Case Study: Digital logic circuits in yeast with CRISPR-dCas9 NOR gates[‡]

Hieu Do[*], Jonathan Mah[*], Zachary McNulty[*]

Amath 422, Autumn 2018

[‡] Original work by Miles W. Gander, Justin D. Vrana, William E. Voje, James M. Carothers, and Eric Klavins.
[*] Denotes equal contribution to work.

1 Introduction

Gene regulatory networks are controlled by proteins known as transcription factors, which control the rate at which DNA is transcribed into RNA, and by extension, the rate at which genes are expressed. These transcription factors include activators, which increase gene expression, and repressors, which inhibit gene expression. The interactions of gene regulatory networks allow cells to execute complex decision making processes in response to internal and external stimuli. The implications of constructing equally complex synthetic digital circuits in living cells would revolutionize ongoing research in synthetic biology – relevant applications include gene therapies that respond to disease, high-specificity cancer diagnostics, and stem-cell induced organ growth (Roybal et al., 2016; Xie et al., 2011; Guye et al., 2016). As discussed in this case study, Gander et. al have developed a novel method for constructing digital logic circuits in yeast cells (Gander et al., 2017).

In prior studies, programmable CRISPR-*dCas9* transcription factors were used to construct simple digital circuits through *dCas9*-mediated repression (Nielsen and Voigt, 2014). This repression mechanism is achieved through steric hindrance, as the *dCas9* protein binds to CRISPR regions of DNA, but does not make a double-stranded cut, resulting in interference with RNA transcription (Gander et al., 2017). However, the resulting CRISPR-*dCas9* transcriptional protein complex did not have a suitably low rate of transcriptional leakage, i.e., there were still moderate levels of gene expression even when fully repressed. As a result, the number of components in these simple CRISPR-*dCas9* digital circuits was limited, since the output signal would decrease with each additional component. Gander et al. improved upon these existing simple digital circuits by incorporating the use of the *Mxi1* transcriptional protein complex, which further represses gene expression to the effect of greatly decreasing the rate of transcriptional leakage (Gander et al., 2017).

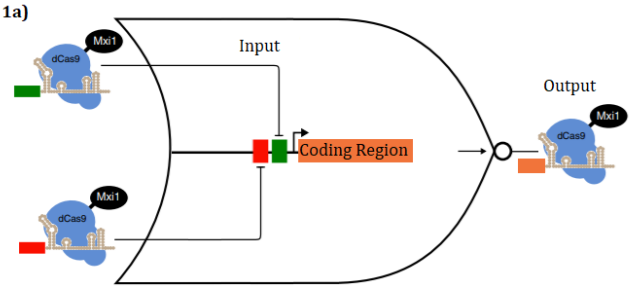
Using this improved CRISPR-*dCas9-Mxi1* repression mechanism, Gander et al. built a universal, single-gene negation-or logical (NOR) gate, in which the two input signals are strings of guide RNA (gRNA) sequences in the promoter region of the NOR gate, and the output signal is a string of gRNA sequences in the promoter region of a different NOR gate further downstream. As a result, if either input signal gRNA of a NOR gate were to be expressed, the CRISPR-*dCas9-Mxi1* transcriptional protein complex would bind to the associated promoter and repress that NOR gate. Similarly, if the output signal of the NOR gate were to be expressed, then the CRISPR-*dCas9-Mxi1* transcriptional protein complex would bind to the

promoter region of a different NOR gate further downstream. Figure 1 includes a graphic representation of a CRISPR-*dCas9-Mxi1* NOR gate.

Notably, NOR gates are functionally complete, meaning that permutations of connected NOR gates can be used to represent any boolean logic function. Using this feature, Gander et al. composed the CRISPR-*dCas9-Mxi1*-mediated NOR gates to form negation logical (NOT) gates, which produce an output signal of opposite truth value to the input signal (Gander et al., 2017). They then sequentially linked NOT gates together to form a "repression cascade", in which each NOT gate represses the subsequent NOT gate, with the final NOT gate expressing green fluorescent protein (GFP), whose fluorescence could be measured in arbitrary units. From this, they designed a mathematical model of cascade behavior, and demonstrated the greatly decreased rate of transcriptional leakage.

In this case study, we reproduce a figure from the original article, analyze the sensitivity of model parameters, expand upon the original mathematical model, and apply the mathematical model to hypothesize the system behavior of CRISPR-*dCas9-Mxi1*-mediated NOR gates used to create the Violacein synthetic cascade (Zalatan et al., 2015).

