Engineering scalable biological systems

Timothy K. Lu

1Synthetic Biology Group; Research Lab of Electronics; Department of Electrical Engineering and Computer Science; Massachusetts Institute of Technology; and 2Broad Institute of MIT and Harvard; Cambridge, MA USA

Synthetic biology is focused on engineering biological organisms to study natural systems and to provide new solutions for pressing medical, industrial and environmental problems. At the core of engineered organisms are synthetic biological circuits that execute the tasks of sensing inputs, processing logic and performing output functions. In the last decade, significant progress has been made in developing basic designs for a wide range of biological circuits in bacteria, yeast and mammalian systems. However, significant challenges in the construction, probing, modulation and debugging of synthetic biological systems must be addressed in order to achieve scalable higher-complexity biological circuits. Furthermore, concomitant efforts to evaluate the safety and bioccontainment of engineered organisms and address public and regulatory concerns will be necessary to ensure that technological advances are translated into real-world solutions.

In the last century, scientists have made giant strides in identifying and studying biological parts such as proteins and nucleic acids, understanding regulatory networks, and constructing engineered organisms using the ever-advancing tools of genetic engineering. In the last decade, synthetic biologists have leveraged the power of modern molecular biology using frameworks translated from traditional disciplines such as electrical engineering, computer science, mechanical engineering and chemical engineering to create a wide range of synthetic biological circuits, including switches, oscillators, digital logic gates, filters and modular and interoperable memory devices, counters, sensors and protein scaffolds. Using these circuits, biological engineers have created synthetic organisms that can be used for bioremediation, biosensing, computation, bioenergy and medical therapeutics (reviewed in ref. 31–33). Despite these advances, the realization of synthetic-biology-based applications will require future breakthroughs in our ability to create sufficiently complex and reliable biological systems. Here, I will discuss current limitations and potential solutions for the construction, probing, modulation and debugging of scalable biological systems as well as hurdles for the deployment of engineered organisms from bacteria to mammalian cells which adds to the discussion of next-generation synthetic gene networks in reference 31 (Fig. 1).

Physical Construction of Scalable Biological Systems

Construction of early synthetic circuits largely relied on restriction enzymes and polymerase chain reaction (PCR)-based techniques to assemble existing genetic components. These methods do not scale well with increasing complexity due to a lack of sufficient unique restriction sites and the need to have physical DNA templates from which to amplify genetic parts. Standards for library construction and the assembly of parts libraries have been integral in circumventing this dependency on templates and restriction sites. However, since these parts must be devoid of restriction sites used in the defined standards and should ideally be optimized for use in one’s organism of choice, the use of whole-gene DNA synthesis is on.
COMMENTARY

what DNA to write (Fig. 2). Just as the decoding of the human genome sequence did not immediately reveal the functions of all human genes, the utility of high-throughput DNA synthesis technology will only gradually become evident as synthetic biologists learn how to create complex systems. For example, future synthesized circuits should be designed with ease of probing, modulating and debugging in mind. These features could be implemented by including validated RNA “handles” that can be easily measured with standard probe sets to determine internal RNA concentrations, gene circuits that allow inducers to modulate synthetic circuit protein levels, and properly situated restriction sites for the rapid cloning of components that need systematic optimization, such as ribosome binding sequences.

Significant advances in well-characterized, interoperable devices are necessary for the construction of higher-order modules that will enable scalable biological systems. These trends are similar to Moore’s law for integrated circuits and suggest that there is substantial room for growth in the field of synthetic circuits.

Using direct chemical synthesis, circuits can be designed in silico and implemented in DNA with significantly less effort from researchers. As DNA synthesis becomes increasingly economical and efficient, it will become possible to construct complex systems with less reliance on restriction enzymes. For example, DNA synthesis productivity has exceeded 1 Mbp per person per day while Venter and colleagues recently succeeded in synthesizing a 1.08 Mbp genome. However, most synthetic gene circuits to date have not exceeded the 50 Kbp level, indicating that there is a large gap between our ability to read and write DNA and knowing the rise. Using direct chemical synthesis, circuits can be designed in silico and implemented in DNA with significantly less effort from researchers. As DNA synthesis becomes increasingly economical and efficient, it will become possible to construct complex systems with less reliance on restriction enzymes. For example, DNA synthesis productivity has exceeded 1 Mbp per person per day while Venter and colleagues recently succeeded in synthesizing a 1.08 Mbp genome. However, most synthetic gene circuits to date have not exceeded the 50 Kbp level, indicating that there is a large gap between our ability to read and write DNA and knowing
development and characterization of compatible biological parts. Specificity in biological systems largely relies on spatial distribution and chemical interactions. This is in stark contrast to electrical engineering, where specificity is achieved through direct electrical wiring. Thus, strategies for achieving inter-part compatibility include targeting circuits to isolated compartments, mutagenesis and directed evolution of existing parts, and using comparative genomics to identify, synthesize, and test homologous proteins or nucleic acids. These efforts may be complicated by unknown global factors (e.g., growth rates, endogenous transcription factors with off-target effects on synthetic circuits, protein-protein interactions, small RNAs) that can confound device testing and render it difficult to use pre-defined parts in a wide range of organisms and environmental conditions without additional alterations and characterization. Therefore, combinatorial methods to test single-component performance, multi-component interactions and biological crosstalk (e.g., cross-activation or cross-repression of transcription, non-specific enzymatic activity, inappropriate triggering of signalling pathways) will be important for parts libraries (Fig. 3). These results should be incorporated into mathematical models to aid future model-based design. Indeed, institutions such as BIOFAB are attempting to systematically assemble and characterize libraries of synthetic devices. However, development efforts for certain platforms that are promising for library construction, such as zinc finger proteins and RNA interference, may be slowed by the presence of existing intellectual property.

