

DEMO: XPLAN – Experiment Planning for Synthetic Biology

Ugur Kuter[†] and Robert P. Goldman[†] and Daniel Bryce[†] and Jacob Beal[‡]

Matthew DeHaven[†] and Christopher S. Geib[†] and Alexander F. Plotnick[†] and Tramy Nguyen[‡] and Nicholas Roehner[‡]

[†]SIFT, LLC

Minneapolis, MN, USA

{dbryce, rpgoldman, ukuter, mdehaven, aplotnick, cgeib}@sift.net

[‡]Raytheon BBN Technologies

Cambridge, MA, USA

jacob.beal@ieee.org, nicholas.roehner@raytheon.com, tramy.nguy@gmail.com *

Abstract

In the field of synthetic biology, as with other emerging fields of engineering, scientific exploration and engineering design are intimately entwined. Unlike established fields of engineering, synthetic biology has only highly uncertain and incomplete mechanistic models. As a result, engineering synthetic biological systems is an incremental process in which the production of designs is closely interleaved with execution of experiments to assess the success of those designs and data analysis to identify factors and mechanisms responsible for design successes and failures. In this demonstration, we show how a Hierarchical Task Network (HTN) planning system automates scientific discovery, closing the loop between multiple machine learning analysis and biological design tools and wet labs to guide the discovery and design process.

1 Research Problem and Motivations

Organization and planning of synthetic biology experiments is currently done almost entirely by hand. Several ongoing developments, however, are rapidly increasing the need for automation assistance in experiment planning. More and more laboratory automation is becoming available, increasing the scale and complexity of experiments that can be performed. Automation and information technology are supporting new business models with laboratory work done by technicians or outsourced to a “lab for hire.” Finally, new “multiplexing” protocols allow many tests to be conducted on a single experimental sample, and multiple experimental samples to be processed in parallel. In all of these cases, the growth in scale and complexity are rapidly outstripping the abilities of humans to create detailed experimental plans and to hand-curate the relationships between those plans and the large collections of data they generate. Furthermore, experiments are still costly both in money and time, and the space a researcher wishes to explore is often much larger than the

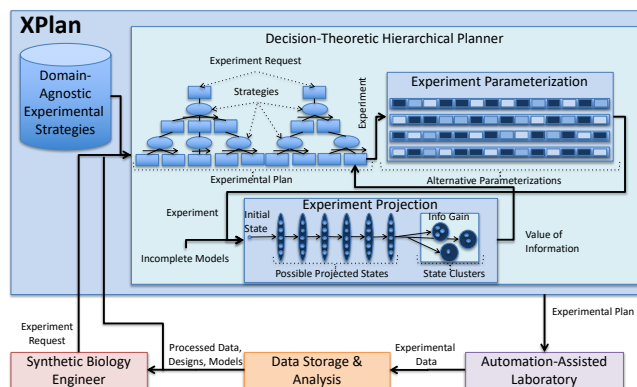


Figure 1: XPLAN combines domain-agnostic strategies and domain-specific knowledge to expand experiment requests into executable plans, which are then parameterized and projected for VOI analysis. Plans are given to laboratories to execute, producing data that results in updated models and designs and new experiment requests.

number of samples that can be tested, so there are opportunities for automation to assist in optimizing the value of information from each sample, potentially even dynamically based on partial results from an experiment in progress.

This demonstration shows how our XPLAN planner, based on Hierarchical Task Network (HTN) planning (Currie and Tate 1991; Wilkins 1988; Erol, Hendler, and Nau 1994; Nau et al. 2003) addresses these issues by providing automation support for experiment planning. Figure 1 shows a high-level architectural diagram of XPLAN. In the next section, we describe our application domain, i.e., the class of discovery and design problems addressed by XPLAN, and the challenges they pose. We then explain the technical highlights of our demonstration, which uses our SHOP2 planner (Goldman and Kuter 2018; Nau et al. 2003) and addresses these challenges.

