Development and application of computer-aided design methods for cell factory optimization

Genetically modified organisms (GMOs) can be used to produce chemicals for everyday applications. Engineering microorganisms is a multidisciplinary task comprising four steps: design, build, test and learn. The design and learn phases rely on computational, statistical models, data analysis and machine learning. The process of creating strains with commercially relevant titers is time consuming and expensive. Computer-aided design (CAD) software can help scientists build better strains by providing models and algorithms that can be used to generate and test hypotheses before implementing them in the lab.

Metabolic engineering already uses computational tools to design and analyze the metabolic and regulatory mechanisms of microorganisms. Genome-scale metabolic models (GEMs) describe the biochemical reactions in an organism and their relationship with the genome, hence they can be used to design microbial cell factories. In this PhD thesis we present cameo, a CAD software for metabolic engineering that uses GEMs. State-of-the-art and novel algorithms are implemented in cameo. These algorithms have been made accessible using a high-level API to enable any user to start running them without having advanced programming skills. Using cameo, we designed a Saccharomyces cerevisiae strain with improved mevalonate production.

In the food industry, recombinant DNA technologies cannot be used because of strict GMO regulations, especially in Europe. This industry relies on classical strain improvement (CSI) and adaptive laboratory evolution (ALE) to create new and better products. Nevertheless, some engineering and design principles can be applied to create strains in this industrial setup. In this work, we present MARSI, a software tool that uses a completely new model-based approach to strain design, focusing on metabolite targets. MARSI designs can be implemented using ALE or CSI.

We used MARSI to enumerate metabolite targets in Escherichia coli that could be used to replace experimentally validated gene knockouts.

Genetic variability occurs naturally in cells. However, the effects of those variations are unpredictable and can impact the performance of production strains. Moreover, strains resulting from CSI and ALE experiments contain a lot of mutations that are not trivial to explain. In this thesis, we explored strategies to integrate re-sequencing data using GEMs. Here, we present a workflow to integrate and analyze data from E. coli wild-type, mutant and closely related strains. In this study, we evaluated the effect of genetic variability on kcats. These parameters can be used to constrain GEMs and produce more accurate predictions. Therefore, using a combination of bioinformatics, chemoinformatics and machine learning tools, we explored the landscape of kcats using multiple enzyme sequences and their chemical reactions.