As an example of combinatorial characterization (Fig. 3), suppose one would like to construct multiple interoperable NOR (NOT-OR) gates to constitute a universal logic system. NOR functionality can be built by placing pairwise combinations of unique operator sites for transcriptional repressors within synthetic promoters. To identify orthogonal repressors, one can encode individual transcription factors under inducible control on one set of plasmids and individual cognate operator sites driving expression of a reporter gene on another set of plasmids. Then, all possible combinations of transcription factor plasmids and reporter plasmids can be co-transformed into cells and tested for single-component performance (e.g., when a transcription factor is co-transformed with its cognate operator site) and potential crosstalk interactions (e.g., when a transcription factor is co-transformed with non-cognate operator sites). Standard induction curves can be derived by varying the concentration of transcription factors using the inducible promoters and measuring the resulting output. Based on these results, an optimal set of non-interacting transcription factors and cognate operators can be selected. To create the NOR gates, all possible pairwise combinations of operators can be constructed in synthetic promoters and co-transformed into cells with all pairwise combinations of transcription factors under independent inducible control. Proper NOR gate functionality and crosstalk can then be determined in a high-throughput fashion for all potential gates by varying inducer levels and measuring reporter gene output. In addition to enabling interoperable gate selection, large-scale experiments such as these should yield substantial data for models that can predict the orthogonality of transcription factors and operators for future circuits (e.g., using heuristic or thermodynamically guided algorithms). Moreover, matrices of cross-repression interactions can be constructed and incorporated into
transcriptional models when cross-interacting transcription factors must be used in other systems.

Model-guided design is crucial for the construction of complicated electrical and mechanical systems. Time-based simulations for electrical and mechanical systems are possible since mathematical models are established and parameters are well known. In contrast, most parameters in biological circuits are unknown and the computational resources required to accurately simulate noise and multiple component interactions are significant. Recent advances in modelling chemical networks, transcription, translation and biological noise using the inherent physics of solid-state electronic devices should enable the construction of large-scale real-time electronic models of synthetic biological systems. Other techniques from control theory such as small-signal linearization and modularization enable tractable modelling and simulations prior to implementation. Biological systems exhibit nonlinearity (e.g., cooperativity) which can be linearized in different regions of operation (Fig. 4A). Frequency-domain analysis in linearized systems allows for block modelling and deeper understanding of system dynamics, such as noise, stability, time constants and performance (Fig. 4B). Small-signal linearization and frequency-domain modelling have not been extensively used for studying and designing synthetic biological circuits even though advances in microfluidics and time-lapse microscopy can now achieve frequency modulation of inputs and long time-scale data collection necessary for frequency-domain analysis. Furthermore, microfluidics devices can be coupled with electronic controllers to stabilize and alter the dynamics of synthetic biological circuits similar to electronic controllers that are used to control mechanical systems. The insights that can be gained from linearized block models of complex systems can complement the accuracy of time-domain state-space representation, time-based mathematical simulations and non-linear control theory. To enable successful time-based and frequency-based modelling of biological systems, accurate parameters will need to be derived by high-throughput in vitro

Figure 4. Control theory techniques for modelling synthetic biological circuits. (A) Small-signal linearization of biological components in different regions of operation enables the development of linear models. (B) Linearization can enable frequency-domain analysis, systems modelling using block diagrams and deeper insights into system dynamics. For example, transcription and translation can be understood as low-pass filters and block diagrams can be drawn for simple negative-feedback loops to yield understanding into system interconnections and responses to different input types. In the block diagram shown, $s$ refers to $j\omega$, where $j$ is $\sqrt{-1}$ and $\omega$ is angular frequency.
and in vivo probes and microarrays, as described below.

