototyping to implement regulatory circuits [20], decision-making logic [18, 23], novel genetic oscillators [42], and metabolic pathways [70, 40, 74]. We have also developed and published improved methods for using TX-TL as a prototyping tool [62, 64] and as an educational resource [22, 23]. Using an acoustic liquid handling system (Labcyte Echo), we are able to test hundreds of circuit and pathway variants at a time, enabling high throughput data collection and design space exploration.

An example of the design workflow for TX-TL-based circuit design is shown in Fig. 4. Starting at the left and proceeding clockwise, the workflow begins with a design concept and mathematical model of the desired circuit. Individual components are characterized and tested, then combined for circuit-level testing and characterization. Finally, the entire circuit is assembled onto a single plasmid for final validation. This design-build-test cycle can be iterated multiple times, if needed. In this project we will adapt this workflow to focus on synthetic cells, with the *in vivo* implementation replaced with a synthetic cell implementation.

Characterization of biological circuits Cell-free systems also play an important role in our ability to develop high fidelity models and characterize individual biological components, whether for future cell-free usage or for cell-based circuits. Examples in my group include characterization of genetic context effects due to supercoiling [71], occupancy effects in integrase circuits [2] and effects of resource limits on circuit performance [58]. These results all rely on a combination of modeling, analysis, and (cell-free) experiments. In addition, we have developed open source modeling toolboxes that capture many of the important details of cell-free systems [66].

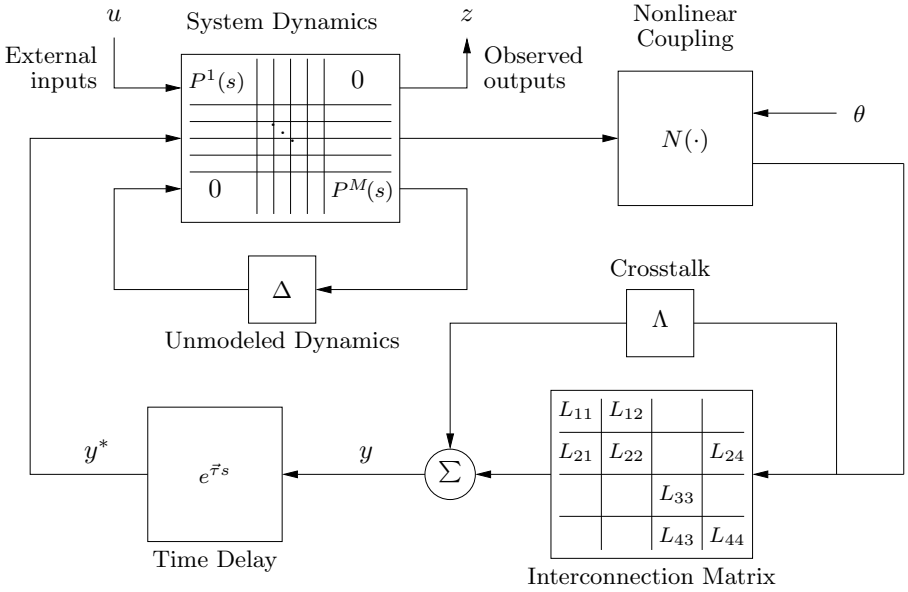


Fig. 5: Interconnection analysis framework.

2.2 Modeling, Analysis, and Design

In order to design the types of synthetic cells that we envision, it will be important to have an underlying mathematical and computational framework for modeling, analysis, and design. We plan to build on concepts from (electrical) circuit theory, control theory, and stochastic systems to provide a set of tools (theory and algorithms) that can take into account stochastic dynamics of biomolecular systems, interactions between components and subsystems, and the need for multi-layered design abstractions that hide complexity of lower layer functions.

Networked systems structure A key challenge in developing models for any class of problems is the selection of an appropriate mathematical framework for the models. Among the features that we believe are important for a wide variety of biological systems is capturing the temporal response of a biomolecular system to various inputs and understanding how the underlying dynamic behavior leads to a given phenotype. The models should reflect the subsystem structure of the underlying dynamical system to allow prediction of results, but need not necessarily be mechanistically accurate at a detailed biochemical level. We are particularly interested in those problems that include a number of molecular "subsystems" that interact with each other, and so our models should support a level of modularity (with the additional advantage of allowing multiple groups to develop detailed models for each module that can be combined to form more complex models of the interacting components). Since we are likely to be building models based on high-throughput experiments, it is also key that the models capture the measurable outputs of the systems. Fig. 5 shows a block diagram representation of one possible modeling framework.

For many of the systems that we are interested in, a good starting point is to use reduced-order models consisting of nonlinear differential equations, possibly with some time delay. Using the basic structure shown in Fig. 5, a model for a multi-component system might be described using a set of input/output differential equations of the form

$$\begin{aligned} \frac{dx_i}{dt} &= Ax_i + N(x_i, Ly^*, \theta) + Bu_i + Fw_i, \\ y_i &= Cx_i + Hv_i \qquad y_i^*(t) = y_i(t - \tau_i). \end{aligned} \tag{1}$$

The internal state of the ith component (subsystem) is captured by the state $x_i \in \mathbb{R}^{n_i}$, representing the concentrations of various species and complexes as well as other internal variables required to describe the dynamics. The "outputs" of the subsystems $y_i \in \mathbb{R}^{p_i}$ describe species (or other quantities) that interact with other subsystems in the cell. The internal dynamics consist of a set of linear dynamics Ax as well as nonlinear terms $N(x, Ly^*, \theta)$ that depend both on the internal state and the outputs of other subsystems, where Ly^* represents interconnections with other subsystems and θ is a set of parameters that represent the context of the system. We also allow for the possibility of time delays (due to folding, transport or other processes) and write y_i^* for the "functional" output seen by other subsystems (in Fig. 5, time delays are represented by their Laplace transform $e^{\vec{\tau}s}$).

The coupling between subsystems is captured using a weighted graph, whose elements are represented by the coefficients of the interconnection matrix L. In the simplest version of the model, we simply combine different outputs from other modules in some linear combination to obtain the "input" Ly^*. More general interconnections are possible, including allowing multiple outputs from different subsystems to interact in nonlinear ways (such as one often sees on combinatorial promoters in gene regulatory networks). The structure of L corresponds to the interactions within a subsystem (L_{ii} blocks) and between subsystems (L_{ij} blocks, where $i \neq j$).

Finally, in addition to the internal dynamics and nonlinear coupling, we separately keep track of external inputs to the subsystem (Bu), stochastic disturbances (Fw) and measurement noise (Hv). We treat the external inputs u as deterministic variables (representing inducer concentrations, nutrient levels, temperature, etc.) and the disturbances and noise w and v as random processes. If desired, the mappings from the various inputs to the states and outputs, represented by the matrices B, F and H can also depend on the system state x (resulting in additional nonlinearities).

The mathematical structure in Fig. 5 and equation (1) captures a large number of modeling frameworks in a single formalism. In particular, mass action kinetics and chemical reaction networks can be represented by equating the stoichiometry matrix with the interconnection matrix L and using the nonlinear terms to capture the fluxes, with θ representing the rate constants. We can also represent typical reduced-order models for transcriptional regulatory networks by letting the nonlinear functions N represent various types of Hill functions and including the effects of mRNA/protein production, degradation and dilution through the linear dynamics. These two classes of systems can also be combined, allowing a very expressive set of dynamics that is capable of capturing many relevant phenomena of interest in molecular biology.

