

5. HOW CODEXIS TECHNOLOGY WORKS

We are innovators in the directed evolution of enzymes and microbes to enable industrial biocatalytic reactions and fermentations via biocatalyst engineering, metabolic pathway engineering and fermentation microbe improvement.

Our approach to developing commercially viable biocatalytic processes begins by conceptually designing the most economically practical manufacturing process for a targeted product. We then develop optimized biocatalysts to enable that process design, using our directed evolution technology, including screening and validating biocatalysts under relevant conditions. Typical design criteria include stability in the desired reaction conditions, biocatalyst activity and productivity (yield), ease of product isolation, product purity and cost.

Alternative approaches to biocatalytic process development typically involve designing and engineering the biocatalytic processes around shortcomings of available biocatalysts, including, for example, biocatalyst immobilization (for stability and/or reuse), special equipment and costly product isolation and purification methods. We circumvent the need for these types of costly process design features by optimizing the biocatalyst for fitness in the desired process environment. As a result, we enable and develop cost-efficient processes that typically are relatively simple to run in conventional manufacturing equipment. This also allows for the efficient technical transfer of our process to our manufacturing partners.

The successful embodiment of our technology platform in commercial manufacturing processes requires well-integrated expertise in a number of technical disciplines. In addition to those directly involved in practicing our directed evolution technologies, such as molecular biology, enzymology, microbiology, cellular engineering, metabolic engineering, bioinformatics, biochemistry, and high throughput analytical chemistry, our process development projects also involve integrated expertise in organic chemistry, chemical process development, chemical engineering, fermentation process development, and fermentation engineering. Our tightly integrated, multi-disciplinary approach to biocatalyst and process development is a critical success factor for our company.

THE ENZYME OPTIMIZATION PROCESS

The enzyme optimization process starts by identifying genes that code for enzymes known to have the general type of catalytic reactivity for a desired chemical reaction. Typically, we identify gene sequences in published databases and then synthesize candidate genes having those sequences. Using a variety of biotechnology tools, we diversify these genes by introducing mutations, giving rise to changes in the enzymes for which they encode. The methods for diversifying these genes, and types of diversity being tested, often vary over the course of a biocatalyst optimization program. For finding initial diversity, methods typically include random mutagenesis and site-directed (included structure-guided) mutagenesis. We also test mutational variations that distinguish related enzymes among different organisms. Once we have identified potentially beneficial mutations, we test combinations of these mutations in libraries made using our proprietary gene recombination methodologies, gene shuffling and multiplexed gene SOEing.

With our proprietary gene shuffling methodology, we generate libraries of genes that have random combinations of the mutations we are testing. The pool of genes is used to transform host cells, which entails introducing the various genes, one each, into host cells. These cells are then segregated and grown into colonies. Cells from individual colonies are cultured in high throughput to produce the enzyme encoded by the shuffled gene in those cells. The enzymes are then screened in high throughput using test conditions relevant to the desired process. The screening results identify individual shuffled genes that produce improved enzymes having combinations of beneficial mutations and weed out enzymes having detrimental ones. Using different test conditions and/or different analytical methods, we can identify variant enzymes that exhibit various improved performance characteristics, such as stability, activity and selectivity, under conditions relevant to the desired chemical process.

In the next step in our optimization process, we use our proprietary software tool, ProSAR, to analyze protein sequence-activity relationships. We initially licensed ProSAR from Maxygen and further developed and customized ProSAR to address our specific needs. ProSAR aids in identifying specific gene and enzyme mutations that are beneficial, neutral or detrimental with respect to the desired performance characteristics. Earlier directed evolution methods did not separately evaluate individual mutations in libraries of variants which carry multiple mutations, where beneficial and detrimental performance characteristics may be mixed in an individual gene or enzyme. Capitalizing on the advent of inexpensive gene sequencing, we are able to determine which particular mutations are present in the genes and proteins we have screened.