1a)



1b)

Truth Value of NOR Gate	
Input Signals (1, 2)	Output Signal
True and True	False
True and False	False
False and True	False
False and False	True

Figure 1: 1a) Simplified rendition of CRISPR-*dCas9-Mxi1* NOR gate based on Figure 1a from original study (Gander et al., 2017). gRNA sequences targeted by the *dCas9* protein of the CRISPR-*dCas9-Mxi1* complex are associated with the promoter region of the NOR gate. If either of the two input gRNA's are expressed, then their targeted promoter region will be repressed. Thus, the coding region of the NOR gate is only expressed when neither of the two input gRNA's are expressed. The resulting output is in turn another sequence of gRNA which will associate with the promoter region of a different NOR gate further downstream. **1b)** Associated truth table of NOR gate, in which a value of "True" indicates that the gRNA signal is expressed. Given the two inputs in the "Input Signals" column, a NOR gate will output the corresponding "Output Signal".

2 Mathematical Model

Gander et. al constructed a kinetic model to capture the behavior of the NOR gate cascades, given by the following system of differential equations (Gander et al., 2017),

$$
\begin{aligned}
\dot{r_D} &= b\left(\frac{v(\frac{u}{K})^{n_u}}{1+(\frac{u}{K})^{n_u}} - r_D\right) \\
\dot{r_d} &= b\left(\frac{v_d(1-L)}{1+(\frac{r_{d+1}}{k_d})^{n}} + Lv_d - r_d\right) \\
d &\in \{1, ..., D-1\} \\
\dot{G} &= B\left(\frac{1-L}{1+r_1^n} + L - G\right)
\end{aligned}
$$

This deterministic model is given by a system of three ordinary differential equations: r_D is the concentration of the input gRNA signal; r_d is the concentration of output at the d^{th} layer of the cascade, indexed by d ranging from $D-1$ to 1 with D being the total number of layers in the cascade; G is an arbitrary measurement of the normalized concentration of green fluorescent protein. Additionally, the system incorporates several parameters, whose values were fit using two optimization strategies: differential evolution and the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm (Storn and Price, 1997). v_d is the maximum steady-state concentration of gRNA from the promoter for each r_d. b is the rate of degradation for gRNA at all layers of the repression cascade. B is the rate of degradation of green fluorescent protein. k_d is the number of repressors required to suppress the associated promoter at layer d to half-strength. V is the maximum transcription, and by extension the maximum expression level of the input promoter. K is the Michaelis-Menten constant, that is, the substrate concentration at which the reaction rate is half of its maximum. n_u is the Hill coefficient of the inducible promoter. u is the concentration in μm of the input β-estradiol. L is the percentage of transcriptional leak observed at maximal repression.

Each layer of the repression cascade is treated as a Hill function which represents the transcription and repression of each successive gRNA-*dCas9-Mxi1* signal. This is notably similar to the mathematical model of the "Repressilator" gene regulatory network (Elowitz and Leibler, 2000).

3 Reproduction of Results

The fit model parameters used in the original paper were kindly provided to us by Gander et. al. We used the provided fit model parameters to recreate the mathematical model of the regression cascade. Numerical integration was performed using the python package, Tellurium (Medley et al., 2018). An example of Tellurium syntax and all code used for modeling is provided in the appendix. In Figure 2, we show a reproduced figure from the original paper.