Computational tools and algorithms The mathematical techniques described above provide a basic framework for modeling, analysis, and design. To be useful, these techniques must be implemented as computational algorithms that can be used by circuit-, subsystem-, and system-level engineers to carry out a design at the appropriate layer of abstraction. Recent results have become available that demonstrate these types of computer aided design tools, such as NUPACK [73], Cello [43] and the GRSM compiler [51].

Previous work at Caltech has developed a diverse set of computational tools for modeling core mechanisms in biological systems in a manner that allow exploration of different levels of fidelity in models of individual subsystems as well as interconnection between subsystems (including preliminary support for multi-compartment models). Fig. 6 shows some of the types of analyses that are possible using a combination of BioCRNpyler, which "compiles" high level descriptions of biomolecular circuits into detailed SBML models, and BioSCRAPE, which performs both simulation and parameter inference. In this project we will expand these tools to allow incorporation of multi-component systems, including mechanisms for compartmentalization, diffusion, and active transport.

In addition to modeling and simulation tools, the BioSCRAPE Inference toolbox can be used to

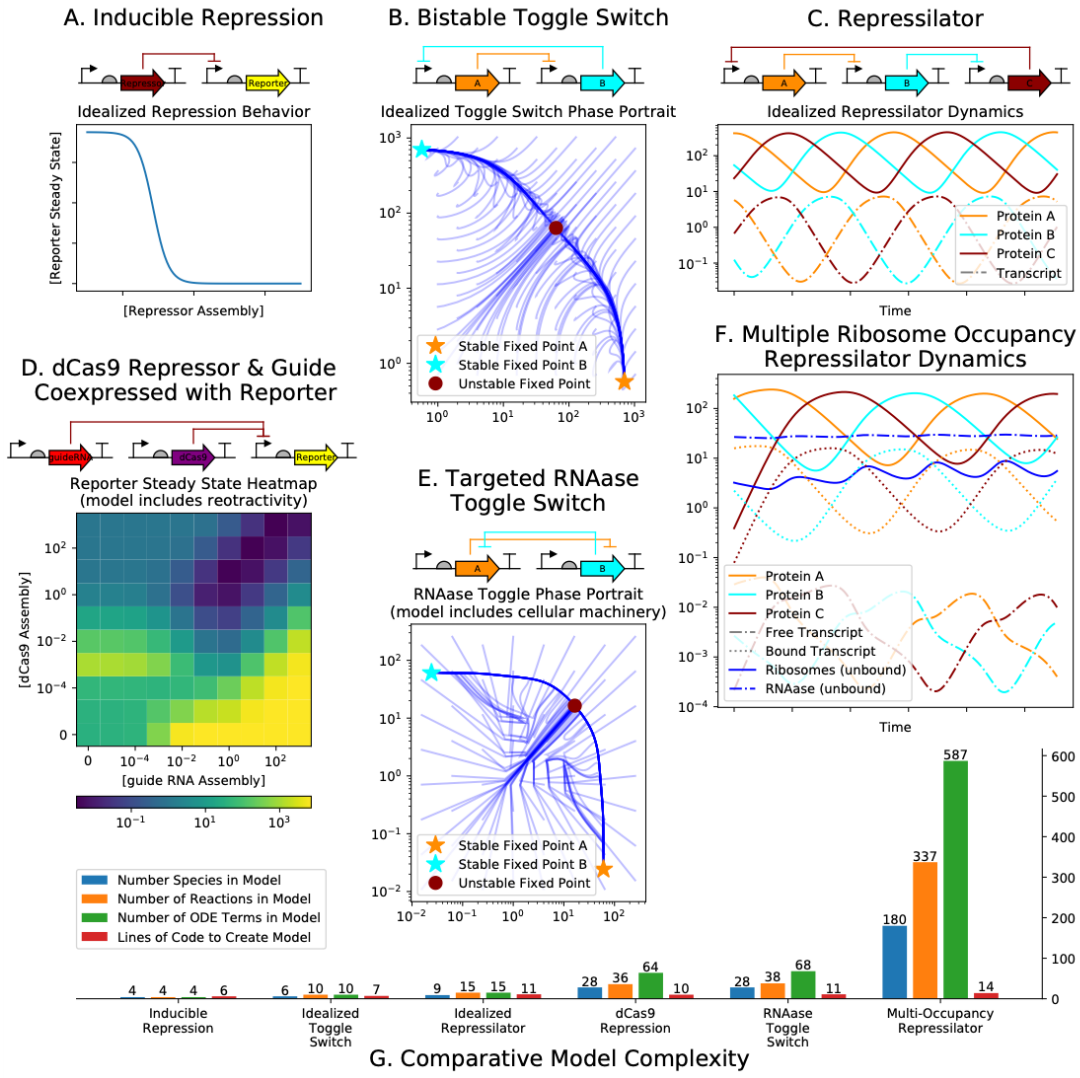


Fig. 6: Examples of modeling and analysis using BioCRNpyler [49]

perform Bayesian inference of parameter values based on experimental data. These techniques can be applied to the experimental results to allow creation of models that recapitulate experimental results and are suitable for use in computational analysis of interaction effects. In addition to models that can be used for design and optimization, Bayesian inference techniques also allow validation of modeling assumptions by investigation of the distributions that are returned for the parameter values in a model. If all parameters have unimodal distributions with low coefficient of variation, this provides evidence of the validity of the model, while multi-modal distributions or distributions with a broad distribution of equally likely parameters can indicate the the model does not capture the experimental data. These results can then be used to hypothesize and test new molecular mechanisms that may be relevant for the experimental conditions.

The combination of modeling and system identification will serve as a basis for testing hypotheses regarding how circuits and subsystems interact with each other, and for providing insights into situations in which (initially) unexpected interactions between subsystems cause failure in the desired capabilities that we are trying to demonstrate.

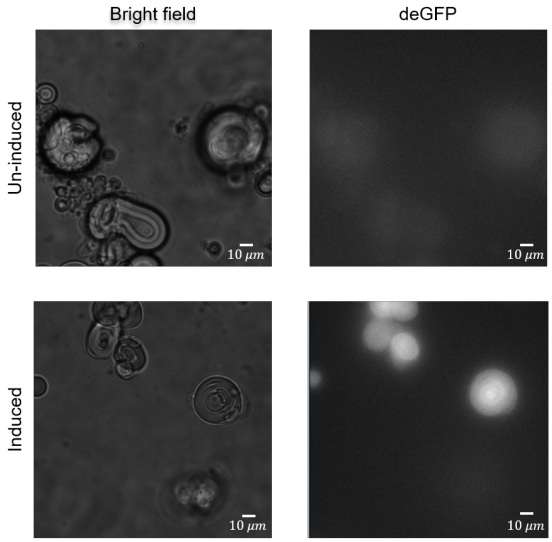
Fundamental research contributions The use of model-based approaches to synthetic biology is still in its infancy. Most groups today use models as a means of explaining their experimental results, rather than as predictive tool for design. The research proposed above will provide new methods for stochastic modeling, system identification, circuit, subsystem and whole-system analysis. While driven by the problem of genetically-programmed synthetic cells, we anticipate that many of the techniques that we develop will be useful in other areas of synthetic and systems biology.