2 Synergistic Discovery and Design (SD2)

Synthetic biology is the systematic engineering of living organisms to perform desired functions. For example, biological sensors have applications in sensing pathogens or biological, chemical, and radioactive weapons; effectors have applications in chemical synthesis and cleanup, and in targeted medical therapies. Because existing models for genetic

*This work was supported by the Air Force Research Laboratory (AFRL) and DARPA under contract FA875017CO184. This document does not contain technology or technical data controlled under either U.S. International Traffic in Arms Regulation or U.S. Export Administration Regulations. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the AFRL and DARPA.

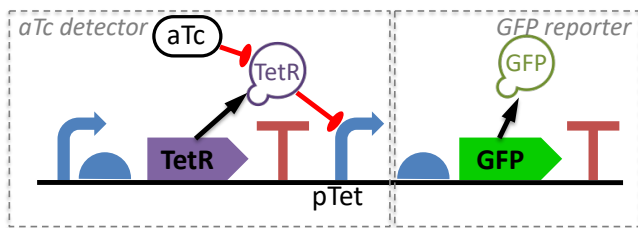


Figure 2: High-level example of SBOL representation of a biological design of GFP-producing module. GFP: Green Fluorescent Protein.

structures, assembly, and expression are still relatively weak, however, synthetic biology necessarily involves both design and experimentation to assess the success of designs and identify factors responsible for success and failure.

DARPA’s Synergistic Discovery and Design (SD2) program seeks to speed scientific and design processes through automated support for experiment planning, automated execution of experimental protocols across laboratories, and high-speed, large-scale exploratory data analysis. Figure 1 shows the high-level architecture of XPLAN, our hierarchical experiment planning system, which is a key part of the overall SD2 project. XPLAN uses HTN planning to generate experimental protocols from synthetic biologists’ expressions of experimental intent. It also translates the protocols it generates into executable forms so that they can be performed at different laboratories, which have different equipment, levels of automation, and processes.

XPLAN also helps with data analysis, by storing information about the protocol in SynBioHub (McLaughlin et al. 2018; Madsen et al. 2016), a standard synthetic biology semantic database. This enables labs performing protocols to accurately and consistently label the resulting data, immensely simplifying the process of data analysis. It also enables the operation of an automated pipeline for preliminary data formatting, labeling, and processing.

3 Demonstration Details

Design of experiments (DOE) for planning While a design is ultimately a sequence of biological constructs (e.g., DNA, RNA, other molecule/protein sequences), we assume that it is specified in high-level language that identifies genetic constructs, their function, and their relationships. Our demonstration shows how in synthetic biology problems, the designs are represented in the Synthetic Biology Open Language (SBOL) standard (Roehner et al. 2016), which captures components and their roles and interactions in hierarchically arranged modules. The Synthetic Biology Open Language (SBOL) represents both structure and function of biological designs, as shown in Figure 2. This shows an example of a system comprising two modules (dashed lines): in the left module, aTc de-represses the pTet promoter by repressing the TetR protein, which regulates the GFP (Green Fluorescent Protein)-producing right module.

Hierarchical planning for biology experiments XPLAN formulates the search for experimental plans task as hierarchi-

cal planning, more specifically, in our SHOP2 planner. In this approach we represent experimental strategies as task decomposition methods in SHOP2’s domain definitions. Experimental strategies specify a way to break-down a task as a set of sub-tasks that accomplish it. There may be several such strategies for a given task.

XPLAN’s plan library is divided into three components, broadly speaking. First is a high-level library of experimental strategies that is not specific to synthetic biology or to particular laboratories. These strategies aim to distribute experiments across laboratories for execution while minimizing variation, validating hypotheses, and determining parameters for designs during planning. Second is an abstract set of protocol components that are specific to synthetic biology, but not to particular lab configurations. Finally, there are methods that are specific to particular labs, and that enable our procedures to be translated into executable form. For example, some of these library components enable an XPLAN-generated experimental protocol to be translated to Autoprotocol. Autoprotocol, developed by Transcriptic¹, is an executable JSON schema providing a domain specific language for automated wet lab operation.