Molecular Probes and Modulators for Scalable Biological Systems

High-throughput methods for probing multiple nodes, such as protein and RNA levels, in complex biological circuits are necessary to achieve reliable and scalable performance. Ideal probe features include high signal-to-noise ratios, specificity, low cost, multiplexability, non-invasiveness and the ability to reveal real-time dynamics. Complementary techniques for the reliable and multiplexed modulation of synthetic circuits are also needed to generate perturbations to obtain system parameters, drive accurate models and monitor system performance.

Analyses of most synthetic circuits rely on fluorescent proteins, which are well-characterized, easy to detect and can be multiplexed, as reporters of combined transcriptional and translational activity. However, fluorescent proteins are large, require folding and maturation and can be too stable to track rapid dynamics without additional modifications. To address these issues, methods have recently been developed to create fast-folding fluorescent proteins, degradation tags and in vivo ligation of fluorophores to proteins. Alternative methods for monitoring protein levels in synthetic circuits include luminescence or colorimetric assays although these methods can be more difficult to multiplex. Advances in high-throughput proteomic characterization may eventually enable global monitoring of protein levels across time without the need for multiplexed reporter proteins. Nucleic-acid aptamer beacons may also play a role in protein detection but have not been extensively applied to in vivo settings and require selection protocols since there are no algorithms that can reliably predict the affinity of binding between aptamers and proteins of interest. Thus, new techniques for multiplexed and global protein detection are needed for scalable biological system design.

Transcriptional activity can be monitored by quantifying RNA in engineered cells, which is easier to do in a multiplexed and global fashion than quantifying protein levels. Multiplexed RNA levels can be assayed with qRT-PCR or microarrays but these techniques require nucleic acid extraction and cellular destruction. Fusing RNAs of interest to aptamers that bind fluorescent proteins or fluorescent small molecules can be used to monitor in vivo RNA levels but requires modifying RNAs of interest. Side-by-side fluorescence resonance energy transfer (FRET) probes, quenched autoligation probes and molecular beacons are sequence-specific tools for detecting RNA levels in vivo with multiplexing capabilities but face challenges such as intracellular delivery, compartmentalization and degradation. Nonetheless, nucleic-acid-based probes are likely to be early enablers for high-throughput in vivo monitoring of transcription in synthetic biological circuits given that they are relatively easy to design for different targets. Thus, methods for achieving efficient delivery of nucleic-acid-based probes are crucially needed to make them widely useful to synthetic biologists for monitoring in vivo RNA levels in real time.

The majority of modulators used in synthetic circuits today are small molecules that induce or repress the activity of existing transcriptional regulators. As increasing number of transcriptional regulators are identified or created, it will be important to identify corresponding inducers that can control in vivo function, characterize crosstalk between different inducers and design interoperable modulators. For example, direct evolution of the AraC transcriptional regulator was necessary to increase compatibility between arabinose-controlled and IPTG-controlled systems. Other modulators for synthetic circuits include aptamers that respond to small molecules and trigger translational activities. These aptamers can be discovered using techniques such as Systematic Evolution of Ligands by Exponential Enrichment (SELEX). Light-inducible proteins provide rapid and well-controlled signal transduction that could potentially be enhanced by developing variants which preferentially respond to different wavelengths of light. An underutilized technology in synthetic biology is RNA interference (RNAi), which allows straightforward design of sequence-specific modulators. However, RNAi is not available in many organisms of interest and must be coupled with efficient delivery mechanisms for nucleic acids. Fortunately, RNAi has been reconstituted in well-studied organisms such as Saccharomyces cerevisiae and significant advances are being made in RNAi-delivery technology.

Debugging Malfunctioning Biological Systems

The harsh reality of engineering synthetic biological circuits is that most designs fail to perform as expected. Debugging of synthetic circuits is a painstakingly iterative process involving the detailed characterization of all available internal nodes, trial-and-error modification of constituent components and modelling. However, debugging will be greatly facilitated by improved techniques for sequencing, probing and modulating synthetic circuits, as described above. For example, an initial debugging step is usually to fully sequence all circuit components to ensure that the intended genetic program has been constructed. Subsequent debugging can include dynamically probing all possible RNA and protein nodes and interactions using techniques such as qRT-PCR, fluorescent protein fusions and circuit-specific methods (e.g., in vitro gel shift assays, protein phosphorylation assays, single-cell microscopy, and so forth). Based on these results, it can often be determined whether system failure is due to inherent circuit topology or due to poor component performance leading to circuit operation in non-functional parameter regions. In the former case, redesigns are needed and can benefit from detailed modelling using insights and measurements from the failed attempts. In the latter case, systematic and randomized mutagenesis and screening techniques are helpful to optimize system performance. For example, synthetic circuits often do not work properly because of mismatched expression levels resulting in outputs that do not fit into the dynamic range of other inputs. Mismatches in dynamic range can usually be corrected by systematically altering the strengths of promoters.
and ribosome binding sequences using mutagenesis. 57,58