2.3 Circuits and Subsystems

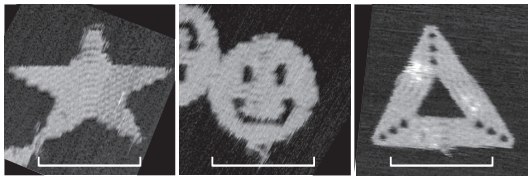
To build a synthetic cell of the sort conceptualized in Fig. 1, a variety of system functions will need to be implemented. We envision that each of these functions would represent a collection of biomolecular constructs (scaffolds, circuits, pathways) that operate to together as a subsystem. In this section we provide an overview of the key subsystems shown in Fig. 1 and how they might be implemented in a synthetic cell. As noted elsewhere, we do not propose here to develop *all* of these technologies, but rather to build on existing results and focus on integrating a subset (3–4) that would demonstrate the key concepts and serve as driver for the modeling, analysis, and design tools described in Section 2.2.

Spatial organization A key element of a synthetic cell will be the ability to spatially localize biomolecular components, in the same way that natural organisms use scaffolds, organelles, cell walls, and other mechanisms. Two possible technologies that can be used to provide spatial organization are phospholipid vesicles and (3D) DNA origami. Both have already been demonstrated to some extent. Technical challenges include efficient means of encapsulating the desired genetically-programmed elements within the container and identifying means of transferring signals and molecules across the container boundaries.

Phospholipid vesicles A starting point for synthetic cells is the development of phospholipid vesicles that can contain TX-TL extract, as shown in Fig. 7a. Building on work by Adamala and Martin-Alarcon et al. [1], we have demonstrated several of the key experimental capabilities needed for inserting functional transmembrane proteins into vesicle membranes. In particular, we have generated vesicles containing TX-TL and DNA encoding the α-hemolysin protein that forms pores in lipid bilayers. We expressed this protein in the vesicles and used a membrane impermeable chemical inducer to generate fluorescence if the pore was formed. Fig. 7a shows bright- and dark-field images of the vesicles with and without IPTG, demonstrating that fluorescence is observed only if the pore was formed in the liposome allowing diffusion of IPTG into the proteoliposome. This result demonstrates that we can express membrane proteins in vesicles and verify that they are functional.



(a) Phospholipid vesicles



(b) DNA origami [52]

Fig. 7: Spatial localization technologies.

DNA origami Over the past decade, structural DNA nanotechnology has allowed us to create molecular programs that self-assemble into arbitrary shapes and patterns. In work at Caltech, Paul Rothemund

has pioneered the technique of DNA origami to create custom patterns and shapes up to 100 nanometers per dimension in size [52], as illustrated in Fig. 7b. These structures arise out of a diffusional dance in solution between 200 short "staple" strands and a single 7000-base "scaffold" strand. Many groups around the world use DNA origami, demonstrating its robustness as a molecular self-assembly platform for diverse nanodevices.

In this project, the potential use of DNA origami is twofold: as a structural platform for implementing a synthetic cell and as a scaffold for providing spatial localization. Both uses have already been partially demonstrated: 3D DNA origami containers have been built that can isolate chemicals from their environment [12], and various DNA scaffolds have been used to localize DNA, RNA, and proteins to fixed locations [8, 55]. My group has active collaborations with Rothemund, and my students have served as TAs for courses at Caltech that make use of DNA origami technology, thus facilitating the integration of this technology into our workflow, as needed.

Metabolism Metabolic subsystems will be responsible for providing both the energy required for the system to operate as well as small molecules, proteins, and other species required in the operation of the synthetic cell. We have already demonstrated the ability to implement simple metabolic processes within TX-TL [70, 40, 74], but scaling this up to provide a (minimal) metabolic network is a major challenge. Cell extract can be used initially to provide the required functionality, with the possibility of re-energizing natural metabolic pathways [28].

An example of the sort of metabolic pathway that can be implemented in a cell-free system is shown in Fig. 8. This figure shows the implementation of a pathway for production of 1,4-butanediol (BDO), which was used as part of a joint project with Genomatica, Inc [70]. This pathway is an example of the sort of small molecule metabolic pathway that is possible today. Other work in using TX-TL for metabolic pathways include demonstration of the polypeptide valinomycin [74] (done by an undergraduate in my group), 2,3-BDO [29] and violacein [40].

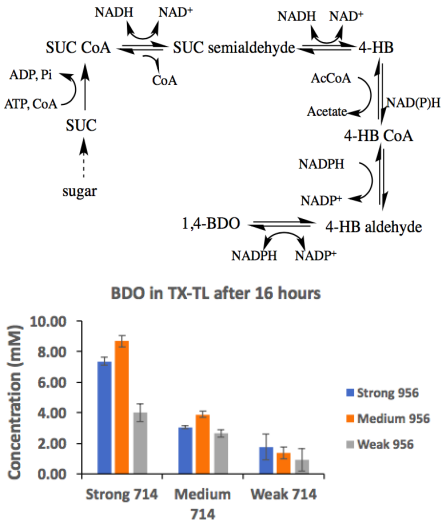


Fig. 8: Cell-free synthesis of 1,4-BDO [70].

Sensing and signaling In order for synthetic cells to interact with the environment, they will need to sense external conditions. Vesicle- or origami-bound proteins are natural way approaches to explore, building off of the various molecular sensors provided in natural biological systems. In addition, signaling cascades will be needed to amplify and process environmental signals.

The incorporation of functional proteins into artificial vesicles for signal transduction has not yet been demonstrated, but many groups are working on such technologies. We have recently collaborated with Amgen, Inc. to demonstrate the ability to express membrane bound proteins in TX-TL using phospholipid "nanodisks" that emulate the cell membrane. We used these nanodisks to explore the functionality of a two-component signaling system operat-