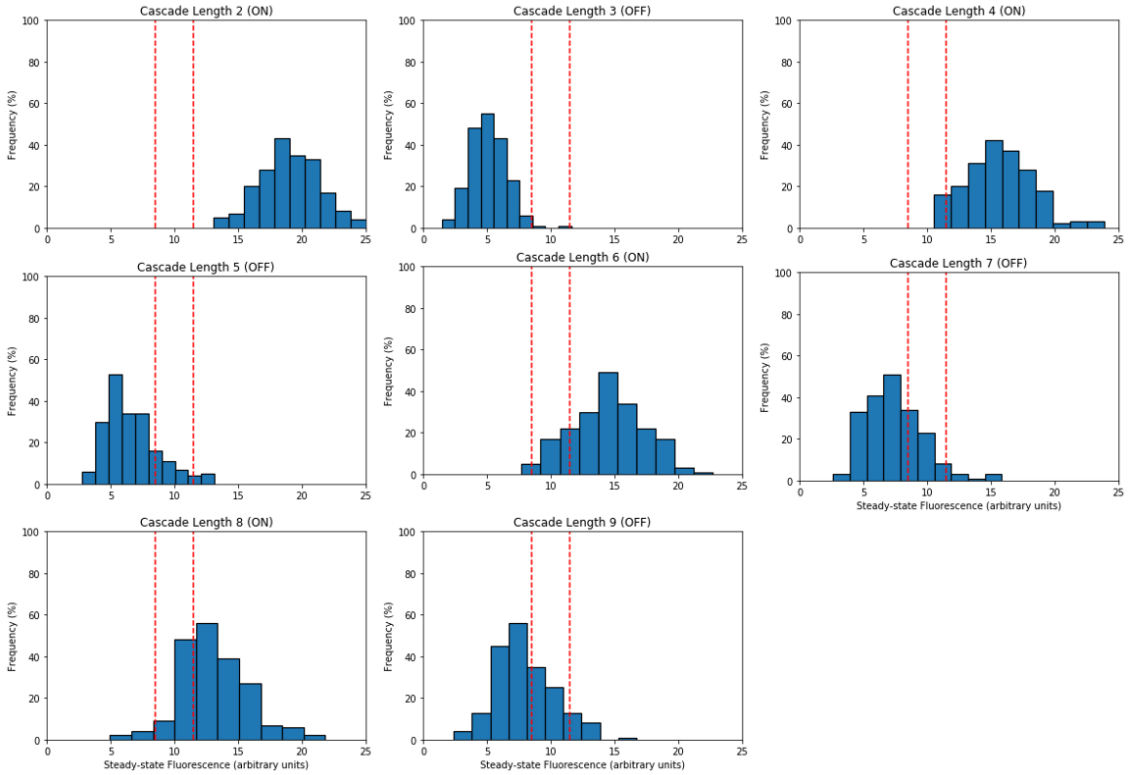


Figure 2: Recreation of Figure 4a from original study (Gander et al., 2017). The eight plots shown describe the steady-state fluorescence of the repression cascade model for two through nine layers. Each histogram represents two-hundred stochastic model simulations. Each additional cascade layer indicates the connection of another NOT gate. Note that in the original study, fluorescence was measured in arbitrary units. In recreating the figure, we also decided to use arbitrary units of the steady-state fluorescence, but have rescaled for readability. The dotted red lines indicate the cutoffs for the consideration of "ON" and "OFF" output signals.

4 Novel Results

4.1 Parameter Sensitivity Analysis

As part of our investigation, we analyzed the sensitivity of the model parameters. The purpose of the regression cascade model is to demonstrate the low rate of transcriptional leakage achieved with the use of CRISPR-*dCas9-Mxi1* logic gates. Thus, we defined "ideal" parameters as parameters which result in a large difference between sequential output values, that is, parameters which make the outputs of successive cascades more obviously different. Figures 3 through 8 describe our analysis of the parameter sensitivity. The mean value of the fit model parameters is shown as a vertical black line. For the six model parameters, L, b, v_d, k_d, n, and concentration of the initial input signal, we perturbed around the mean parameter value and held all other parameters constant. Solid lines indicate the steady-state value of fluorescence for an even number of layers in the repression cascade, that is, a repression cascade set to "On". Dotted lines indicate the steady-state value of fluorescence for an odd number of layers in the repression cascade, that is, a repression cascade set to "Off".

Figure 3: Sensitivity analysis of the model parameter, L, which describes the rate of transcriptional leakage. Notably, as we lower the rate of transcriptional leakage, then the difference between "On" and "Off" signals increases, which is expected given that transcriptional leakage is what causes the signal to degrade.

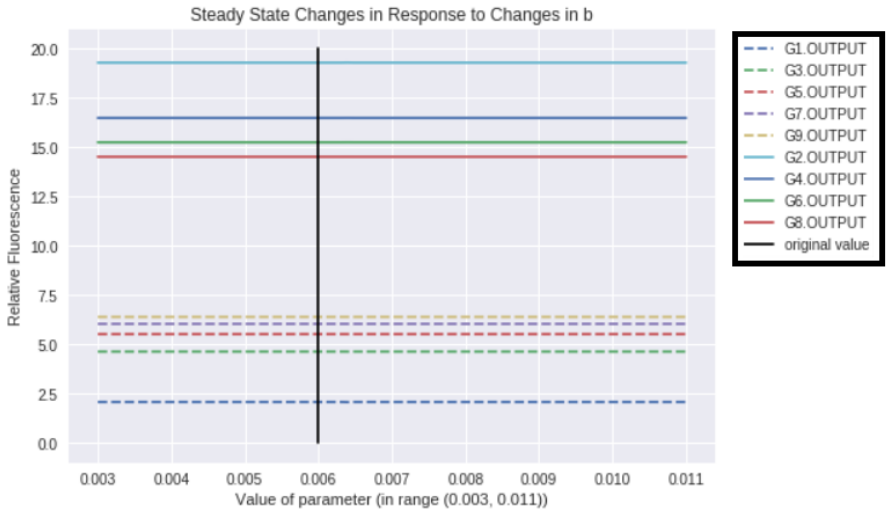


Figure 4: Sensitivity analysis of the model parameter, b, which describes the degradation rate of gRNA across all layers of the repression cascade. Perturbing the degradation rate of gRNA across all layers does not appear to affect the stead-state value of fluorescence. This may be because the model parameter b affects all layers equally, and mainly contributes to the time-frame of fluorescence, rather than the steady-state value.

