Introduction

In industrial biotechnology, enzymes are produced by over-expressing genes in production hosts such as Aspergillus niger. Ideally, the produced protein should be secreted, such that it can easily be recovered from a reactor. Previously, we exploited a dataset of proteins to predict successful high-yield secretion. Here, we develop a method to rationally redesign the sequence of a low-yield protein to make it more similar to that of high-yield proteins, with the aim to improve secretion yields.

Constraints

To reduce the probability that a mutation affects enzymatic activity, both buried (white) and ligand-binding residues (orange) are fixed.

Design method

A genetic algorithm is used to optimize the objectives under the given constraints.

4) In silico validation

Increasing sequence similarity compared to a protein from the positive set

As an in silico test, a negative protein (red) is redesigned with the number of mutations set to 30. The redesigned version (green) is afterwards compared to a positive protein (blue) with a similar structure. Mutated positions are shown below. In blue the residues at the mutated positions in the structurally aligned positive protein. The dots indicate identical residues, clearly showing an increasing number for the redesigned sequence, thereby showing an increasing sequence similarity with the positive protein.

Objective 1

Prediction outcome

The first objective is to increase the prediction outcome, i.e. promote mutations from amino acids with a negative weight to amino acids with a positive weight. As illustrated in the histogram above, the aim is to increase the prediction from 0.30 to around 1.10.

Objective 2

Often observed mutations

The second objective promotes mutations that are often observed in homologous proteins. A multiple sequence alignment is used to create a position frequency matrix.

Amino acid composition

The third objective aims to optimize the similarity between the amino acid composition of the redesigned protein to the average amino acid composition of the proteins in the positive set.

Objective 3

Amino acid composition

To test this, we developed a protein design method that aims to alter the amino acid composition of a protein without affecting protein structure and enzymatic activity. The designs will be tested in the lab.