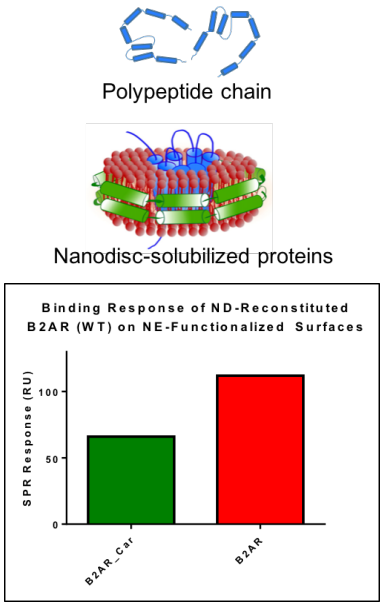
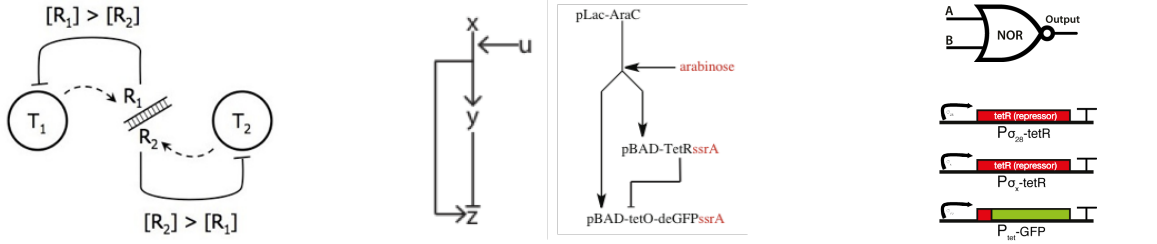
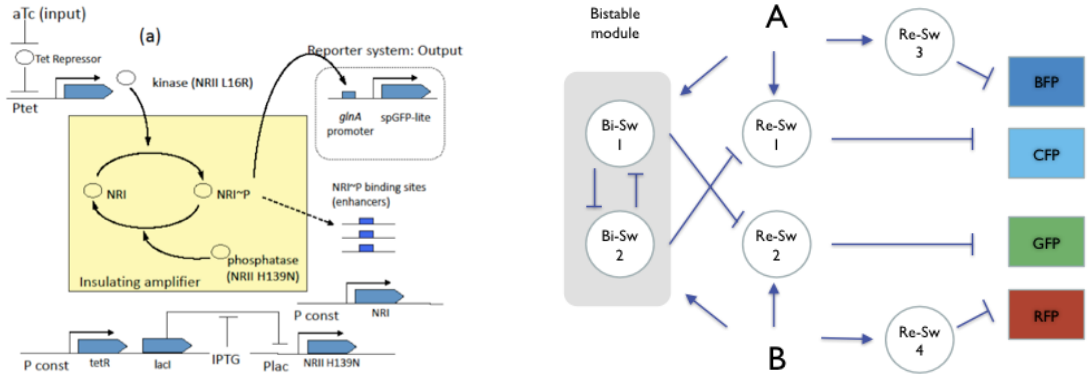


Fig. 9: Expression of membrane-bound proteins using nanodisks.



(a) Genelet rate regulator [15] (b) Incoherent feedforward loop [18] (c) TX-TL logic gate [23]



(d) Phosphorelay-based insulator [20] (e) A then B event detector [72]

Fig. 10: Cell-free regulatory and computational circuits.

ing in TX-TL [19], as shown in Fig. 9. Specifically, we have tested $\beta 2$ adrenergic receptor (BCAR) proteins binding to an SPR surface coated with norepinephrine, with and without carazolol, demonstrating proper protein function through a fluorescence-based carazolol binding assay.

We have also demonstrated functionality of sensing and signaling-like circuits using genelet-, transcription-, and phosphorylation-based components and transcriptional regulation. Genelet components include an incoherent feedforward loop allowing adaptation to signal levels [32] and an insulation circuit that allows minimization of circuit loading [14]. Transcriptional circuits that might be used in signaling pathways include a molecular sensor to detect vanillin [9] and a fold-change detection circuit [18]. Finally, a phosphorylation-based "insulator" was tested in TX-TL and demonstrated the ability to reduce circuit loading [20]. Each of these circuits serves as a source of experience that can be utilized for implementing more complex sensing and signally subsystems.

Regulation and computation Central to the operation of the synthetic cell is the ability to regulate the various operations of its subsystems and to perform computations that enable interesting behaviors and functions. This is perhaps the area that is most well-studied in my group's past research, although the complexity of circuits we are currently able to build is far below what would be required for even a minimal synthetic cell.

Several preliminary results are available demonstrating implementation of regulatory and computational elements in cell-free settings, as illustrated in Fig. 10. Genelet technology has been used to create oscillators [35], oscillators driving a set of DNA "tweezers" [14], a feedback controller to regulate the the rates of production of RNA [15], and the previously-mentioned incoherent feedforward loop [32]. We have also demonstrated the implementation of a set of two input logic gates in TX-TL, as part of a demonstration project at the DARPA "Wait, What?" conference [23] (now being transitioned for use in high school laboratories as part of the BioBuilder curriculum). circuits.

In addition to the use of cell-free circuits for regulation and computation, we also anticipate that many cell-based technologies will be amenable to implementation in synthetic cells. Work in my group that is directly relevant includes work with scaffold proteins to implement a concentration tracker [26], use of integrase circuits for temporal logic gates [27], and ongoing work in the use of CRISPR-based guide RNA logic (done jointly with Niles Pierce's group at Caltech).

Actuation and locomotion In some instances, it may be useful for synthetic cells to be able to move in their environment, perhaps using cilia- or flagella-like mechanisms. Another possibility is to control the shape of cells using filaments or other cytoskeletal structures. This is an area of high risk, since the molecular actuators that are present in biological systems are much more complicated than anything that has been demonstrated in cell-free environments to date.

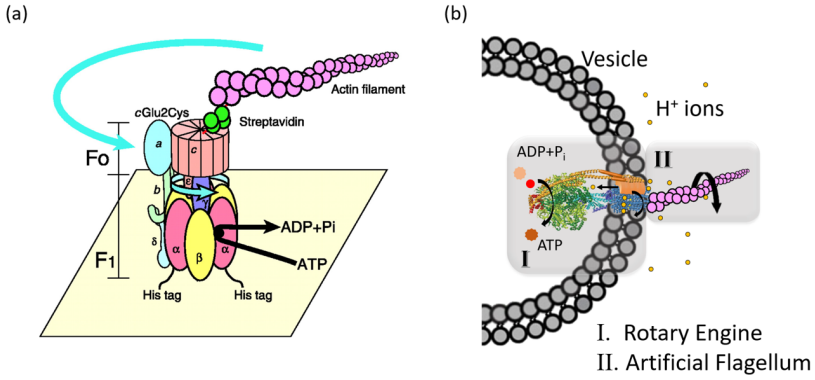


Fig. 11: (a) Diagram of ATPase reconstituted to rotate an actin filament [53]. (b) Overview of the proto-flagella design with ATPase crystal structure [3].

There are multiple approaches that might be taken to move forward this particular subsystem technology. One possibility would be to make use of DNA nanostructures such as the tweezers that were used in previous joint work with Eric Winfree and Fritz Simmel [14]. A much more complicated approach would be to attempt to implement the flagellar machinery of a microorganism, perhaps first by using nanodisks for prototyping the membrane-bound machine that is found in living cells.

We have received funding under the NSF EAGER program to pursue to development of a proto-flagellar motor that builds on recent efforts to reproduce assembly of a bacterial flagellum starting from partially complete structures [65]. We are developing a flagella-inspired minimal propulsion complex consisting of an ATP-driven rotary engine fused to self-assembling protein filaments that act as external propellers (Fig. 11b).

Materials input/export and communications Another useful function for synthetic cells is export of small molecules or proteins into the environment. These can be used for communication between cells, materials production, or export of enzymes. Although there are a variety of export proteins available in nature, none have been demonstrated to work in synthetic vesicles and (3D) DNA origami approaches have so far demonstrated limited capabilities of this sort.

Given our ability to prototype membrane-bound circuits in cell-free systems, we anticipate the use of various chemical transporters and efflux pumps as a means of exploring this direction. We are also pursuing two independent projects in the use cell-based multi-cellular communication using AHLs [11, 48]. The use of microfluidically controlled environments and phospholipid vesicles provides a starting put for moving these cell-based technologies into a synthetic cell setting.

Fundamental research contributions The concepts listed above represent an approach to developing a collection of circuit and subsystem functions in a synthetic cell context. Some of them will involve transitioning concepts from other settings into a synthetic cell environment. Others will undoubtedly evolve into more fundamental research projects that focus on the core mechanisms involved and provide new insight into how biomolecules function and how we can engineer novel behaviors using existing and new components. Importantly, it is the combination of the analytical

approaches described in Section 2.2 with the experimental efforts in this section where we see the most potential for new fundamental research advances.

