

How can AI help Synthetic Biology?

AI Topics

Presenter: Fusun Yaman or Aaron Adler from BBN Technologies. The synthetic biology work highlighted in this talk was done by the presenters, Dr. Jacob Beal at BBN, and collaborators at MIT and BU. Contributions from others in the field will also be highlighted as referenced in the citations below.

Keywords: Applications, Multidisciplinary Topics

Highlighted Papers: (Yaman et al. 2012; Beal, Lu, and Weiss 2011; Hillson, Rosengarten, and Keasling 2012)

Introduction

Our primary goal in this talk is to draw the attention of the AI community to a novel and rich application domain, namely Synthetic Biology. Synthetic biology is the systematic design and engineering of biological systems. Synthetic organisms are currently designed at the DNA level, which limits the complexity of the systems. In our talk we will introduce the domain, describe the current workflow used by synthetic biologists, and demonstrate the feasibility of progress in this domain. Problems specific to each AI topic area will be highlighted.

Over the past decade, biologists have begun to establish engineering control over the genetic machinery of cells. Biologists identified and created many DNA sequences which can be used as building blocks of novel cellular programs. These programs are then injected into cells and executed. Synthetic biology holds the potential for revolutionary advances in medicine, environmental remediation, and many more. For example, some synthetic biologists are trying to develop cellular programs that will identify and kill cancer cells (Xie et al. 2011), while others are trying to design plants that will extract harmful pollutants like arsenic from the ground (Antunes et al. 2011). However, the field has reached a complexity barrier that AI researchers can help it overcome.

The state-of-the-art techniques in synthetic biology require practitioners to design organisms at the DNA level. This low-level, manual process becomes unmanageable as the size of design grows. This is analogous to writing a computer program in assembly language which also becomes difficult quickly as the size of the program grows.

In this talk, we will first review the state-of-the-art in synthetic biology, to explain the context in which these artificial intelligence problems arise – discussing how individual biological elements are composed together and the current typical workflow by which synthetic biology engineering is carried out. Then we will highlight the AI problems that we and others have made progress on as well as open problems in the field. Finally we will discuss how AI researchers can get involved through:

- conferences, e.g., SBx.0, IWBDA;
- journals, e.g., Nature, ACS SynBio, PLoS ONE;
- contests, e.g., iGem;
- online tools, e.g., Clotho, j5, and TASBE tools.

Currently the synthetic biology engineering workflow is mostly manual and relies heavily on domain expertise, a limited amount of which is shared through publications. There are several points in the workflow where informed decision making would improve the efficiency and reduce the time to engineer an artificial genetic circuit. Such decision points are where the biologist can use the help of AI researchers. In addition to the areas highlighted below, we will also discuss how AI researchers can help synthetic biologist with reasoning, multiagent systems, and robotics.

Constraints and Satisfiability

In a top-down design approach practitioners design organisms using higher level descriptions (as in Figure 1 (top)). These descriptions will be mapped to a composition of primitive motifs, producing an *abstract genetic regulatory network* (AGRN): one which defines relationships between parts, but leaves the actual identities of those parts unspecified (Figure 1 (middle)). To realize this network, one must solve the part selection problem: mapping abstract features to a collection of particular standardized biological parts that preserve the relationships between features prescribed by the network. Prior work has demonstrated the design of AGRNs from high-level programs (Beal, Lu, and Weiss 2011) and automated assembly of DNA sequences from standardized biological parts such as BioBricks (Densmore et al. 2010; Hillson, Rosengarten, and Keasling 2012). MatchMaker (Yaman et al. 2012) fills a critical gap in the actual selection of particular biological parts to implement the design. We demonstrated in MatchMaker that transforming high-level organism descriptions to DNA sequences involves solving several constraint satisfaction and optimization problems. The key steps in this transformation are: 1) *Feature Matching*: finding compatible features (functional DNA elements) that have the same regulatory relationship as defined in the AGRN, 2) *Signal Matching*: choosing a specific set of parts within the family of required features such that the chemical concentration levels produced are compatible with each

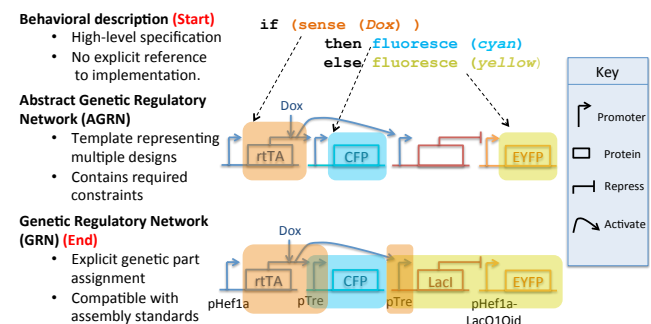


Figure 1: (Top) High-level design describing organism-level behavior which can be compiled to Abstract Genetic Regulatory Networks (AGRN). (Middle) Design at the AGRN level, none of the parts are assigned to DNA sequences. (Bottom) Design at the GRN level, fully specified and suitable for assembly.

other to ensure robust system performance, and 3) *Part Matching*: finding standard composable parts that when put together result in the required function with the desired performance. Each step is NP-Complete.