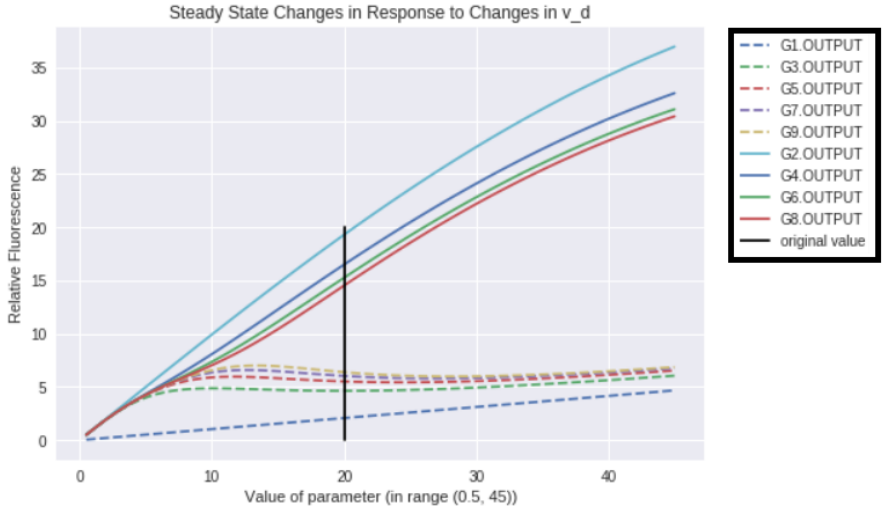


Figure 5: Sensitivity analysis of the model parameter, v_d, which describes the maximum steady-state concentration of gRNA for the d^{th} layer of the repression cascade. Increasing v_d also increases the maximum steady-state concentration at the output layer, which likely contributed to the rising maximum steady-state fluorescence for both "On" and "Off" signals.

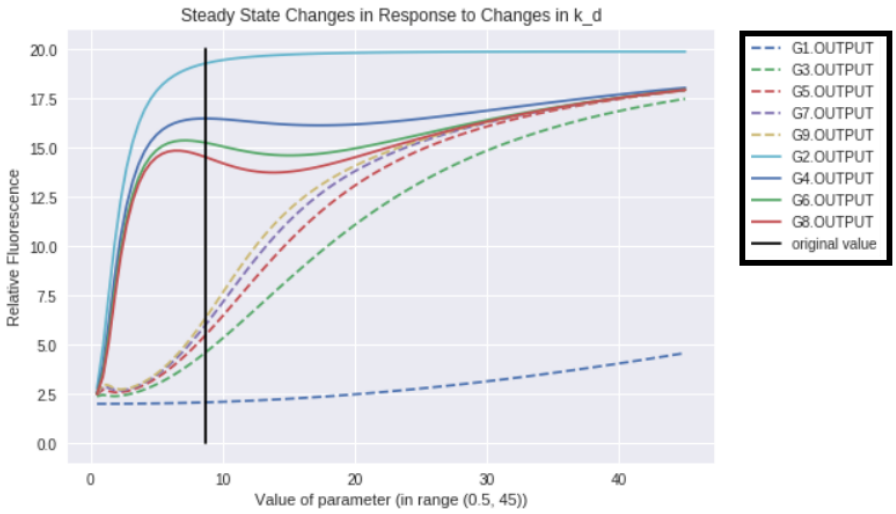


Figure 6: Sensitivity analysis for the model parameter, k_d, which describes the number of repressors required to supress the d^{th} promoter to half-strength. When k_d is too low, that is, when only a small number of repressors can repress the promoter to half strength, then the rate of transcriptional leakage becomes a large enough factor to completely degrade the output signal. When k_d is too high, that is, when a large number of repressors are required to repress the promoter to half strength, then outputs from successive gates are not strong enough to differentiate "On" and "Off" signals.