In addition to the specific circuits and subsystems, an underlying need is to develop methods for design of *programmable* genetic elements for implementing biomolecular circuits. Three leading technologies are genelet-based circuits (or other RNA-based methods, such as those developed by Julius Lucks [36]), integrase-based circuits, which can be programmed through arrangement of (possibly nested) attachment cites [4, 27, 59], and CRISPR-based circuits, which can be programmed through the use of guide-RNAs whose sequence and/or secondary structure can be used to modulate the dynamics of the circuit [17, 44, 45, 50]. The overriding feature that is needed is the ability to program the operation and interaction between circuits and subsystems, which will serve as an enabler for building systems with hundreds of engineered biological components (built by multiple groups, using a interoperable approach).

2.4 Systems Engineering and Technology Demonstrators

In addition to the individual subsystems and functions described above, two major challenges will be the integration of the subsystems and assembly of the overall artificial cell. It is likely that compartmentalization will be required to limit the interactions between subsystems (similar to the spatial organization present in natural cells). How to assemble the various subsystems into a functioning artificial cell is also a major challenge. One possibility would be to encapsulate the DNA encoding the various functions into a vesicle, as described above, and then establishing an "assembly" process by which the subsystems would self assemble and integrate into the chassis wall (where appropriate).

To address these issues and to demonstrate the results of the theoretical framework (Section 2.2) and circuit/subsystem designs (Section 2.3), we plan to explore the development of two or more "system-level" technology demonstrators. Each demonstrator will bring together multiple subsystems in a way that creates a more complex set of functions than currently available and provides insights into how more complex artificial cells and multi-cellular machines can be created. In this section we describe our initial ideas for demonstrators that we could pursue, although these are likely to change as the project unfolds and new technologies are developed (by us and by others).

Distributed event detection One broadly relevant application of biologically engineered systems is the detection of small molecules (chemical signatures) and the monitoring and logging of sequences of events. An "event detector" is a circuit that allows the detection of a pattern of chemical inputs that might vary in terms of species combinations, relative magnitudes, and temporal timing. An "event logger" is a circuit that records a sequence of events (or environmental states) in a manner that can recovered at a later time. Preliminary research includes the detection of small molecules [9, 16, 60], implementation of logical functions [5, 38, 59, 68], detection of sequences of events [27, 51], and methods for implementing long-term memory [59]. My group has experience with all of these technologies (as described in Section 2.3) and the capability to integrate the works of others, as appropriate.

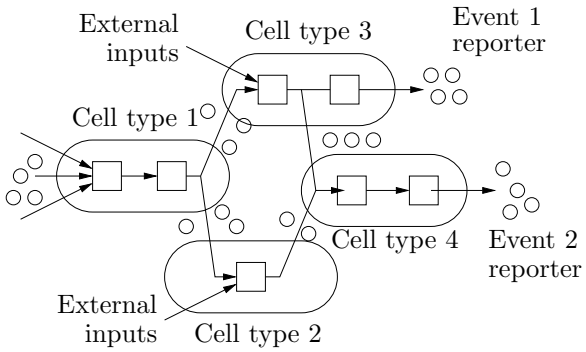


Fig. 12: Distributed sensing concept.

As a demonstration of the concept of synthetic cells, we will encapsulate components of event detection and logging circuits in individual phospholipid vesicles and use small molecules as a means for components that are in different cells to communicate with each other. This would be similar to the conceptual diagram shown in Fig. 12, where the square boxes represent individual functions

(detectors, logic, memory) and the small circles represent chemical signals. Preliminary demonstrations of vesicle-based distributed operations have recently been demonstrated by Adamala and Martin-Alarcon et al. [1], and our own work [27] has demonstrated the use of integrase-based circuits as a means of both implementing temporal event detection ("a then b" logic) and providing long term memory.

There are several advantages of using multi-cellular event detection techniques. By creating "modules" that implement common subfunctions, a variety of behaviors can be created by controlling which circuit elements are combined together in a multi-cellular machine. In addition, by exposing a population of artificial cells to environmental conditions and measuring the distribution of the response, it is possible to achieve new types of measurement and control functionality. For example, our recent work using integrase-based event detectors demonstrated the ability to infer timing, duration and amplitude of pulses of chemical events by looking at the distributional response of a population of cells [27].

Flagellar-controlled locomotion As a second demonstration, and one that is considerably higher risk than the others, we will integrate various subsystem technologies that are already demonstrated or under development to implement a chemotaxis-like mechanism for a synthetic cell.

The implementation of chemotaxis in natural systems is well understood and the basic operation is illustrated in Fig. 13. Examples of chemotaxis include the ability of organisms to move in the direction of nutrients or move away from toxins in the environment. Many chemotaxis mechanisms are stochastic in nature, with biased random motions causing the average behavior to be either positive, negative or neutral (in the absence of stimuli). The chemotaxis system in *E. coli* consists of a sensing system that detects the presence of nutrients, an actuation system that propels the organism in its environment, and control circuitry that determines how the cell should move in the presence of chemicals that stimulate the sensing system.

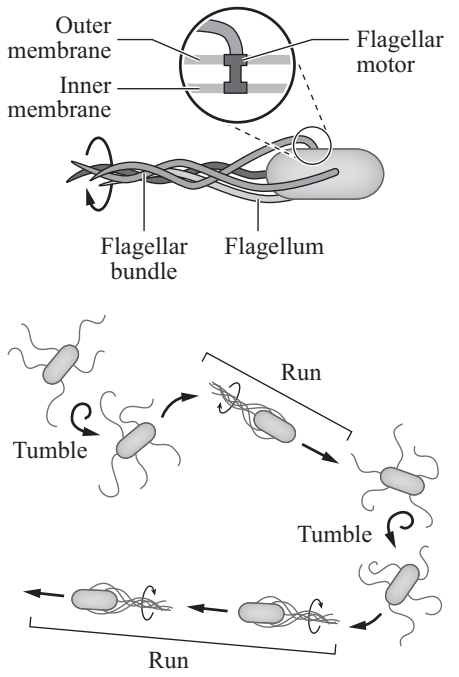


Fig. 13: Bacterial chemotaxis (from Del Vecchio and Murray [10, Fig 5.14]).

We will attempt to implement a chemotaxis-like mechanism in a synthetic cell by coupling a sensing subsystem, a decision-making subsystem, and an actuation subsystem, each of which are either already available or being developed under separate projects. While getting some of the individual technologies is complex, our focus here is on figuring out how to integrate multiple technologies in a bio-compatible manner that allows them to operate together.

Fundamental research contributions While the examples described above could be considered as just demonstrations of a set of technologies, proper execution of these demonstrations will serve as a basis for several different fundamental research contributions in cell-free (and likely cell-based) synthetic biology. We briefly summarize three such potential contributions here.

Managing crosstalk and uncertainty A fundamental challenge in synthetic biology is the creation of circuits with hundreds or thousands of components. How can we maintain robust performance in the presence of unwanted interactions caused by shared resources and other sources of cross-talk in systems with large numbers of of uncertain components? The system-level demonstrations proposed here will force us to confront these issues and find ways to manage crosstalk and uncertainty.

