→ A rational protein redesign method for improved secretion yields in *Aspergillus niger*

B.A. van den Berg[*1,3,4], M.J.T Reinders[1,3,4], J.M. van der Laan[2], J.A. Roubos[2], and D. de Ridder[1,3,4]

[1] The Delft Bioinformatics Lab, Faculty of Electrical Engineering, Mathematics & Computer Science, Delft University of Technology, Delft, The Netherlands, [2] DSM Biotechnology Center, Delft, The Netherlands, [3] Netherlands Bioinformatics Centre, Nijmegen, The Netherlands, [4] Kluyver Centre for Genomics of Industrial Fermentation, Delft, The Netherlands

* b.a.vandenberg@tudelft.nl

Introduction

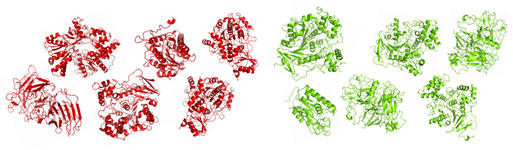
In industrial biotechnology, enzymes are produced by over-expressing genes in production hosts such as *Aspergillus niger*. Ideally, the produced protein should be secreted, such that it can easily be recovered from a reactor. Previously, we exploited a dataset of proteins to predict successful high-yield 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.

1) Data set

345 *A. niger* proteins

negative set
167 proteins: low secretion yield

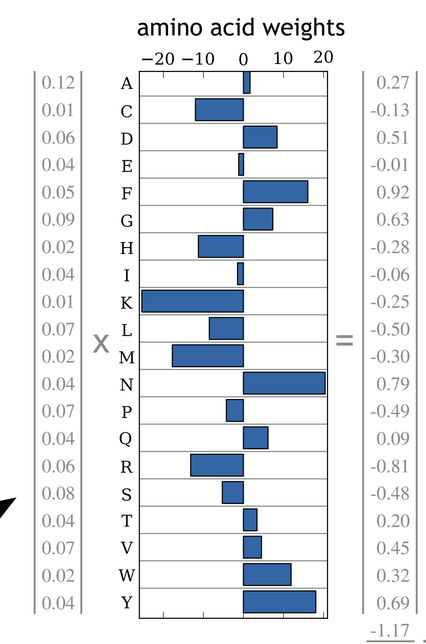
positive set
187 proteins: high secretion yield



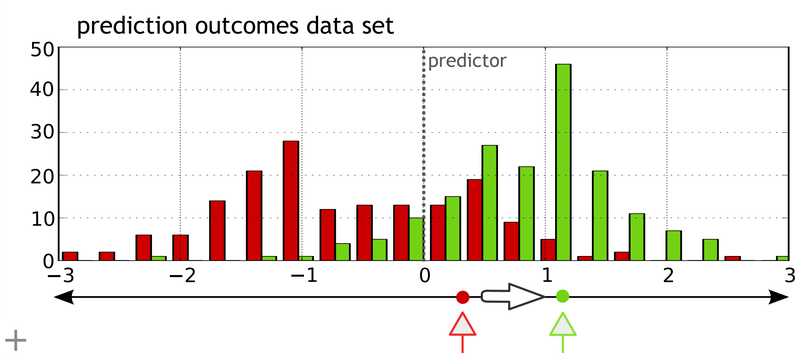
All proteins are predicted to be secreted, i.e. contain a signal peptide and do not contain an ER-retention signal or transmembrane regions. A high secretion yield is assigned in case of a high extracellular concentration after overexpression (visible band on gel), low secretion yield otherwise.

2) Predicting high secretion yield

Amino acid composition most predictive



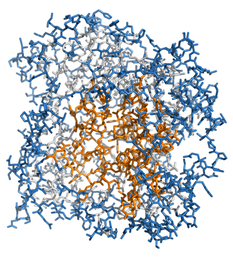
To predict high-yield secretion, many features have been explored. The amino acid composition was found to be most predictive. High-yield secretion prediction is calculated by taking the sum of a protein's weighted amino acid composition. The amino acid weights are retrieved from a linear support vector machine that was trained using the proteins in the data set.