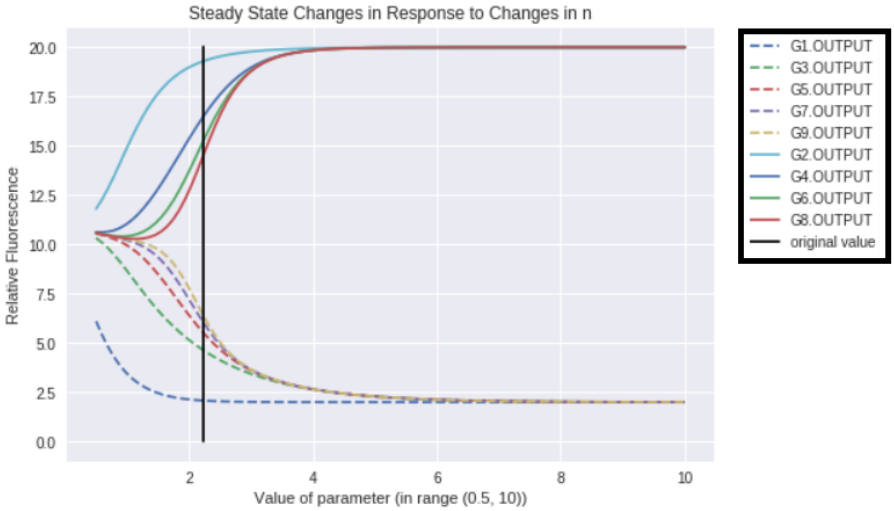


Figure 7: Sensitivity analysis for the model parameter, n, which is the Hill coefficient of the promoter. Large values of n increase the "hill-like" behavior of the hill function, meaning that "On" and "Off" signals are more distinct as noise is filtered out better.

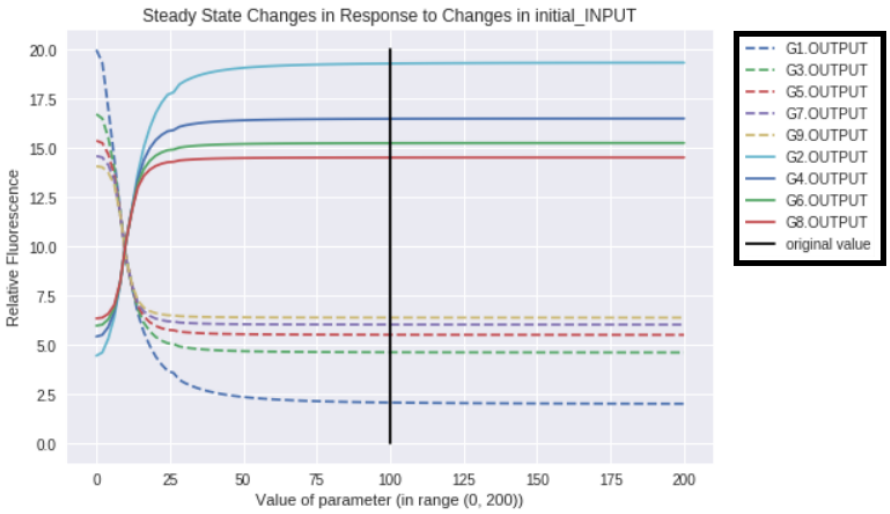


Figure 8: Sensitivity analysis for the concentration of the initial input signal. It appears that, after the concentration of the initial input signal reaches some minimum threshold, the system displays consistent steady-state behavior. This minimum threshold is likely the minimum amount of input signal to produce an output signal.

4.2 Expansion of Mathematical Model

The original mathematical model for the repression cascade describes the system behavior of a NOT gate composed of the CRISPR-*dCas9-Mxi1* NOR gates. Using the same fit model parameters, we expanded this original model from a one-input system to a two-input system – effectively forming a mathematical model for a NOR gate, rather than a NOT gate. The following differential equation describes the concentration of the output signal for a single NOR gate.

$$\dot{r_d} = b\left(\frac{v_d(1-L)}{1+\left(\frac{INPUT1}{k_d}\right)^n+\left(\frac{INPUT2}{k_d}\right)^n}+Lv_d-r_d\right)$$

Rather than describe the output signal of each "layer" of a repression cascade, in this expanded mathematical model, r_d now describes the concentration of the output signal for the d^{th} NOR gate. The other fit model parameters, v_d, L, k_d, and b are defined as before. The input parameters "INPUT1" and "INPUT2" describe the concentration of the input signal for two inputs. This form was chosen to preserve the limiting behavior of the original model. Additionally, the symmetry of the two input variables and the use of the same fit model parameters allows us to directly model the system behavior of CRISPR-*dCas9-Mxi1* NOR gates.

4.3 Applying NOR Gates to Real System

Note that NOR gates are functionally complete, and possess arbitrary connectivity. Using our expanded mathematical model of a NOR gate and the fit model parameters from the original paper, we were able to compose mathematical models for all forms of boolean logic. We then applied these mathematical models to hypothesize the system behavior of the Violacein biosynthesis pathway. The Violacein pathway has previously been shown to work when expressed using CRISPR-*dCas9* NOR gates in yeast, so we sought to investigate whether it would continue to work using CRISPR-*dCas9-Mxi1* NOR gates (Zalatan et al., 2015). Given varying concentrations of three input signals, termed "VioA", "VioC", and "VioD", the Violacein biosynthesis pathways produces fluorescence of four different colours. Additionally, given no input signal, the pathway does not produce any color.