Multi-element ensembles for distributed computation Line natural cells, synthetic cells will likely exhibit distributional responses and carry out operations across a population of cells. How can we carry out distributed operations with highly stochastic elements and design population-level responses that operate across a broad set of operating environments?

Testbeds and characterization Based on past experience, a substantial amount of activity is likely to take place in the construction of test environments for artificial cells and multi-cellular machines and the tools used to characterize their behavior. How can we construct, test, and characterize synthetic cell systems in a high-throughput, quantitatively-accurate fashion to enable rapid design, build, test, learn (DBTL) cycle?

3 Broader Impacts

Team-based, undergraduate research The proposed project will be carried out by a combination of 2–4 graduate students (2 supported by NSF), a research technician (partially supported by NSF), and 6–8 undergraduate researchers. Undergraduate researchers will be an integral part of the activity and will focus on the development of individual subsystems and, working with the graduate students, the integration of those subsystems in various combinations, leading toward the final demonstrations. An initial implementation of this type of activity was carried out in the summer of 2020, when 8 undergraduates and 2 graduate students participated in a team-based summer project aimed at computational design, implementation, and integration of synthetic cell subsystems (the original plans for experimental work were not possible due to the COVID-19 pandemic).

All of the students working on this project will obtain broad experience with many aspects of biological engineering, and will able to work in many areas outside of cell-free systems (including more traditional approaches in synthetic biology). We also anticipate the participation of one or more visiting scientists (graduate students, postdocs, faculty and other researchers), who would participate in the project with support from their home institutions.

Dissemination and Interaction In addition to the development of the basic technology, a major element of the proposed effort will be the dissemination of tools for engineering biology to the broader community. We will use a variety of mechanisms, including establishing collaborations with leading research groups who are interested in the component and system technologies, visiting and interacting with other researchers to identify new research opportunities, offering short courses to expose interested researchers to the underlying technologies, and maintaining "open source" repositories of algorithms, protocols, and techniques. These activities are already a central part of our current activities in cell-free synthetic biology, including running three TX-TL workshops, maintaining all protocols on the OpenWetWare web site, and providing past support for approximately 15 groups (academic, industry, and government) who are using the biomolecular breadboards that we have helped to develop.

Build-A-Cell and fabriCELL interactions The technical work in this proposal will be carried out in the context of two ongoing projects in the US and the UK focused on development of synthetic cells (Build-A-Cell and fabriCELL). We will make use of our involvement with those projects, including two NSF-supported collaborations with researchers at Imperial College in London, to establish a stronger bridge between them and bring together technologies not available in concert elsewhere in the world. We will also develop and disseminate standardized protocols for cell-free systems operating in vesicles, which are required in order to allow a larger community to build on this work. This will enable others to engineer motility and event detection functions into their systems as well as exploiting the tools and technologies of this project to build systems exhibiting other behavioral modules. Personnel exchange between Caltech and Imperial College, funded via

other ongoing projects, will facilitate both the standardization and bridging objectives, as well as provide valuable experiences for students and postdocs. Finally, we will make use of existing activities at Caltech and Imperial to maintain a strong public discourse regarding synthetic biology and to engage the broader public in our work.

4 Work Plan

The following high-level milestones will be used to accomplish the overall project objectives.

Year 1 Establish the initial approach to design and implementation of circuits and subsystems in synthetic cells, as well as connections to other researchers: • Implement at least two input/output circuits or pathways in a synthetic cell and characterize their performance. • Make use of the mathematical framework in Section 2.2 to model all circuits and subsystems. • Establish connections with at least three non-Caltech labs that can serve as partners for research in synthetic biology, and run a "boot camp" on cell-free synthetic biology.

Year 2 Demonstrate the ability to design and implement a "subsystem" using prior work and results from Year 1: • Implement and characterize at least one biomolecular subsystem capable of operating in a spatially-isolated, cell-free environment (vesicle- or origami-based). • Expand the mathematical framework of Section 2.2 to account for uncertain behavior, including unmodeled dynamics Δ and crosstalk Λ. • Host one or more visitors for multi-week (ideally multi-month) visits to participate in the project, and run a second "boot camp" on cell-free synthetic biology.

Year 3 Design and implement multiple subsystems, and prototype a method of assembling them into a multi-cellular machine: • Implement three or more biomolecular subsystems operating in spatially-isolated, cell-free environments and characterize their robustness properties. • Demonstrate one or more methods for assembling multiple biomolecular subsystems into a multi-element ensemble with structured interactions between the individual subsystems. • Make use of the expanded mathematical framework from Year 2 as an integral element of the design of all subsystems. • Host multiple visitors for a multi-week or longer visits and run a workshop on synthetic cells.

Year 4 Optimize designs and establish robustness properties of multi-cellular machines: • Develop and demonstrate a suite of design tools (mathematical framework and supporting software) for modeling, analysis, and design of circuits, subsystems and (synthetic) cells. • Demonstrate at least two multi-subsystem machines that validate the proposed design framework, subsystem implementations, and assembly techniques. • Run a workshop on cell-free synthetic biology (artificial cells and multi-cellular machines) that involves collaborators from government labs and academia, with the goal of highlighting work in the field and identifying future research challenges.

5 Results of Prior NSF Support

Molecular Programming Architectures, Abstractions, Algorithms, and Applications (R. M. Murray, NSF 1317694; $2M total/year [for 10 co-investigators], 1 Oct 2013–30 Sep 2019). The goal of the Molecular Programming Project (MPP) is to develop abstractions, languages and tools needed to design and implement artificial molecular programs. • Intellectual Merit: Murray's work focused on the development of *in vitro* circuits that demonstrate the principles of feedback in biomolecular systems and the application of cell-free assays as a "biomolecular breadboard" for molecular programming. • Broader Impact: Broader impact included development of a textbook on biomolecular feedback systems [10] providing undergraduate research opportunities for approximate 10 students. • Publications: Two conference publications [21, 24], four journal publications [32, 37, 54, 57], a technical report [33], and a textbook [10] have appeared based on work supported by this grant.

References

[1] K. P. Adamala, D. A. Martin-Alarcon, K. R. Guthrie-Honea, and E. S Boyden. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nature Chemistry*, 2016.

[2] G. Artavanis, V. Hsiao, C. A. Hayes, and R. M. Murray. The role of single occupancy effects on integrase dynamics in a cell-free system. In *Synthetic Biology: Engineering, Evolution and Design (SEED) Conference*, 2016. Available at `biorxiv.org/content/early/2016/06/18/059675`.

[3] C. von Ballmoos, A. Wiedenmann, and P. Dimroth. Essentials for ATP synthesis by F1F0 ATP synthases. *Annual Review of Biochemistry*, 78:649–672, 2009.

[4] J. Bonnet, P. Subsoontorn, and D. Endy. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, 109:8884–8889, 2012.

[5] J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy. Amplifying genetic logic gates. *Science*, 340(6132):599–603, 2013.

[6] Build-A-Cell Consortium. `http://buildacell.io`.

[7] F. Caschera and V. Noireaux. Compartmentalization of an all-E. coli cell-free expression system for the construction of a minimal cell. *Artificial Life*, 22(2):185–195, 2016.

[8] Y.-J. Chen, B. Groves, R. A. Muscat, and G. Seelig. DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology*, 10(9):748–760, 2015.

