How Effective Diversity Generation Drives Rapid Forward Engineering

For optimizing the biological performance of proteins and metabolic pathways, forward engineering is a powerful approach, particularly when the genomic rules are largely unknown. Effective forward engineering allows scientists to adjust parameters based on the performance needed as well as on the resources and time available.

A hallmark methodology of forward engineering is the use of iterative cycles of the design, generate, test, and learn (DGTL) process for navigating the complexities inherent to synthetic biology. The principle of effective diversity generation is key to the DGTL process because it recognizes that genetic sequence space is large and that knowing the appropriate intervention points is a central challenge for research focused on engineering biological systems.

The purpose of diversity generation is to build and test sufficient numbers of high quality genetic variants to fuel a subsequent combinatorial optimization step, which focuses on efficiently searching over the previously generated benificial diversity space. A crucial component of the forward engineering combinatorial optimization approach is that it leverages evolutionary optimization principles from the natural world to obtain the best outcomes. Qualifying genetic diversity can include insertions, deletions, and substitutions, and may span sequence lengths ranging from SNPs and single amino acid changes, all the way up to full gene knockouts. Additionally, these changes may be made in proteins, pathways, or across the genome.

Diversity generation typically depends on two primary strategies (Figure 1). The first is focused on generating ideas for variation based on existing studies and published literature and testing them to see if they yield the desired improvement in phenotype. This strategy focuses on screening a smaller group of well-characterized targets. The chances of success are high, but the strategy often results in few candidate hits. A second strategy is to screen a wider pool of targets, most of which researchers know little or nothing about. With this approach the overall success rate is lower, but it generates a greater number of hits for the subsequent optimization steps. Both strategies have value and should be pursued in tandem to obtain a large collection of beneficial diversity.

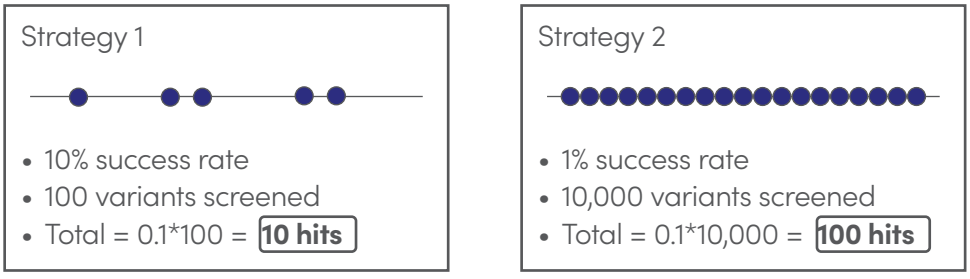

Figure 1: Diversity generation strategies

The Need for Scale and Variety

Introducing useful diversity requires thinking about both the number of genomic locations that can be targeted using a given technology, and the variety of edit types possible. Techniques such as genome-wide base editing, genome-wide knockout, and classical mutagenesis can introduce diversity into many locations in the genome. However, these approaches are limited in the variety of edit types delivered. In contrast, precision genome engineering technologies now introduce a much richer variety of edit types, but for the most part at only a limited number of genomic locations.

Greater Odds of Reaching the Goal

For scientists with limited test capacity, identifying an effective strategy for maximizing the number of shots on goal is crucial. One approach is to deeply sample small libraries. However, the math shows that it is actually preferable to sample larger libraries less completely. This approach shifts the focus from maximizing the fraction of possible hits per library to maximizing the number of unique hits per variant screened, which is more pertinent for forward engineering. Working with a smaller library places a cap on the number of possible unique variants that may be observed in the population. In contrast, larger library sizes offer access to a greater number of unique variants for the same testing capacity. The objective of effective diversity generation is to recover as many unique beneficial variants as possible in order to deliver a higher likelihood of forward engineering success in the optimization step. (Figure 2)

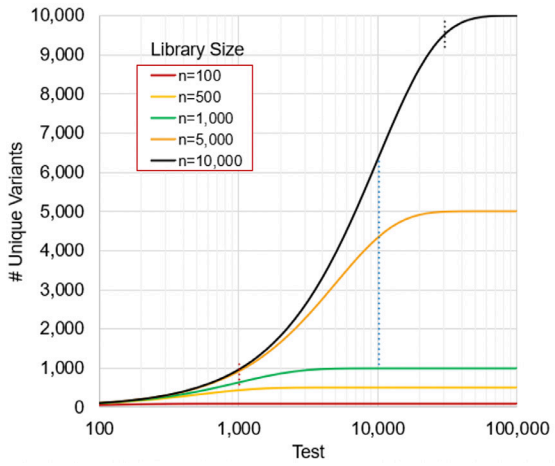

Figure 2: Using larger libraries to screen more variants can enhance a diversity-generation process.

