Genome Regulation using a Newly Designed Neural Network

Lukáš Kuhajda[1], Daniel Georgiev, Jan Švec[2]

1 Introduction

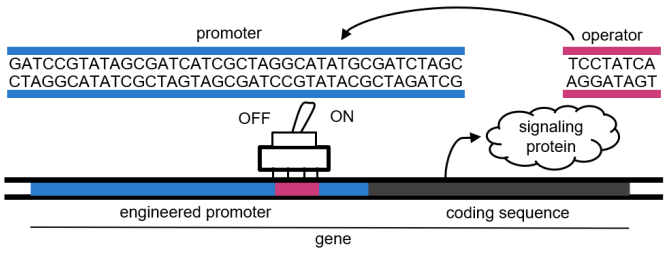
Purpose is to create a genetic toggle switch by which gene expression can be turned ON or OFF. A novel neural network (NN) architecture is designed and applied to solve the problem. Performance of the method was tested by genetic engineering of the yeast species *Saccharomyces cerevisiae*.

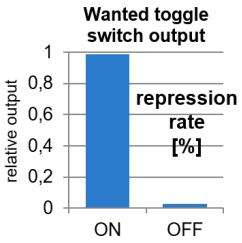
2 Problem Formulation

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 proteins 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.

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.



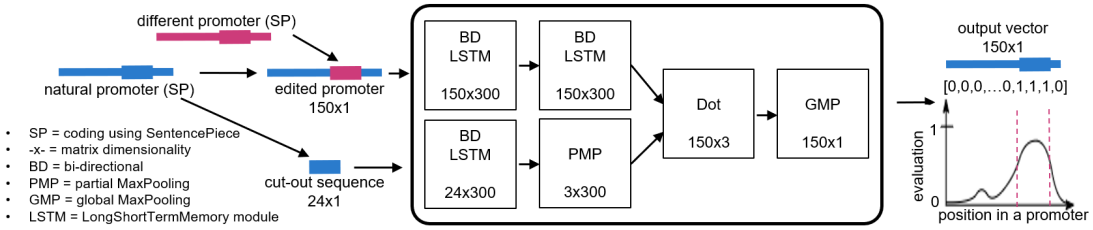


3 Neural Network Method – Place-Back

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 the same-length sequence from a non-related promoter, 3) performance of the NN is evaluated based on its ability to place-back the removed sequence.

[1] PhD student of Applied Sciences and Informatics, field of study Cybernetics, specialization Artificial Intelligence Usage for Synthetic Biology, e-mail: kuhajdal@ntis.zcu.cz
[2] University of West Bohemia, Faculty of Applied Sciences, Dept. of Cybernetics, email:georgiev@kky.zcu.cz, honzas@kky.zcu.cz

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

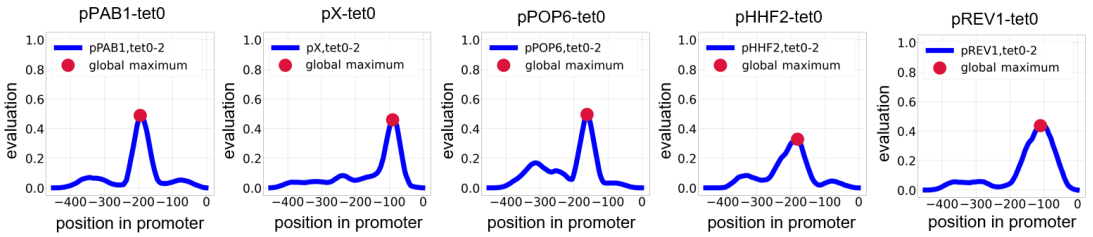


Given a natural promoter and operator sequence Place-back evaluates each base along the promoter with a value on the interval [0, 1]. Area in the promoter with amplified evaluation shows position for an operator placement.

4 Results

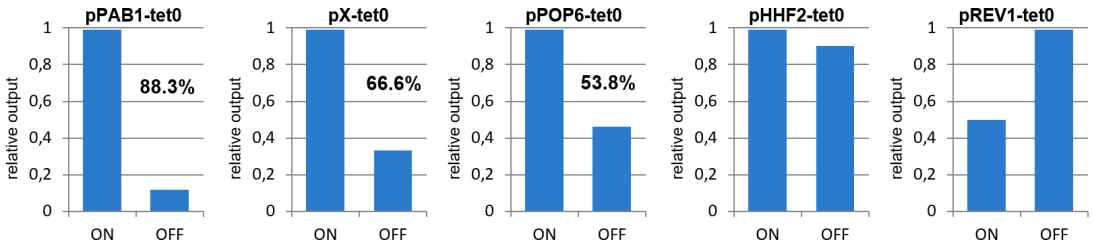
4.1 In silico

Exhibition of five chosen promoters for biological experiments. Figures show curve how NN would place TetR regulatory site into given promoters with marked optimal estimate.



4.2 In vitro

According to NN outputs corresponding genes were assembled, multiplied in bacteria, extracted and transformed into *Saccharomyces cerevisiae*. Genetically engineered strains were tested on toggle-switch behavior.



5 Conclusion

It is the first experiment of its kind and only one iteration of NN results was experimentally tested. Even so significant repression rate was obtained in 3/5 cases. The highest repression rate was 88.3 % and in the promoter toggle switch behavior was never tried on before.

Acknowledgement
This work was supported by the University of West Bohemia grant, project No. SGS-2019-027