[9] E. L. C. de los Santos, J. T. Meyerowitz, S. L. Mayo, and R. M. Murray. Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: Developing a vanillin sensor. *ACS Synthetic Biology*, 5:287–295, 2016.

[10] D. Del Vecchio and R. M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, 2014.

[11] Joy Doong, James Parkin, and Richard M. Murray. Length and time scales of cell-cell signaling circuits in agar. *bioRxiv*, 2017.

[12] S. M. Douglas, I. Bachelet, and G. M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

[13] fabriCELL Consortium. `http://fabricell.org`.

[14] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, and F. C. Simmel. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences*, 108(40):E784–E793, 2011.

[15] E. Franco, G. Giordano, P.-O. Forsberg, and R. M. Murray. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synthetic Biology*, 3(8):589–599, 2014.

[16] C. E. French, K. de Mora, N. Joshi an A. Elfick, J. Haseloff, and J. Ajioka. *Institute of Medicine (US) Forum on Microbial Threats. The Science and Applications of Synthetic and Systems Biology: Workshop Summary*, chapter A5. Synthetic Biology and the Art of Biosensor Design. National Academies Press, 2011.

[17] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S Qi. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451, 2013.

[18] S. Guo, Y. Hori, and R. M. Murray. Systematic design and implementation of a novel synthetic fold-change detector biocircuit in vivo. In *Winter q-bio Conference*, 2015. `www.cds.caltech.edu/~murray/preprints/ghm15-wqbio_s.pdf`.

[19] S. Guo, A. Vaish, Q. Chen, and R. M. Murray. Expressing biologically active membrane protein in a cell-free transcription-translation platform. Technical report, bioRxiv, 2017. Submitted, *Synthetic and System Biotechnology*. Available at `biorxiv.org/content/early/2017/01/30/104455`.

[20] S. Guo, E. Yeung, K. Soli Nilgiriwala, D. Del Vecchio, and R. M. Murray. Implementation and simulation of phosphorylation-based insulator in transcription-translation platform. Technical report, California Institute of Technology, 2014. `www.cds.caltech.edu/~murray/preprints/guo+14-wqbio_s.pdf`.

[21] W. Halter, F. Allgower, R. M. Murray, and A. Gyorgy. Optimal experiment design and leveraging competition for shared resources in cell-free extracts. In *Proc. American Control Conference*, pages 1872–1869, 2018.

[22] C. Hayes and N. Kuldell. Cell-free circuits: Molecular diagnostics outside the cell. BioBuilder Laboratory Notes, 2017. In preparation.

[23] C. Hayes and R. M. Murray. DARPA "Wait, What?": Build a DNA circuit. Web site, September 2015. `http://txtlbio.org/waitwhat`.

[24] C. A. Hayes, E. L. C. de los Santos, E. Yeung, S. R. Sanchez, S. Y. Lee, and R. M. Murray. Rapid in vitro engineering of 16 two-input logic gates. In *Winter q-bio Conference*, 2016.

[25] C. E. Hodgman and M. C. Jewett. Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering*, 14(3):261–269, 2012.

[26] V. Hsiao, E. L. C. de los Santos, W. R. Whitaker, J. E. Dueber, and R. M. Murray. Design and implementation of a synthetic biomolecular concentration tracker. *ACS Synthetic Biology*, 4:150–161, 2014.

[27] V. Hsiao, Y. Hori, P. W. K. Rothemund, and R. M. Murray. A population-based temporal logic gate for timing and recording chemical events. *Molecular Systems Biology*, 12(5):869, 2016.

[28] M. C. Jewett and J. R. Swartz. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1):19–26, 2004.

[29] J. E. Kay and M. C. Jewett. Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metabolic Engineering*, 32:133–142, 2015.

[30] A. S. Khalil and J. J. Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367, 2010.

[31] J. Kim, J.J. Hopfield, and E. Winfree. Neural network computation by in vitro transcriptional circuits. *Advances in Neural Information Processing Systems (NIPS)*, 17:681–688, 2004.

[32] J. Kim, I. Khetarpal, S. Sen, and R. M. Murray. Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Research*, 42(9):6078, 2014.

[33] J. Kim, J. F. Quijano, E. Yeung, and R. M. Murray. Synthetic logic circuits using RNA aptamer against T7 RNA polymerase. Technical report, bioRxiv, 2014. Available at `biorxiv.org/content/early/2014/09/04/008771`.

[34] J. Kim, K. S. White, and E. Winfree. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology*, 2:68, 2006.

[35] J. Kim and E. Winfree. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology*, 7:465, 2011.

[36] Julius B. Lucks, Lei Qi, Vivek K. Mutalik, Denise Wang, and Adam P. Arkin. Versatile rna-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*, 108(21):8617–8622, 2011.

[37] X. F. Meng, A.-A. Baetica, V. Singhal, and R. M. Murray. Recursively constructing analytic expressions for equilibrium distributions of stochastic biochemical reaction networks. *Journal of the Royal Society, Interface*, 14(130):20170157, 2017.

[38] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature News*, 491(7423):249–253, 2012.

[39] N. Nandagopal and M. B. Elowitz. Synthetic biology: Integrated gene circuits. *Science*, 333(6047):1244–1248, 2011.

[40] P. H. B. Nguyen, Y. Y. Wu, S. Guo, and R. M. Murray. Design space exploration of the violacein pathway in *Escherichia coli* based transcription translation cell-free system (TX-TL). Technical report, bioRxiv, 2015. Available at `biorxiv.org/content/early/2016/01/26/027656`.

[41] H Niederholtmeyer, V Stepanova, and S. Maerkl. Implementation of cell-free biological networks at steady state. *Proceedings of the National Academy of Sciences*, 110(40):15985–15990, 2013.

[42] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl. Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife*, 4:e09771, 2015.

[43] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. Genetic circuit design automation. *Science*, 352(6281), 2016.

[44] A. A. K. Nielsen and C. A. Voigt. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology*, 10(11):763–763, 2014.

[45] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, 2014.

[46] V. Noireaux, Y. T. Maeda, and A. Libchaber. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proceedings of the National Academy of Sciences*, pages 3473–3480, 2011.

[47] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin, and J. J. Collins. Paper-based synthetic gene networks. *Cell*, 159(4):940–954, 2014.

[48] James M. Parkin and Richard M. Murray. Long-distance communication in synthetic bacterial consortia through active signal propagation. *bioRxiv*, 2018.

[49] William Poole, Ayush Pandey, Andrey Shur, Zoltan A. Tuza, and Richard M. Murray. Biocrnpyler: Compiling chemical reaction networks from biomolecular parts in diverse contexts. *bioRxiv*, 2020.

[50] L. S Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013.

[51] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu. Synthetic recombinase-based state machines in living cells. *Science*, 353(6297):aad8559, 2016.

[52] P.W.K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[53] Y. Sambongi, Y. Iko, M. Tanabe, H. Omote, A. Iwamoto-Kihara, I. Ueda, T. Yanagida, Y. Wada, and M. Futai. Mechanical rotation of the c subunit oligomer in ATP synthase (F0F1): direct observation. *Science*, 286(5445):1722–1724, 1999.