Our ProSAR bioinformatics software relates the screening results to the mutations and ranks the individual mutations with regard to their degree of benefit or detriment, relative to whichever process parameter(s) the screening tested. Using that information, we can bias the pool of mutational diversity in the next iteration to further the accumulation of beneficial diversity and cancel out detrimental diversity in the individual genes in the resulting shuffled library. The ProSAR results also help us develop ideas about new diversity to test. ProSAR, combined with efficient gene synthesis and high quality library generation methods, has led to a significant increase in the efficiency and speed of enzyme improvement and optimization.

In another step of our optimization process, we take the best variants we have identified and prepare enough of each to test in the desired chemical process at laboratory scale, for in-process confirmation. This optimization routine is done iteratively, typically adding new diversity to the pool in each iteration. The gene that codes for the best performing enzyme in one iteration is used as the starting gene for the next iteration of shuffling and screening. As the enzymes improve over these iterations, the screening conditions are made increasingly more stringent. In this way, enzymes are rapidly optimized until all in-process performance requirements have been achieved and the economic objectives for the desired process have been met.

Multiplexed gene SOEing is our new proprietary methodology for rapidly generating gene variants. Using multiplexed gene SOEing, we rapidly generate collections of individual gene variants that have predetermined, as opposed to random, combinations of mutations we are testing. It is based on a biotechnology technique, which we refer to as SOEing, or Splicing by Overlap Extension, generally used

to make a hybrid, or spliced, gene from fragments of two genes and/or to introduce a specific mutation into a splice between fragments of one gene. We have automated the process to robotically make, in parallel, one hundred to several hundred variants, each with a predetermined combination of the mutations we are testing. The variants are introduced into host cells, and the encoded enzyme is produced and screened in high throughput, as described above.

Using multiplexed gene SOEing, we can test many mutations and combinations thereof in parallel, and because the mutation incorporation is controlled and predetermined before screening, as opposed to random incorporation and selection after screening, the resulting data set can be more optimal for ProSAR analysis.

We believe using multiplexed gene SOEing to quickly survey many mutations, followed by ProSAR-driven shuffling of beneficial mutations, is a particularly effective approach, providing rapid gains in enzyme performance.

ABOUT CODEX BIOCATALYST PANELS

Our Codex Biocatalyst Panels were initially developed to speed our own internal process for identifying enzymes with desired characteristics for further optimization. Each Codex Biocatalyst Panel is comprised of variants of one or more enzymes that catalyze one type of a generally useful chemical reaction. We assemble, on one or more microtiter sample plates, variants of a parent enzyme that we pre-optimize for stability in industrial chemical processes and for ready manufacturability. The variants are diversified to react to a variety of chemical structures that are susceptible to that type of chemical reaction.

Either we or our innovator pharmaceutical customers use the Codex Biocatalyst Panels to screen a new chemical structure against the assembled variants to rapidly identify variants that react with the new chemical structure. For some new structures, a variant on the panel could enable production of the desired product. We can also analyze the data from the panel screen using ProSAR to identify the mutations that are beneficial for the reaction of the new structure and further optimize the enzyme as needed using the enzyme optimization techniques described above. In cases where a customer wishes to screen a proprietary new chemical structure itself, we can produce a custom panel of new variants on a sample plate produced by multiplexed gene SOEing.

We may also use our Codex Biocatalyst Panels in our bioindustrial programs. In our biofuels research and development collaboration with Shell, we are developing a library of cellulases that have the potential to convert a wide variety of cellulosic biomass sources into fermentable sugars. The cellulosic biomass that we expect will be used to produce advanced biofuels is highly variable from region to region and can change over time. To optimize the local and seasonal conversion of cellulosic biomass to fermentable sugars, we expect to produce a Codex Biocatalyst Panel of cellulases that we or Shell can use to customize the biocatalysts that Shell uses at each advanced biofuel production facility. This technical innovation may ultimately make our sugar platform feedstock agnostic. Similarly, there is regional variation in coal. We may develop a Codex Biocatalyst Panel that we or our customers can use to tailor our carbon capture biocatalysts to the specific characteristics of the coal used in each energy facility that adopts our carbon capture technology.