With that goal in mind, there are some important principles to consider. The first is that library size should be similar to the test capacity — a good match is often between 0.5X to 2X the library size. In this regime most of the library members tested are unique, thereby maximizing the number of shots on goal for a finite test capacity. Secondly, oversampling libraries is not the most efficient use of test resources because it does not yield a corresponding gain in the number of unique variants and, crucially, total hits. It is important to acknowledge that oversampling can indeed return more unique hits and potentially more beneficial mutations; however, the obligatory disproportionate investment in test capacity may not be the best use of resources. Lastly, working with very large libraries and limited test capacity wastes build resources without offering a comparable gain in the number of unique variants and hits.

Metabolic Engineering of *E. coli* for Increased Lysine Biosynthesis

An example of how highly scalable genome engineering can be used for diversity generation in a forward engineering project offers a look at how this novel approach can make a difference for protein and pathway engineers. The project aimed to optimize the flux toward lysine in *E. coli*. In addition to suggestions from the literature about enzymes involved in this process, scientists performed a massive, unparalleled saturation mutagenesis campaign across all enzymes in the lysine pathway. These two approaches yielded roughly 175,000 single amino acid mutations that were designed and created. In addition, scientists introduced systematic loss-of-function edits across the *E. coli* genome as part of a genome-wide knockout; they also knocked in promoters of varying strengths in front of all the *E. coli* genes. Combined, these approaches yielded about 200,000 edits for evaluation in the search for ways to boost lysine production in the microbe (Figure 3).

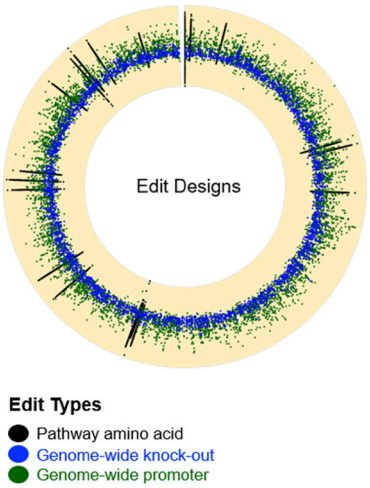

Figure 3: Representation of the edit designs in the genome. In a single experiment, several edit types were introduced to span the entire genome of *E. coli*.

Initial, shallow sampling of the edit space — about 10% of the 200,000 edits generated — produced numerous hits across all 24 libraries investigated. Follow-up testing of the suggested hits confirmed numerous beneficial targets across the lysine pathway — both known and novel ones — as well as 17 new targets outside of the lysine pathway, including both loss-of-function and gain-of-function hits in the promoter libraries searched. Pairing an amino acid variant identified in the lysine pathway with a high-strength promoter hit on the same gene resulted in a 10,000-fold increase in lysine production. Additional combinations on top of this initial pairing resulted in even higher yields (Figure 4). Deeper screening of these libraries to a recommended 0.5X depth would likely result in many more hits that could be stacked to improve performance further.

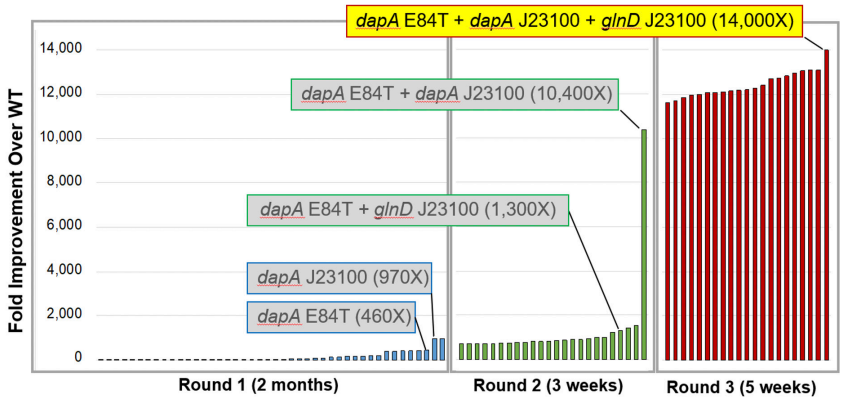

Figure 4: The biggest improvements in *E. coli* lysine production came from combining multiple hits produced in the diversity generation workflow.

The Way Forward

Despite decades of research in model microorganisms such as *E. coli* and *S. cerevisiae*, our current limited understanding of the rules governing the behavior of biological systems continues to hamper our capacity to predictably manipulate or design genomes from first principles. Until that day arrives, the ability to efficiently run many experiments in parallel through the iterative forward engineering process remains our most powerful approach for biological optimization. Finally, by continuing to leverage novel tools and technologies to most effectively generate and test high quality genetic diversity, an investigator will be well positioned to most rapidly benefit from the vast potential of the genome. Read more about the forward engineering process in our combinatorial optimization white paper (Doc # 1002367).

Learn more at **INSCRIPTA.COM**

Document 1002366 RevB 03232022