Our demonstration shows SHOP2’s SD2 plan libraries in their entirety and details how they work to generate plans for biology experiments. In particular, planning for biology experiments require explicitly reasoning about objects in the planning problem, by successively transforming initial set of objects to new ones in iterative phases during planning. This is difficult to accomplish in well-known classical planning formalisms, such as PDDL (Fox and Long 2003; McDermott 1998). Our demonstration shows how we extensively use SHOP2’s expressive HTN domain description language that allows arbitrary numeric and semantic reasoning over objects and allows creation of new objects on the fly during planning.

SynBioHub: Semantic Data Store for Synthetic Biology

SynBioHub is a repository designed for the sharing of biological constructs, experiments, and data. Designs for DNA, proteins, strains, media, etc., can be uploaded, then organized and shared with others either publicly or via a “private” link. By combining data from a variety of sources, SynBioHub also facilitates searches for information about pre-existing parts, designs, and experiments. Our demonstration features the Web Interface of SynBioHub, and shows how to access biological information related to the experiments described above.

4 Conclusions

This abstract summarizes our work on hierarchical planning to design and generate protocols for conducting synthetic biology experiments. We will demonstrate our DOE approach, hierarchical planning models and how solution protocols are generated by our planner SHOP2, and finally, our semantic data storage, SynBioHub. We will also perform a video demonstration by Transcriptic Labs, showing how XPLAN experiment plans are executed in lab setting to conduct the biology experiments.

¹<https://www.transcriptic.com/>

References

- [Currie and Tate 1991] Currie, K., and Tate, A. 1991. O-Plan: The open planning architecture. *Artificial Intelligence* 52(1):49–86.
- [Erol, Hendler, and Nau 1994] Erol, K.; Hendler, J.; and Nau, D. S. 1994. HTN planning: Complexity and expressivity. In *Proc. National Conf. on Artificial Intelligence (AAAI)*.
- [Fox and Long 2003] Fox, M., and Long, D. 2003. PDDL2.1: An extension to PDDL for expressing temporal planning domains. *Journal of Artificial Intelligence Research*.
- [Goldman and Kuter 2018] Goldman, R. P., and Kuter, U. 2018. Explicit stack search in shop2. Technical Report SIFT-TR-2018-1, SIFT, LLC, Minneapolis, MN.
- [Madsen et al. 2016] Madsen, C.; McLaughlin, J. A.; Misirli, G.; Pocock, M.; Flanagan, K.; Hallinan, J.; and Wipat, A. 2016. The SBOL Stack: A Platform for Storing, Publishing, and Sharing Synthetic Biology Designs. *ACS Synthetic Biology* 5(6):487–497.
- [McDermott 1998] McDermott, D. 1998. PDDL, the planning domain definition language. Technical report, Yale Center for Computational Vision and Control.
- [McLaughlin et al. 2018] McLaughlin, J. A.; Myers, C. J.; Zundel, Z.; Misirli, G.; Zhang, M.; Ofiteru, I. D.; Goñi-Moreno, A.; and Wipat, A. 2018. SynBioHub: A Standards-Enabled Design Repository for Synthetic Biology. *ACS Synthetic Biology* 7(2):682–688.
- [Nau et al. 2003] Nau, D.; Au, T.-C.; Ilghami, O.; Kuter, U.; McDermott, W.; Wu, D.; and Yaman, F. 2003. SHOP2: An HTN planning system. *JAIR* 20:379–404.
- [Roehner et al. 2016] Roehner, N.; Beal, J.; Clancy, K.; Bartley, B.; Misirli, G.; Grünberg, R.; Oberortner, E.; Pocock, M.; Bissell, M.; Madsen, C.; Nguyen, T.; Zhang, M.; Zhang, Z.; Zundel, Z.; Densmore, D.; Gennari, J. H.; Wipat, A.; Sauro, H. M.; and Myers, C. J. 2016. Sharing Structure and Function in Biological Design with SBOL 2.0. *ACS Synthetic Biology* 5(6):498–506.
- [Wilkins 1988] Wilkins, D. E. 1988. *Practical Planning: Extending the Classical AI Planning Paradigm*. San Mateo, CA: Morgan Kaufmann.