Constraints

Fixed residues

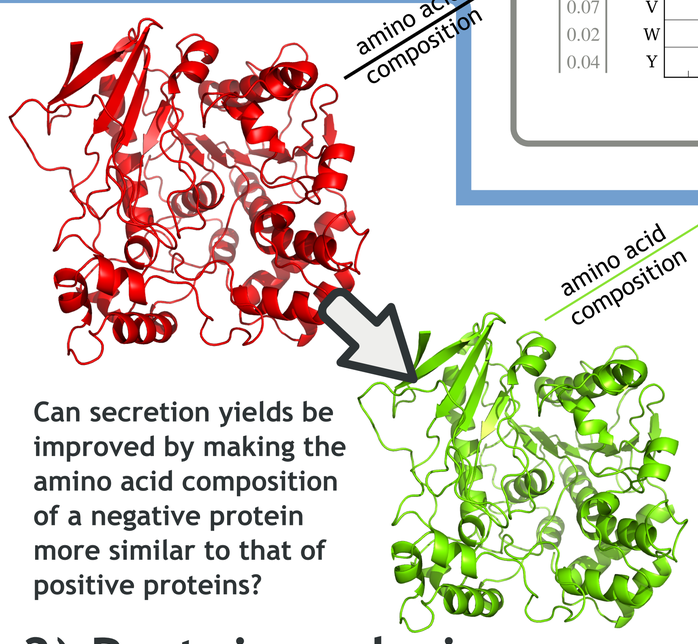
To reduce the probability that a mutation affects enzymatic activity, both buried (white) and ligand-binding residues (orange) are fixed.



Design method

Genetic algorithm

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



Can secretion yields be improved by making the amino acid composition of a negative protein more similar to that of positive proteins?

3) Protein redesign

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.

Objective 1

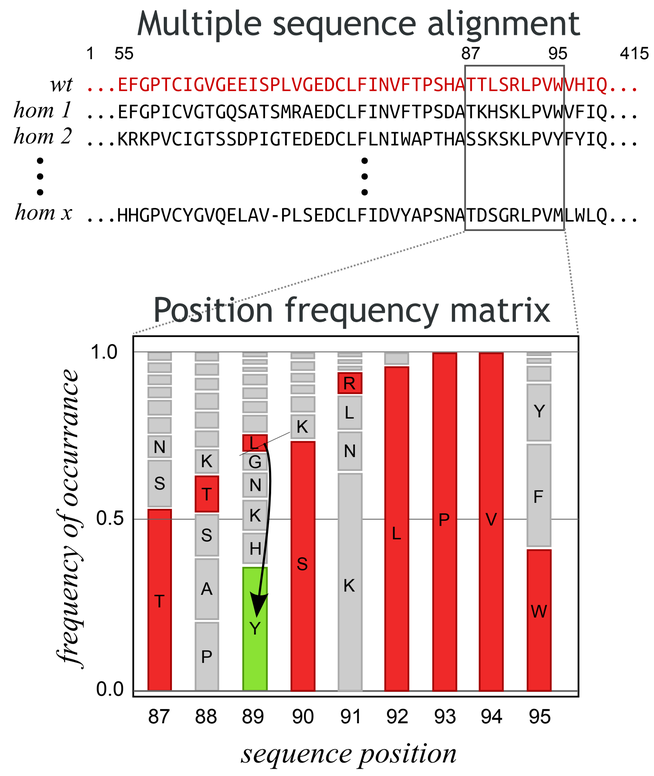
Prediction outcome

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 for a set of homologous proteins is used to create a position frequency matrix.

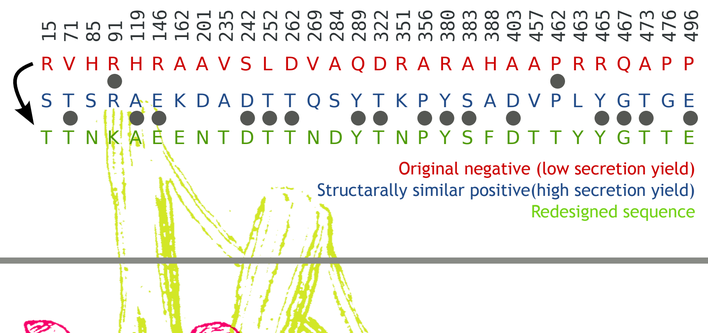


Mutations from rarely observed amino acids to frequently observed amino acids are prefered. For example, a mutation at position 89 from leucine (L) to tyrosine (Y) would be promoted, because Y occurs more often in homologous proteins at this position.