[54] S. Sen, D. Apurva, R. Satija, D. Siegal, and R. M. Murray. Design of a toolbox of rna thermometers. *ACS Synthetic Biology*, 6(8):1461–1470, 2017.

[55] A. Shaw, V. Lundin, E. Petrova, F. Fördös, E. Benson, A. Al-Amin, A. Herland, A. Blokzijl, B. Høgberg, and A. I. Teixeira. Spatial control of membrane receptor function using ligand nanocalipers. *Nature Methods*, 11(8):841–846, 2014.

[56] J. Shin and V. Noireaux. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and synthetic cell. *ACS Synthetic Biology*, 1(1):29–41, 2011.

[57] D. Siegal-Gaskins, E. Franco, T. Zhou, and R. M. Murray. An analytical approach to bistable biological circuit discrimination using real algebraic geometry. *Journal of the Royal Society, Interface*, 12(108):20150288, 2015.

[58] D. Siegal-Gaskins, Z. A. Tuza, J. Kim, V. Noireaux, and R. M. Murray. Resource usage and gene circuit performance characterization in a cell-free 'breadboard'. *ACS Synthetic Biology*, 3:416–425, 2014.

[59] P. Siuti, J. Yazbek, and T. K. Lu. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31:448–452, 2013.

[60] S. Slomovic and J. J. Collins. DNA sense-and-respond protein modules for mammalian cells. *Nature Methods*, 12(11):1085–1090, 2015.

[61] M. T. Smith, K. M. Wilding, J. M. Hunt, A. M. Bennett, and B. C. Bundy. The emerging age of cell-free synthetic biology. *FEBS Letters*, 588(17):2755–2761, 2014.

[62] Z. Z. Sun, C. A. Hayes, J. Shin, R. M. Caschera, F.and Murray, and V. Noireaux. Protocols for implementing an *escherichia coli* based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments (JoVE)*, (e50762), 2013.

[63] Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, and R. M. Murray. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synthetic Biology*, 3(387–397), 2014.

[64] M. K. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. J. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, and J. B. Lucks. Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synthetic Biology*, 4:503–515, 2014.

[65] H. Terashima, C. Tatsumi, A. Kawamoto, K. Namba, T. Minamino, and K. Imada. In vitro autonomous construction of the flagellar axial structure in inverted membrane vesicles. *Biomolecules*, 10(1):126, 2020.

[66] Z. A. Tuza, V. Singhal, J. Kim, and R. M. Murray. An in silico modelling toolbox for rapid prototyping of circuits in a biomolecular "breadboard" system. In *Proc. IEEE Control and Decision Conference*, pages 1401–1410, 2013.

[67] A. Vaish, S. Guo, R. M. Murray, P. J. Grandsard, and Q. Chen. On-chip membrane protein cell-free expression enables development of a direct binding assay: A curious case of potassium channel KcsA-Kv1.3. *Analytical Biochemistry*, 556:70–77, 2018.

[68] B. Wang, R. I. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2:508, 2011.

[69] W. R. Whitaker, S. A. Davis, A. P. Arkin, and J. E. Dueber. Engineering robust control of two-component system phosphotransfer using modular scaffolds. *Proceedings of the National Academy of Sciences*, 109(44):18090–18095, 2012.

[70] Y. Y. Wu, S. Culler, J. Khandurina, S. Van Dien, and R. M. Murray. Prototyping 1,4-butanediol biosynthesis pathway in a cell-free transcription-translation system. In *Synthetic Biology: Engineering, Evolution, and Design (SEED)*, 2015. Available at `biorxiv.org/content/early/2015/04/09/017814`.

[71] E. Yeung, J. Dy, K. B. Martin, A. H. Ng, D. Del Vecchio, J. L. Beck, J. J. Collins, and R. M. Murray. The effect of compositional context in synthetic gene networks: Biophysical constraints arising from relative gene orientation. *Cell Systems*, 5(1):11–24.e12, 2017.

[72] Enoch Yeung, Jongmin Kim, Jorge Gonçalves, and Richard M Murray. Global network identification from reconstructed dynamical structure subnetworks: Applications to biochemical reaction networks. In *Proc. IEEE Control and Decision Conference*, pages 881–888. IEEE, 2015.

[73] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, 32(1):170–173, 2011.

[74] T. Zhou. Prototyping a valinomycin biosynthesis pathway within a cell-free transcription-translation (TX-TL) system. Technical report, bioRxiv, 2016. Available at `biorxiv.org/content/early/2016/12/04/091520`.

Biographical Sketches

Data management plan

All research data generated as a result of this proposal will be made available to the scientific community in a timely and responsible manner and with as few restrictions as possible. This will occur no later than the release through publication of the main findings in conference proceedings or journals. Where appropriate, metadata (contextual information and documentation) will accompany the data to provide the necessary details for a secondary user to interrogate the data without the possibility of misinterpretation or confusion.

Caltech maintains a secure institutional repository that provides access to material that is of long-term value to the academic community, including datasets. The Caltech Collection of Open Digital Archives (CODA) `<http://coda.caltech.edu>` makes use of the open source EPrints software that is installed on three high-availability Linux servers managed by the Library and utilizing fiber-attached SAN (storage area network). CODA metadata conform to international standards. Documents and metadata are backed up to disk and tape, with off-site tape storage in case of disaster. All records are given Persistent URLs that do not change with system upgrades. Caltech CODA is indexed by all major Internet search engines. All publications (articles, theses, technical reports, conference papers), datasets used for publications, and metadata arising from the project will be archived in Caltech CODA.

Data generated at Imperial will use similar procedures to insure availability of the results from this research. Protocols will be shared publicly though publication and on the applicant's websites. Experimental data (microscopy imaging, EM imaging) will be generated in appropriate proprietary format for that instrument. The formats and software used will enable sharing and the long-term validity of the data. Raw data (enzyme kinetics, microscopy, VIB data) will be processed and stored on primary instrumentation computers as well as back-up hard drives. Metadata associated with other experiments (enzyme kinetics, microscopy and all other data sources) will be recorded as required electronically and written in laboratory books. Access to data can be requested by contacting the PI (or Co-I depending on the data).

Software developed at Caltech for the analysis of data will be licensed under BSD or GNU open source licenses and will be made available to interested researchers through GitHub or similar source code repositories.

The lead PI for the project (Murray) will be responsible for insuring that all datasets, publications, and software arising from this research are managed according to this data management plan. In the event of the departure of one or more of the PIs, Caltech will maintain access to all publications and datasets stored in CODA. Source code will be left in public repositories independent of the status or affiliation of the PIs.

Facilities, Equipment, and Other Resources

Laboratory The Murray laboratory possesses all resources for design, implementation, and testing of synthetic circuits in engineered bacteria. Laboratory facility of approximately 1800 sq. feet are available in the Keck Laboratory for Engineering Sciences. The laboratory contains equipment for microscopy (stereoscope, fluorescence microscope), molecular biology (multimode plate readers, standard thermocycler, qPCR thermocycler, gel documentation system), and bacterial culture (shakers, incubator).

Computing For simulation and computational analysis, we will make sure of a high performance computing cluster maintained by Caltech's Information Management Systems and Services (IMSS). This facility provides access to large numbers of CPU and GPU cores, at rates that are considerably lower than commercial cloud computing providers.

Office Office space is available for all researchers, including visiting researchers, who will work on this project. Conference rooms and meeting rooms are available for shared building usage.