In Figure 9, we demonstrate the default "Off" state of the Violacein biosynthesis pathway, in which no fluorescence is produced. In Figure 10, we demonstrate the four possible colors of fluorescence given varying concentrations of input signals.0

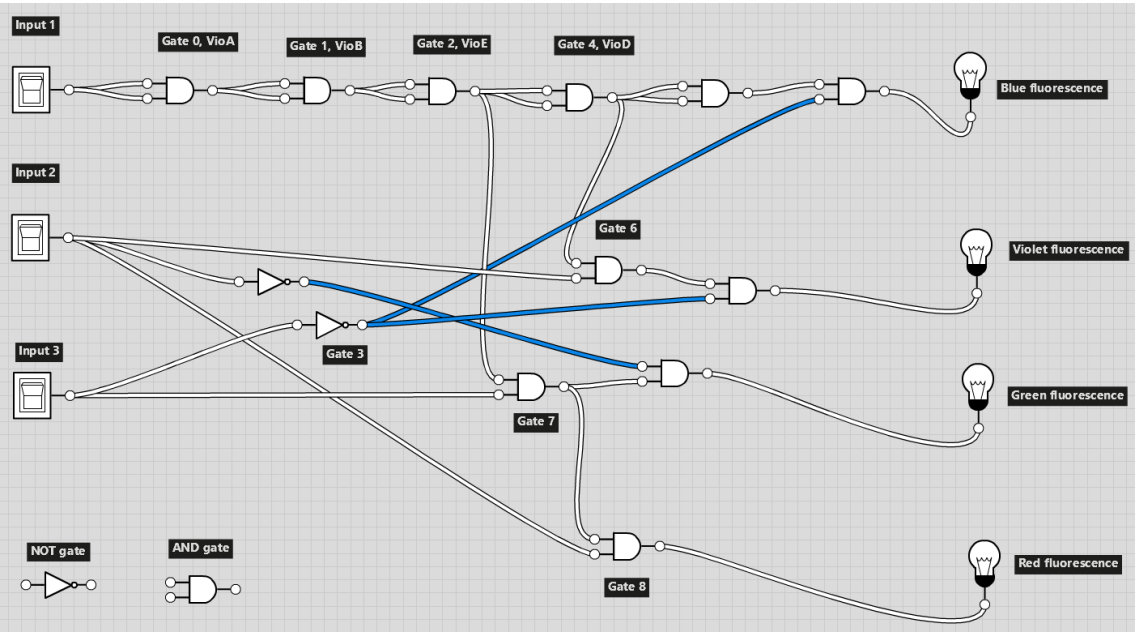


Figure 9: Logic gate representation of Violacein biosynthesis pathway. Given varying concentrations of three input signals, the pathway can produce fluorescence of four different colours. Shown here, given no input signal, the pathway does not produce fluorescence. The additional AND gates near the top were added to track the concentrations of various molecules as part of our simulation, but are not logically necessary.