In addition, optimization of synthetic genes can help resolve problems associated with the expression of heterologous genes in foreign hosts including DNA instability and poor translational efficiency. 59 Techniques for the high-throughput alteration of genetic circuits should also help speed up debugging cycles by enabling testing across larger parameter spaces. 60 Modelling can generate insights into potential failure points in synthetic circuits but is inadequate at identifying problems with components or parameters that are not included in the models themselves. For example, global effects of host factors can change the performance of synthetic circuits. 41 To model these effects, high-throughput data for global transcriptional responses to defined perturbations should be obtained from large promoter libraries and incorporated into whole-cell simulations. 61 Moreover, test platforms based on well-characterized host organisms or minimal organisms may be ideal backgrounds for validating system designs prior to deployment in final vehicles. 31,62 Finally, given the significant time and effort that is spent on fixing biological systems and the wealth of information that can be gleaned from malfunctioning designs, it would be beneficial to the general synthetic biology community to establish repositories where results from debugging cycles and system failures can be disseminated.

Public and Regulatory Considerations for Engineered Biological Systems

In addition to technological advances, efforts to address public and regulatory concerns over synthetic biology are necessary to ensure the successful translation of scalable biological systems to real-world applications. The synthetic biology community has recognized that safety, security and ethical issues must be addressed in an open and earnest fashion. 63,64 Although there is little data to suggest that genetic engineering or synthetic biology have produced harmful constructs over the last few decades, concerns from the public and the media are often voiced when significant technological milestones in synthetic biology are achieved. Some of these concerns may be rooted in inadequate scientific understanding or communication of the current state of synthetic biology. Thus, enhanced outreach and education efforts between researchers and the public are necessary to ensure the field’s continued progress. Furthermore, global variations in cultures and attitudes towards engineered life should be recognized and addressed by researchers and advocates. For example, differences in opinion between the United States and the European Union regarding the use of Genetically Modified Organisms (GMOs) as food products may translate to synthetic organisms. Thus, it is imperative that synthetic biologists act and communicate their roles as critics of the field to ensure that safety, reliability and the realization of broad social benefits are maintained as the utmost research priorities.

The regulatory issues associated with deploying synthetic biological organisms vary greatly depending on application. The areas associated with the lowest regulatory hurdles include metabolic engineering and bioenergy where biocontainment is straightforward to achieve, direct contact with humans is limited and ultimate products are chemical compounds produced by engineered cells. For environmental deployment of synthetic organisms in the United States, the Environmental Protection Agency and the Food and Drug Administration can become involved. Regulatory issues for environmental applications are similar to those involved with GMOs, such as the use of recombinant DNA, biocontainment and impact on natural ecosystems. Standardized methods and technologies must be developed with input from regulatory agencies to evaluate these concerns. The most difficult regulatory hurdles pertain to the use of synthetic biological systems for human therapeutics such as adoptive immunotherapy, 45 cancer-seeking bacteria, 46 engineered phage targeting bacterial biofilms 57 and antibiotic-resistant organisms. 58 These hurdles include immunogenicity, biocontainment, recombinant DNA and manufacturing purity. Stem-cell-based therapies face similar hurdles and therefore it will be informative to follow the path of stem cells into the clinic. 48 Reliable methods to eliminate engineered organisms as needed, 70 stringent techniques to evaluate safety in animals and humans, techniques for encapsulating and isolating engineered cells, 71 assays for mutation rates in deployed organisms, utilization of probiotic strains as substrates for engineering microbe-based therapeutics and close communication with regulatory agencies will be necessary to translate synthetic biological organisms into the important frontier of human treatments. Furthermore, synthetic biologists interested in human therapeutics should focus their efforts on areas of significant unmet need to optimize risk-benefit tradeoffs and should be open to proof-of-concept applications in areas with lower regulatory hurdles such as environmental or veterinary use.

Conclusions

The past decade has delivered significant advances in the design and construction of basic synthetic circuits. In the upcoming decade, novel technologies for composing, probing, modulating and debugging scalable biological circuits will enable the robust performance of useful tasks by engineered organisms. Scientific advancements must be accompanied by concomitant efforts to address societal and regulatory concerns over synthetic biology. These endeavors should yield exciting new solutions for real-world problems in critical areas of medical, industrial, environmental and energy applications.

Acknowledgements

I appreciate the insights that the editor and anonymous reviewers have provided. I would also like to thank Michael Koeris and Ahmad Khalil for their critical reading of this manuscript.

References