THE GENE OPTIMIZATION PROCESS

For fermentation microbes, we enhance metabolic pathways by using gene optimization to improve the production and/or productivity of one or more enzymes in a series of in vivo reactions that make a desired product. We optimize the gene/enzyme using either in vitro or in vivo screening. For fermentation applications, the microbes containing the improved gene(s) are directly evaluated in laboratory scale fermenters.

The metabolic pathway may naturally exist in the microbe, but productivity and/or selectivity improvements are needed to economically produce more of the desired natural product and/or less of an undesired by-product. We can also introduce a new metabolic pathway to produce a desired product using our gene shuffling technology in combination with synthetic biology, a type of metabolic engineering in which new genes are introduced into a microbe.

We are using our gene/enzyme optimization methodologies in our biofuels program to optimize fermentation microbes, including optimization of:

- native and introduced (non-native) cellulase genes for increased productivity in our cellulase production microbes;
- an introduced (non-native) pathway in yeast for the conversion of xylose, a cellulose-derived sugar, to ethanol; and
- an introduced (non-native) pathway in a microbe for the production of our biohydrocarbon fuel molecule.

THE WHOLE GENOME SHUFFLING PROCESS

In addition to our gene optimization technology for enzymes, we have another complimentary technology in our platform for the optimization of fermentation microbes called Whole Genome Shuffling. Whole Genome Shuffling allows us to improve the performance of a fermentation microbe by shuffling unidentified mutations in unidentified genes across the genome. We start with a diversity of mutational variants of a fermentation organism, generated by conventional means such as random mutagenesis.

Our Whole Genome Shuffling involves introducing the entire genome of two or more such cells into a single cell, in which the genetic machinery of the combined cell recombines, or shuffles, the genomes. In one method, this is accomplished by protoplast fusion, in which the cell walls are removed to leave the cells' contents contained only by their cell membranes. The cell membranes of these protoplasts in the diverse population are induced to fuse together into fusants containing the genome of two or more of the parent cells. From these fusants, we regenerate normal cells, each with one copy of a hybridized genome.

Microbial colonies are then grown and screened for their performance in the fermentative production of the desired product. This process can be repeated, including with the introduction of new mutations, until the desired performance in the fermentation process is achieved. One of our collaborators is

operating a fermentation process for a generic pharmaceutical product using microbes we developed by Whole Genome Shuffling.

We are using our Whole Genome Shuffling technology in our biofuels program to optimize fermentation microbes, including optimization of:

- enzyme production hosts for increased production of cellulase enzymes;
- ethanol-producing yeasts for improved xylose utilization, ethanol productivity, and tolerance to higher ethanol concentrations; and
- our biohydrocarbon producing strain for increased productivity.

METABOLIC ENGINEERING AND SYNTHETIC BIOLOGY

In addition to our proprietary enzyme and microbe optimization technologies, we have built expert capabilities in a suite of new metabolic engineering technologies for the development and optimization of fermentation microbes. These technologies are generally applicable to our pathway and strain engineering programs. Genomics, transcriptomics, proteomics and metabolomics all provide more in-depth analyses of the metabolic functioning of fermentation microbes, and differences between variants, to guide further improvements. In many cases, these analyses help to identify enzymes that need to be modified (removed, increased, stabilized, or otherwise modified) in order to increase the overall productivity and performance of the strain.

Synthetic biology involves the design, synthesis and introduction of new genetic programming to organisms for new biological functions. This field has rapidly developed in recent years as DNA synthesis and sequencing costs have rapidly dropped. Using synthetic biology, we are taking advantage of the exploding publicly available gene and genome sequence information in our gene and metabolic pathway optimization projects. This information is being leveraged by our ProSAR software and multiplexed gene SOEing methodologies. For example, we use synthetic biology in our biofuels program to introduce non-native pathways for xylose utilization and for biohydrocarbon production and to optimize these pathways.