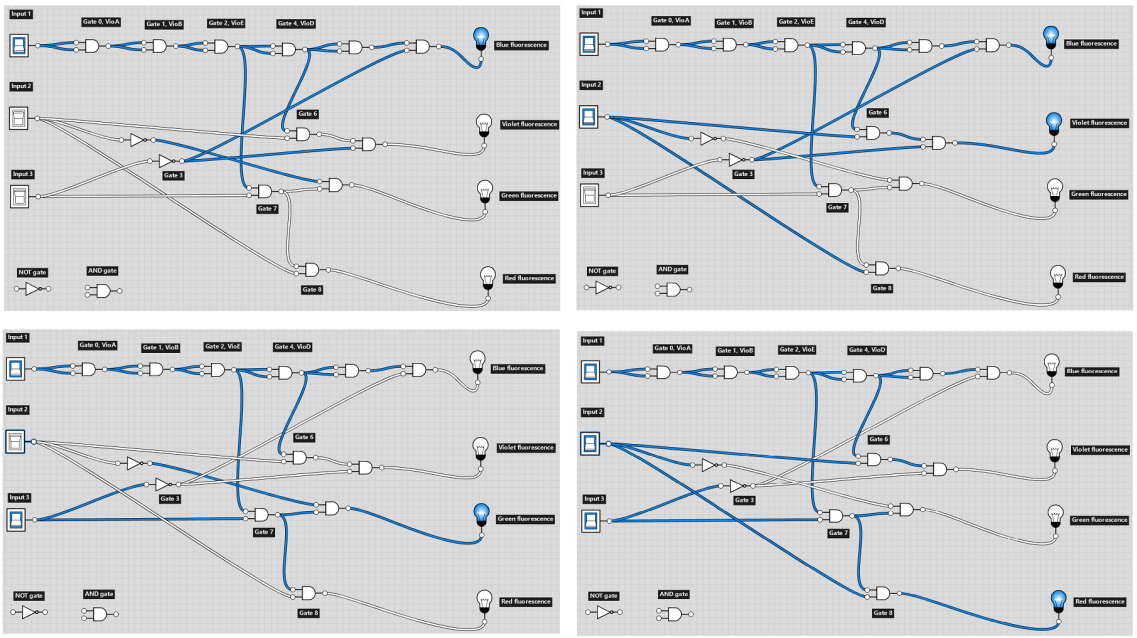


Figure 10: Logic gate representation of Violacein biosynthesis pathway for four different combinations of input signal, resulting in four distinct output signals.

In Figure 11, we juxtapose the experimental results published by Zalatan et. al and our simulated results for the five distinct outputs of the Violacein biosynthesis pathway (Zalatan et al., 2015). Note that there is general agreement between experimental results from literature and our simulated results, which suggests that we successfully simulated the macrostate of expected outputs from the Violacein biosynthesis pathway. However, our simulated results do not exactly match the experimental results. This is expected, as the experimental results are based on the use of CRISPR-*dCas9* NOR gates rather than the CRISPR-*dCas9-Mxi1* logical gates that our mathematical model is based on.

