Signal-to-noise limitations of genetic toggle switches

Lukáš Kuhajda, supervisor: Doc. M.Sc. et M.Sc. Daniel Georgiev, Ph.D.

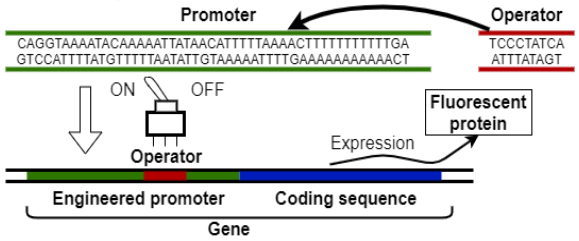
FAKULTA APLIKOVANÝCH VĚD
ZÁPADOČESKÉ UNIVERZITY
V PLZNI

Abstract

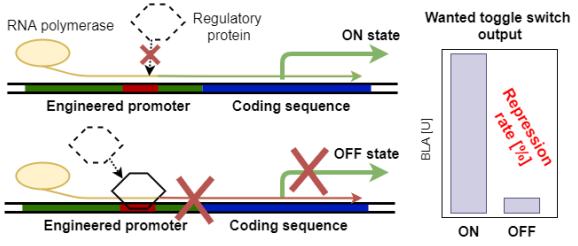
Purpose is to create a genetic toggle switch by which gene expression can be turned ON or OFF. A novel neural network (NN) architecture is designed and applied to solve the problem. Performance of the method was tested by genetic engineering of the yeast species *Saccharomyces cerevisiae*.

Problem formulation

Functional DNA units are called genes and they consist of two relevant parts: a coding sequence (what is expressed) and promoter (when it is expressed). Genetic engineer can insert an operator sequence into a promoter that permits expression to be toggled between ON/OFF states by the addition of chemical inducers. The state is measured by including fluorescent protein in the coding sequence.



Problem is to find the most effective site for the operator in a given promoter. Effectivity of the placement should optimize two criteria: 1) minimize disruption of the natural promoter denoted by the ON expression level and 2) maximize the difference between the ON/OFF expression levels.



Motivation

Current practice is to place operators into promoter sequences purely on the basis of expert knowledge. Finding the right location is time consuming and costly and it rarely works. Moreover, the results are not transferable to other promoter sequences.

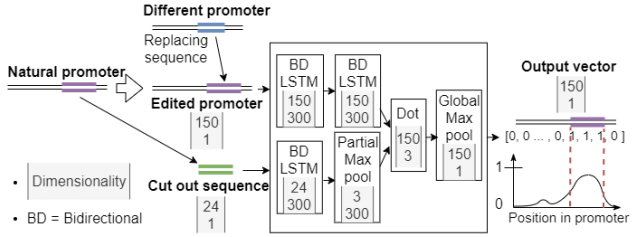
Machine learning method – Place-back

First attempts to solve the problem with NN were unsuccessful. A novel recurrent NN called Place-back was proposed. Place-back estimates optimal operator placement based on learned nucleic acid patterns in fungal promoters.

Training NN:

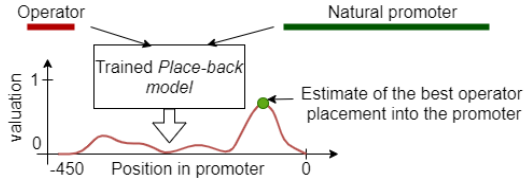
A database of 100 000+ fungi promoters was collected and annotated. The Place-back method was implemented in three steps: 1) a selected natural promoter is encoded by SentencePiece tokenizer, 2) 8-24 long fragment is removed and replaced with a sequence from a non-related promoter, 3) performance of the NN is evaluated based on its ability to place-back the removed sequence.

Place-back validation metrics: Precision = 0.61 Recall = 0.27, F1 = 0.37



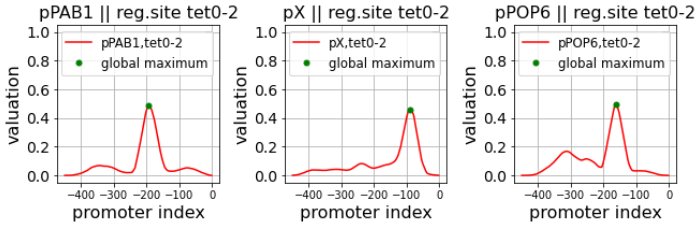
Example output of NN:

Given a natural promoter and operator sequence Place-back evaluates each base along the promoter with a value on the interval <0, 1>.



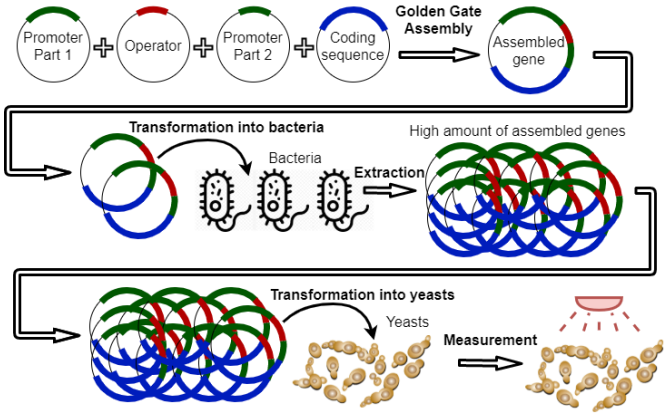
Results in silico

The NN method was applied to five different promoter sequences (3 shown below). Figures show placement scores for the *TetR* operator with the optimal location marked in green.



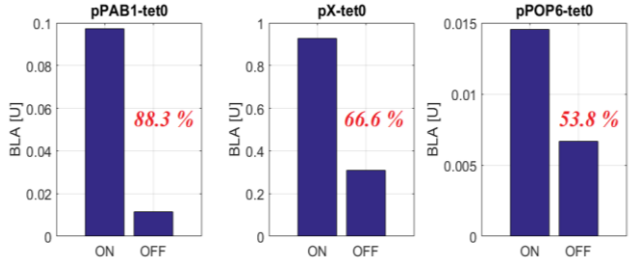
Genetic engineering method

According to NN results corresponding genes were assembled using Golden Gate Cloning method. Assembled genes were multiplied in bacteria, extracted and transformed into *S. cerevisiae*. Than genetically engineered strains were measured by enzymatic assay in a standard plate reader.



Results of biological experiment

Experimental measurement of NN designed promoters switches (3 shown below).



Conclusion

Promoter DNA sequences were designed by a novel machine learning method and shown to be functional by genetic engineering of yeast cells. Significant repression rate was obtained without any re-design in 3/5 never before tested promoters (highest repression rate = 88.3%). To the best of our knowledge this is the first successful demonstration of operator placement by a NN.