EXPLORING DESIGN SPACES WITH BAYESIAN OPTIMIZATION: A TOOL FOR THE SYNTHETIC BIOLOGY OF MODULAR PROTEINS

SUPERVISORS

Prof. dr. ir. Yves Briers

Dr. ir. Michiel Stock

TUTOR

Ir. Bjorn Criel

BACKGROUND

C&G, C, M

MORE INFO

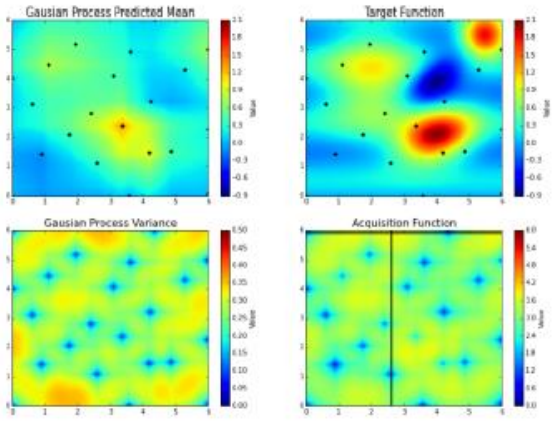
michiel.stock@ugent.be

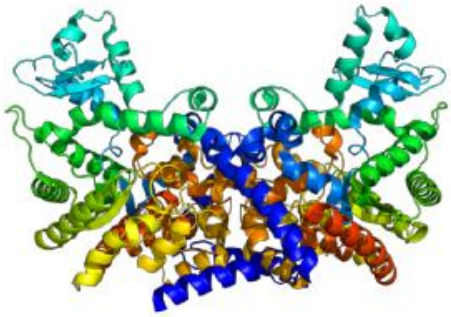
bjorn.criel@ugent.be

Background

In the laboratory of applied biotechnology, we develop modular proteins (composed of multiple domains) for various applications. We can engineer them with **synthetic biology** by recombining their domains and so, create new modular variants with improved properties. One of the applications we focus on is the development of a new class of enzyme-based **antibiotics** (**enzybiotics**).

This thesis will focus on the development of an enzybiotic against *Enterococcus faecalis* in artificial urine (to mimic a urinary tract infection). By recombining a set of natural building blocks, we aim to find a modular variant that is highly active against the targeted bacterium under the desired conditions. However, when the number of available building blocks increases, the size of the combinatorial design space, i.e. all possible modular variants, explodes. It quickly becomes impossible to screen all combinations.





Scope of the thesis

The goal of this thesis is to develop a **Bayesian optimization** toolbox to efficiently find the best candidate in a large combinatorial design space. To achieve this, a **predictive model** will be trained on a fraction of the design space to predict the activity of all the other modular variants. Next, Bayesian optimization will be applied to identify the modular variants that are most likely to be more active than the current best variant. To achieve this, the algorithm will account for both the predicted activity of each variant and the uncertainty on this prediction. The top ranked variants will be screened, and the model is updated. An iterative repetition of this approach will efficiently lead us to the best modular variant without screening the complete design space.

While the proposed approach will be developed to find the best modular enzybiotics against *E. faecalis*, it has a much **wider application potential**. Besides other modular proteins, it will be applicable to all sorts of combinatorial optimization problems (e.g. selecting microbial ecological communities or making new beers).