EBRC Engineering Biology Research Consortium

Engineering Biology

A Research Roadmap for the Next-Generation Bioeconomy

June 2019

roadmap.ebrc.org



5885 Hollis Street, 4th Floor, Emeryville, CA 94608 Phone: +1.510.871.3272 Fax: +1.510.245.2223

This material is based upon work supported by the National Science Foundation under Grant No. 1818248.

© 2019 Engineering Biology Research Consortium



<u>Appendix</u>

A primer on DBTL for Engineering Biology

Engineering biology is a rapidly advancing discipline in which biological circuits and biochemical pathways with predicted functionality are implemented in living systems using systematic engineering workflows. A major difference between engineering/synthetic biology and classical engineering disciplines lies in the fact that engineered systems have been constructed from man-made and well-characterized building blocks in a "bottom-up" design strategy. In contrast, engineering biology often relies on partly characterized biological components that are implemented in extremely complex and dynamic living environments (cells and organisms) that are poorly understood. Because of this complexity, classical engineering approaches are only partly applicable to engineering biology. An iterative Design-Build-Test-Learn (DBTL) cycle has been developed that relies on data analytics and mathematical models with the goal of characterizing and controlling for the host response. Currently, the DBTL cycle is closely connected to the synthetic biology ecosystem, with many different companies working in different parts of the cycle.

The **DESIGN** process encompasses both biological design and operational design. For example, biological designs can specify desired cellular target functions, such as a cell that produces a complex natural product or that generates a detectable signal in response to an extracellular analyte. For operational design, the experimental procedures and protocols requires design. For example, the optimal amount of sample required to execute a specific experimental protocol to achieve required data capture. Required performance specifications must also be captured so that the process has a set of quantitative objectives to meet. To implement these functions in an organism then requires identifying appropriate biological parts (e.g., enzymes, reporters, regulatory sequences, etc.) that can be assembled to implement the desired function. Because the universe of biological parts is large and growing, standard registries that characterize these parts under a variety of different biological contexts and environmental physiological conditions and host organisms will be necessary. New approaches will be needed to specify effective design functions that can be used to drive the assembly of these components into functional assemblies. New mathematical and computational tools will be needed to solve these optimization problems and to specify appropriate constraints. Lastly, these optimal mathematical solutions will need to be implemented using optimal genetic parts to effectively map the space of potential solutions to the space of solutions that can actually be engineered. Design-of-experiment, or DoE, approaches could play an important role in efficiently searching for and assembling genetic parts and circuitry to enable the specified design with DNA sequences derived from either databases or the literature. As the search space is vast, DoE approaches still require choices to be made on what to search. Also, DoE approaches must be supplemented by computational methods to speed up the search for optimal genetic parts. The end of the design process is one or more DNA sequence(s) comprised of multiple genetic parts that generate the desired functions in a targeted biochemical, cellular, organismal, or biome context.



The **BUILD** process primarily consists of DNA assembly, incorporation of the DNA assembly in the host, and verification of the assembled sequence in the expected genetic context. The DNA build process iteratively assembles the DNA sequence specified in the Design process. The DNA assembly process uses molecular biology techniques, often aided by robotic automation, to combine multiple DNA fragments together and generally requires transformation into a host organisms for screening and verification of proper assembly. Build constructs are verified by DNA sequencing, restriction enzyme digests, and other techniques directed by software tools. Many design constructs require multiple hierarchical rounds of DNA assembly. For instance, round one may be used to assemble individual transcriptional units or large genes, round two may be used to assemble multiple individual transcriptional units to generate a biosynthetic pathway. The result of the DNA build process is a physical DNA molecule or, increasingly, a pooled library of DNA molecules that comprises the specified DNA sequence(s).

Delivery and verification of the DNA build within the desired host, or host build, is the second build process. This involves delivering the build genetic construct into the host organism, either as an independent genetic entity (e.g., a circular DNA plasmid or artificial chromosome), or by integration into a host chromosome. This is accomplished using standard molecular biology tools and is termed transformation. The efficiency of the transformation and selection of cells that contain the desired genetic sequence is often optimized through automation and a high-throughput screening process. When working with unstudied hosts, identifying amenable conditions for transformation and integration can require significant research. For example, host-onboarding and host optimization can require significant genetic manipulations of the host before testing, to remove adverse phenotypes and improve a host's utility for a specific design process. This could include the removal of the host's restriction endonuclease system or endogenous toxins, alterations to the membrane to improve phage susceptibility or alter immune modulation, or even inserting 'kill switches' or other biosafety features depending on the specific application.

The **TEST** process involves assessing whether the desired specified biochemical/cellular functions encoded in the designed DNA sequence have been achieved by the host organism or biome. This could also include testing genetic designs in multicellular transgenic organisms, although the scale and complexity of measurements required is challenging. For unicellular organisms, this requires growing the organism and assaying for the desired function (e.g. quantifying production of the desired product). Full validation of proper function and debugging non-functional designs may require substantially more intensive analysis, including tools such as proteomics, liquid chromatography-mass spectrometry, cas chromatography-mass spectrometry, and next-generation DNA/RNA sequencing. Measurements of, for example, product titer and yield, enzyme activities, cell phenotype, sensing thresholds and dynamic ranges, allows an assessment of the efficacy of the current design against the user-defined optimal target function. For bioprocessing, a major challenge is in scaling, which in a Test context requires measurements at small volume to inform large volume fermentation, an area of active research.

The **LEARN** process utilizes measured data and mathematical (statistical or mechanistic) models of the engineered biochemical, cellular, organismal, or biome context to



obtain actionable insights that can be used to generate better designs in the next iterations. For example, the integration of multi-omics data with metabolic models has been used to identify genetic interventions that improve titer, rate, and yield of engineered pathways. The cycle is then repeated until the user-defined target function is achieved.

The DBTL cycle thus provides an overall and iterative design framework to enable systematic design of biological systems at the genetic level as well as the elucidation of potential genetic design rules.