Machine Learning

There are a wide range of applications in which it would be useful to have a small synthetic biology circuit that could reliably classify cell state. For example, in (Xie et al. 2011), the authors propose a cancer therapy based on a circuit that uses miRNA (micro RNA) markers to test whether a cell belongs to a particular type of cancer and then kills only those cells. The authors then demonstrate an miRNA classifier that can distinguish between HeLa cells (cervical cancer cell line) and several other cell lines.

This is a perfect opportunity to apply machine learning techniques to select the relevant features and use supervised learning algorithms. However, there is a catch. Not every classifier/algorithm can be implemented in a cell because cells don't have the same computational machinery as computers. For example, the classifier in (Xie et al. 2011) is just a boolean formula. (Beal and Yaman 2012) have developed an information-based technique for selecting the markers in an effort to create classifiers as described in (Xie et al. 2011). Classifiers with auto-selected features are shown in simulation to be as effective as the biologist-created classifier.

Knowledge-Based Systems

Our BioCompiler (Beal, Lu, and Weiss 2011) uses design motifs to capture well known design constructs, such as NOT, AND, OR like primitives. These motifs can be considered as AGRN fragments with well defined composition hooks. Given a high level description using these motifs, the BioCompiler composes them into an AGRN and optimizes using standard compiler techniques, such as dead code elimination and copy-propagation.

Knowledge Representation and Reasoning

The synthetic biology community has been working on developing standards that can capture details about system designs. For example, the Synthetic Biology Open Language (SBOL) group (SBOL 2013) has been developing a data exchange standard for describing genetic parts, devices, modules and systems. This work also encompasses efforts to model and visually represent these designs.

Reasoning Under Uncertainty

Many biological processes are inherently stochastic and there are many sources of noise in cellular systems, however accurate enough models are needed so that synthetic biology can be an engineering discipline. A cornerstone of synthetic biology is that predictive design is possible – putting together two parts whose behavior is known will have a predictable outcome. Towards this end, we (Beal et al. 2012) and others (e.g., (Canton, Labno, and Endy 2008; Centre for Synthetic Biology and Innovation 2012)) have invested significant effort to characterize parts and make predictions about composite behavior taking uncertainty into

account. Better characterizing the uncertainty will enable tools to accurately reason and make predictions about the system.

Heuristic Search and Optimization

In a wet-lab equipment and resources are constrained and present opportunity for heuristic search. Most of the experiments that are run concurrently might benefit from sharing intermediate products. For example, if two circuits that are being built contain a shared DNA substring, we can coordinate the assembly steps so that this shared substring is produced only once. Assembly Planner (Densmore et al. 2010) for BioBrick assembly is a fine example of such an effort.

References

- Antunes, M. S.; Morey, K. J.; Smith, J. J.; Albrecht, K. D.; Bowen, T. A.; Zdunek, J. K.; Troupe, J. F.; Cuneo, M. J.; Webb, C. T.; Hellinga, H. W.; and Medford, J. I. 2011. Programmable ligand detection system in plants through a synthetic signal transduction pathway. *PLoS ONE* 6(1):e16292.
- Beal, J., and Yaman, F. 2012. Toward automated design of cell state detectors. In *4th International Workshop on Bio-Design Automation*.
- Beal, J.; Weiss, R.; Yaman, F.; Davidsohn, N.; and Adler, A. 2012. A method for fast, high-precision characterization of synthetic biology devices. Technical Report MIT-CSAIL-TR-2012-008, MIT. Technical Report: MIT-CSAIL-TR-2012-008 <http://hdl.handle.net/1721.1/69973>.
- Beal, J.; Lu, T.; and Weiss, R. 2011. Automatic Compilation from High-Level Biologically-Oriented Programming Language to Genetic Regulatory Networks. *PLoS ONE* 6(8):e22490+.
- Canton, B.; Labno, A.; and Endy, D. 2008. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* 26:787–93.
- Centre for Synthetic Biology and Innovation. 2012. Part characterisation. <http://www3.imperial.ac.uk/syntheticbiology/parts>.
- Densmore, D.; Hsiao, T. H. C.; Kittleson, J. T.; DeLoache, W.; Batten, C.; and Anderson, J. C. 2010. Algorithms for automated DNA assembly. *Nucleic Acids Research* 38(8):2607–2616.
- Hillson, N. J.; Rosengarten, R. D.; and Keasling, J. D. 2012. j5 DNA assembly design automation software. *ACS Synthetic Biology* 1(1):14–21.
- SBOL. 2013. Synthetic Biology Open Language. <http://www.sbolstandard.org>.
- Xie, Z.; Wroblewska, L.; Prochazka, L.; Weiss, R.; and Benenson, Y. 2011. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* 333(6047):1307–1311.
- Yaman, F.; Bhatia, S.; Adler, A.; Densmore, D.; and Beal, J. 2012. Automated selection of synthetic biology parts for genetic regulatory networks. *ACS Synthetic Biology* 1(8):332–344.