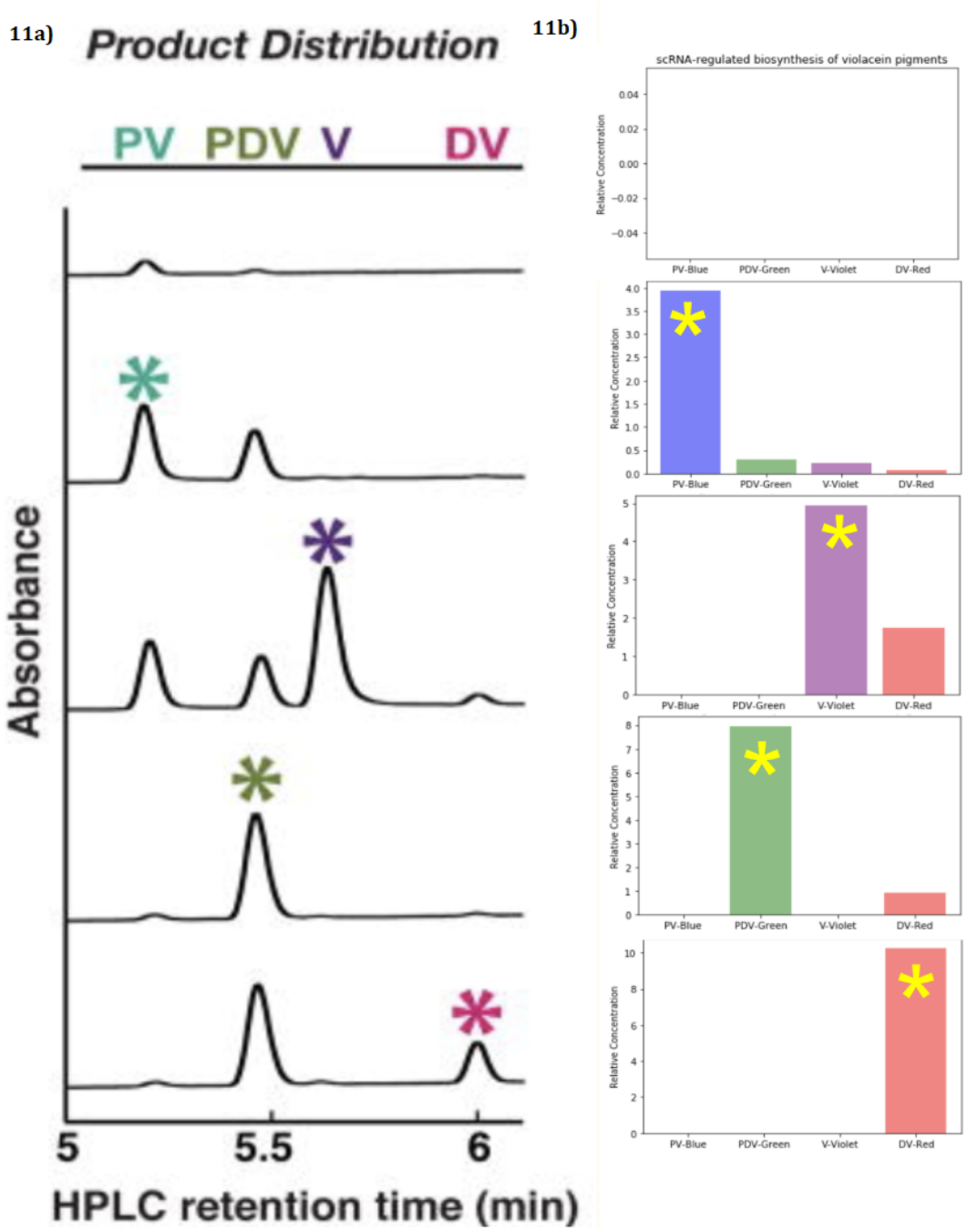


Figure 11: Fluorescence produced given five combinations of inputs, one of which being no input. **11a)** Experimental results published by Zalatan et. al, who used CRISPR-*dCas9* NOR gates to construct the Violacein biosynthetic pathway in yeast. **11b)** Simulated results based on the mathematical models for CRISPR-*dCas9-Mxi1* logical gates.

5 Discussion

We used fit model parameters to recreate the mathematical model of the regression cascade from the original paper, and demonstrated that our recreated mathematical model was able to reproduce a figure from the original paper. From this representation of a NOT gate, we expanded the mathematical model to a two-input system to represent a NOR gate. Since NOR gates are functionally complete, we were then able to compose mathematical models for all forms of boolean logic, and successfully applied these models to hypothesize the macrostate system behavior of an existing biological pathway. These results are promising, and may suggest that CRISPR-*dCas9-Mxi1* logical gates offer a greatly improved alternative over previous methods of synthetic digital circuit construction, in agreement with the findings of Gander et al.

Of note, our analysis is purely computational – further work is necessary to more rigorously interrogate the use of CRISPR-*dCas9-Mxi1* logical gates. For example, experimental results run using CRISPR-*dCas9-Mxi1* logical gates to construct the Violacein pathway may provide insight towards the accuracy of our simulations. Additionally, our simulations were run with the fit model parameters, derived from a repression cascade. A more complex digital circuit like the Violacein biosynthesis pathway may be more aptly described with different parameter values, or even a different set of parameters. Experimental results would also provide data for further estimation of model parameters.

Alternatively, rather than continuing to explore the current model, improvements could be made to existing models and methods. As shown in the parameter sensitivity analysis, as transcriptional leak approaches zero, the "On" and "Off" output signals become more distinct. This suggests that a reduced transcriptional leak rate might result in system behavior more closely resembling electronic circuits, which is the ideal goal of synthetic digital circuits. Also, a current weakness of the CRISRP-*dCas9-Mxi1* NOR gate approach is the need for a large number of individual components to compose boolean logic. Even with a low rate of transcriptional leakage, there is still a limit to the number of components that can be involved in CRISPR-*dCas9-Mxi1* digital circuits before signal degradation distorts the expected result. NOR gates are functionally complete, however, they are not highly specific. It may be possible to directly construct other forms of basic boolean logic using the CRISPR-*dCas9-Mxi1* transcriptional protein complex. Expanding the number of fundamental units to more than just NOR gates would greatly reduce the number of individual components necessary to construct complex digital circuits, which could allow for even more complex synthetic digital circuits.

6 Appendix

All of the code used for recreating figures and performing novel analysis is attached to the back of this case study. Code was written in Python, and is formatted as a Jupyter notebook for readability.

References

Elowitz MB, Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*. 403:335.

Gander MW, Vrana JD, Voje WE, Carothers JM, Klavins E. 2017. Digital logic circuits in yeast with crispr-dcas9 nor gates. *Nature communications*. 8:15459.

Guye P, Ebrahimkhani MR, Kipniss N, Velazquez JJ, Schoenfeld E, Kiani S, Griffith LG, Weiss R. 2016. Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using gata6. *Nature communications*. 7:10243.

Medley JK, Choi K, König M, Smith L, Gu S, Hellerstein J, Sealfon SC, Sauro HM. 2018. Tellurium notebooksan environment for reproducible dynamical modeling in systems biology. *PLoS computational biology*. 14:e1006220.

Nielsen AA, Voigt CA. 2014. Multi-input crispr/cas genetic circuits that interface host regulatory networks. *Molecular systems biology*. 10:763.

Roybal KT, Williams JZ, Morsut L, Rupp LJ, Kolinko I, Choe JH, Walker WJ, McNally KA, Lim WA. 2016. Engineering t cells with customized therapeutic response programs using synthetic notch receptors. *Cell*. 167:419–432.

Storn R, Price K. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*. 11:341–359.

Xie Z, Wroblewska L, Prochazka L, Weiss R, Benenson Y. 2011. Multi-input rnai-based logic circuit for identification of specific cancer cells. *Science*. 333:1307–1311.

Zalatan JG, Lee ME, Almeida R, et al. (11 co-authors). 2015. Engineering complex synthetic transcriptional programs with crispr rna scaffolds. *Cell*. 160:339–350.