4) *In silico* validation

Increasing sequence similarity compared to a protein from the postive set

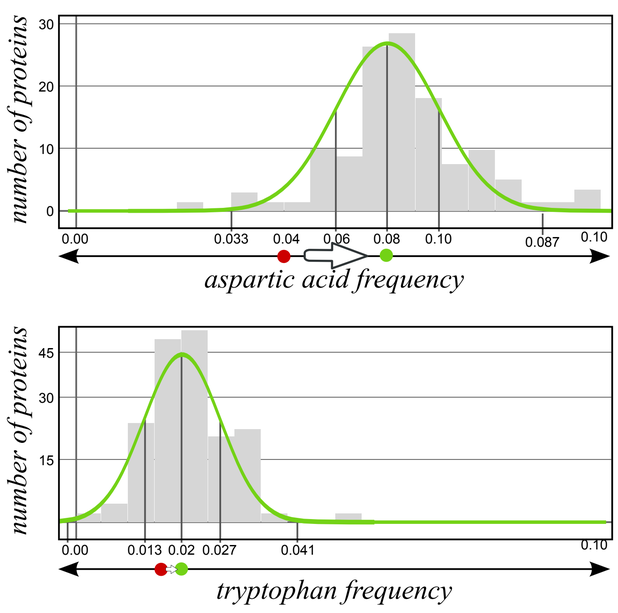
As an *in silico* test, a negative protein (red) is redesigned with the number of mutations set to 30. The redesigned version (green) is afterwards compared to a positive protein (blue) with a similar structure. Mutated positions are shown below, in blue the residues at the mutated positions in the structurally 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.

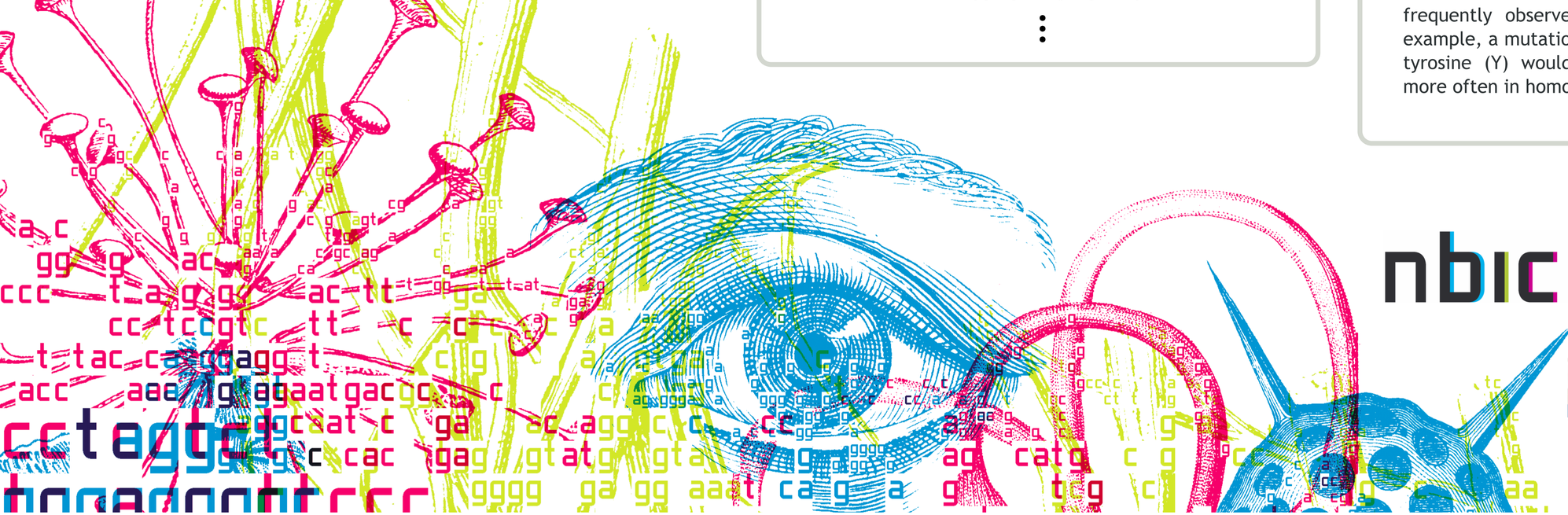


Objective 3

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.





nbic netherlands bioinformatics centre

TU Delft

DSM BRIGHT SCIENCE. BRIGHTER LIVING.

Kluyver CENTRE | Kluyver Centre for Genomics of Industrial Fermentation