AI Challenges in Synthetic Biology Engineering

Fusun Yaman, Aaron Adler, Jacob Beal
fusun.yaman@raytheon.com, aaronadler@alum.mit.edu, jakebeal@ieee.org
BBN Technologies, 10 Moulton Street, Cambridge, MA 02138

Abstract

A wide variety of Artificial Intelligence (AI) techniques, from expert systems to machine learning to robotics, are needed in the field of synthetic biology. This paper describes the design-build-test engineering cycle and lists some challenges in which AI can help.

Introduction

Engineering the behavior of cells by modification of their genetic machinery holds the potential for revolutionary advances in many important application areas, including medical therapies, vaccination, manufacturing of proteins and other organic compounds, and environmental remediation. As capabilities and potential applications grow, the complexity and cross-disciplinary knowledge required to employ them is also growing rapidly. Managing the complexity of biological engineering is thus a problem of increasing importance. The rapid pace of advancement makes it important to have good methods for integration of new knowledge and procedures into organism engineering workflows.

Synthetic biology, the systematic design and engineering of biological systems, is not the first field facing the challenge of managing engineering complexity. In areas as disparate as the engineering of software, electronics, and mechanical systems, the common response has been to use artificial intelligence (AI) techniques to capture human expert knowledge and embed it into assistive tools. (The AI foundation of successful work may be forgotten, e.g., programming languages are the distilled result of AI research into "automatic programming.") This paper refers to AI in the sense of knowledge-based computing systems, including the representation of knowledge (e.g., semantic networks, frame representations), the acquisition of knowledge (e.g., machine learning, hypothesis generation), its employment in planning and decision making (e.g., expert systems, constraint-based reasoning, planning under uncertainty), and also in automated action (e.g., robotics).

Although biological organisms are complex and not entirely understood, there are many opportunities for AI techniques to make a major difference in the efficacy of organism engineering (an impact that has already begun). As it has in other fields, AI is likely to have transformative impact not through the construction of a single dramatic system, but through the deployment of a wide variety of AI-based tools.

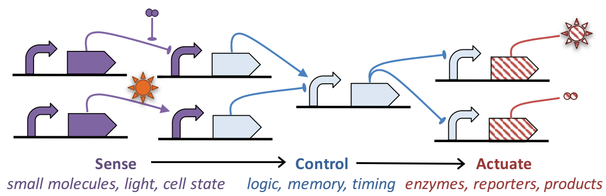

Figure 1: Abstract network illustrating sense-control-actuate paradigm for synthetic genetic networks.

Organism Engineering

Rather than attempting to cover the whole field, we will narrow the focus of discussion for this paper to one important and widely addressed class of synthetic biology systems—genetic regulatory networks. Many synthetic biology systems can be viewed as three components: 1) sensing of environmental or cell state, 2) a control system that processes the signals from these sensors to determine appropriate cellular behavior, and 3) actuators that convert the control signals into actions such as enzymatic pathway regulation or reporter expression (Figure 1).

A typical synthetic biology workflow for organism engineering may be viewed as a cycle of three stages: *design* maps a behavior specification to a nucleic acid sequence intended to realize this behavior; *build* draws on synthesis and/or assembly protocols to fabricate said nucleic acid sequence; and *test* assays (measures) the behavior of cells modified to include the sequence, feeding this information back into the design step completing the cycle (Figure 2).

We discuss each of these steps emphasizing opportunities for improvements in the typical current workflow.

Design At the most abstract level, the engineer must determine the arrangement of sensors, actuators, regulatory relationships, and/or enzymatic pathways that will be used to implement a desired behavior. An arrangement is then mapped onto the set of DNA or RNA components that are available, or new components are engineered with the desired specifications while ensuring that there are not con-

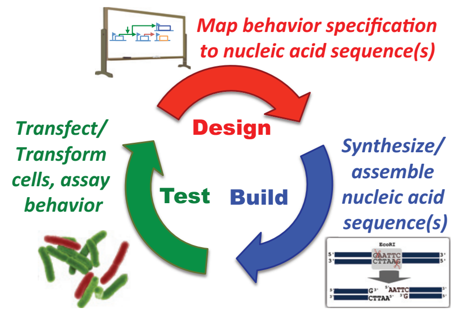

Figure 2: Typical synthetic biology workflow.

flicts between the components selected in the arrangement. Finally, the components in the arrangement must be linearized, i.e., an order must be determined for genes to appear in the DNA sequence. At present, the selection and arrangement of components is typically carried out largely by hand, with little usable characterization data (component measurements under specific conditions) to guide component selection and poor models to quantitatively predict the behavior of the resultant composite system.

