

# Predicting Transcriptional Output of Synthetic Multi-Input Promoters

David M. Zong<sup>1</sup>, Selahittin Cinar<sup>2</sup>, David L. Shis<sup>3</sup>, Krešimir Josić<sup>2</sup>, William Ott<sup>2</sup>,  
Matthew R. Bennett<sup>3,4</sup>

**Short Abstract** — Recent advances in synthetic biology have led to a wealth of well characterized parts. Considering these developments, the potential to create novel multi-input inducible promoters with a greater number of inputs has become simpler. However, the potential combinations of inputs rapidly outpace our capability to fully characterize every possible combination. In this study, we describe two methods to predict the output of multi-input systems as a function of varying concentrations of multiple inducing ligands by combining characterization data of single input systems. These methods can be used by synthetic biologists to better design systems that utilize multi-input promoters.

## I. PURPOSE

AS our ability to design and construct sophisticated synthetic circuits continues to grow, so too must our ability to predict the performance of such circuits *in silico* [1]. In our poster, we will present results of two predictive modeling frameworks. The naïve framework is philosophically ideal in that our naïve models accurately predict multi-input system responses using only single-input data (and knowledge of promoter architecture). The energy modeling framework provides an even higher level of accuracy but requires a small amount of multi-input data.

Importantly, both modeling frameworks scale to systems with large numbers of inputs. In particular, the amount of multi-input data needed to train our energy models scales linearly with the number of inputs, while the number of inducer combinations that such models predict scales exponentially: Suppose we wish to predict a  $D$ -input system, where each input takes  $V$  possible values. Our energy modeling framework accurately predicts system output for all  $V$  input combinations, while requiring only  $D(V-1)+2$  values for model specification. When using a small amount of multi-input data to train our energy models, one natural question arises: How have we chosen the particular ‘one-dimensional’ subset of the multi-input data? Would not another subset work just as well, or perhaps better? Answering the first question, by selecting the ‘one-dimensional’ subset of multi-input data by varying one

inducer at a time while holding the others at full induction, we thereby probe the full dynamic range of each individual inducer. To validate our choice, we compared the error associated with our energy model prediction to the error produced by using all of the multi-input data to fit the energy model. Importantly, these error values essentially match for all of the systems we tested. (This observation leaves open the possibility that other subset choices may work just as well).

## II. CONCLUSION

Practitioners should consider the following when selecting a method. The naïve model is ideal when the user has a vast library of well-characterized single-input devices and wishes to evaluate the digital (on/off) behavior of potential designs without having to perform additional lab work. The naïve framework provides predictions that will aid the user in narrowing down the large design space of potential multi-input combinations and select candidates to build or analyze further. The energy model excels when the user has already constructed a multi-input system and wishes to probe the entire induction space. By collecting a small set of induction data, the rest of the induction space can be predicted to a high degree of accuracy.

Our predictive energy modeling framework captures the analog nature of the inputs/output of multi-input promoters. Such promoters are often treated as digital devices, because it can be too resource intensive to test the entire input space. Digital inputs/output approximation works in certain situations. However, when designing microbes for complex environments such as the gut microbiome or soil, relevant signals may be in constant flux. An analog predictive approach is therefore necessary, as it facilitates the design of circuits that can accommodate a range of signals [2]. The analog approach assists with the parts problem as well: Analog circuits can require fewer parts than their digital counterparts to compute a given function [3].

Overall, our predictive methodologies facilitate the design of synthetic microbes that can operate in complex, dynamic environments.

## REFERENCES

- [1] Shis, D., Hussain, F., Meinhardt, S., Liskin, S., and Bennett, M. (2014) Modular, Multi-Input Transcriptional Logic Gating with Orthogonal
- [2] Venturelli, O. S., Egbert, R. G., and Arkin, A. P. (2016) Towards Engineering Biological Systems in a Broader Context. *Journal of Molecular Biology*, 428(5), 928–944.
- [3] Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (5, 2013) Synthetic analog computation in living cells. *Nature*, 497(7451), 619–62

Acknowledgements: This work was supported by National Science Foundation Graduate Research Fellowship Program [1450681].

<sup>1</sup>Graduate Program in Systems, Synthetic, and Physical Biology, Rice University, 6100 Main Street, Houston, TX 77005 USA

<sup>2</sup>Department of Mathematics, University of Houston, 4800 Calhoun Road, Houston, TX 77004 USA

<sup>3</sup>Department of Biosciences, Rice University, 6100 Main Street, Houston, TX 77005 USA

<sup>4</sup>Department of Bioengineering, Rice University, 6100 Main Street, Houston, TX 77005 USA