Build The build stage creates organisms modified with the designed nucleic sequence(s). First, the sequence(s) are synthesized (created) or assembled to produce actual physical samples, and the host organisms are cultured (grown) to be ready to receive these sequences. The sequences are then delivered to the organism by one of a variety of protocols. Both of these stages have a number of issues in yield and quality assurance. Many protocols require a "magic touch" by which some practitioners get reliable results and others frequently build systems with problematic flaws. Next-generation sequencing may help to address issues of quality control, but planning, resourcing, and executing build protocols effectively is still an open and challenging problem.

Test Finally, the behavior of the newly constructed organism or organisms is assayed (measured) to determine how well it corresponds with the original specification, and to help debug misbehavior such that the next iteration of the design can be closer to the desired behavior. Here, one of the biggest challenges is in relating assay data to the original specification: many assays produce data in great volume, but the mapping back to the original specification is often qualitative or relative. Likewise, it is often not clear how to relate the observed behavior to predictive models that can provide principled guidance in how to adjust the design phase in order to produce improved results.

Potential AI Contributions

Currently most organism engineering workflows have little automation and rely heavily on domain expertise, only some of which is shared in publications. Tools that support or carry out information integration and informed decision making can improve the efficiency and speed of organism engineering, and enable better results. A number of these points are summarized in Table 1.

Bioengineering Challenge	Key AI Techniques
Machine-assisted gene circuit design	expert systems, constraint-based reasoning, heuristic search, optimization, machine learning, multi-agent systems
Flexible protocol automation	robotics, planning under uncertainty
Assay interpretation and modeling	machine learning, qualitative reasoning
Represent/exchange designs	semantic networks, ontologies
Represent/exchange protocols	semantic networks, schemas

Table 1: Summary of bioengineering challenges for which there is a high potential for AI techniques to contribute to the solution.

Other Challenges

A number of non-technical challenges need to be addressed in pursuit of the synthesis of AI and synthetic biology.

- Much expert knowledge is not explicitly written down or relies on a human reader to make "common sense" assumptions. Other expertise may be transmitted by apprenticeship, particularly for complex physical processes. Capturing such knowledge requires investment and cooperation from experts in synthetic biology and AI.
- Many aspects of organism engineering and an organization's engineering workflow may be proprietary or subject to intellectual property claims, or use closed systems that do not easily integrate with automation. In computer science, these types of barriers have been mitigated by strong movements in the scientific and business communities that promote open exchange of knowledge.
- Biological organisms are complex, and many critical pieces of information are still unknown. While this is potentially a serious limitation in some areas, recent results in improving the modeling and predictability of composition in synthetic biology systems, give evidence that some areas of organism engineering are mature enough to support the application of AI techniques. AI techniques can also reveal areas of biology that need further study.
- The continuing rapid advancement in both knowledge and methods may rapidly render specific AI-enabled methods obsolete. Also, the adoption barriers are likely large as most labs have complex and highly customized processes in place. Impactful AI applications should deliver large enough benefits to overcome adoption cost and need to focus on providing somewhat more general frameworks for the rapid capture and automation of methods.

Summary and Recommendations

From an AI perspective, there are many interesting problems for applications, particularly given the massive scope and complexity of biological organisms and the problems encountered in their engineering. Complementarily, from a biology perspective, there are many potentially large benefits from integration of AI techniques. Realizing these benefits is likely to require tight collaboration between practitioners of both disciplines. We thus strongly recommend that practitioners interested in realizing these benefits seek out complementary colleagues.