REPORT

# An Experimentally Validated Mathematical Model for AHL Induced GFP Expression in Genetically Engineered Bacteria

Group # 3. A.B Silva, V. J. Tien, A. J. Clark, and S. Vasudevan

**Abstract:** Here, we develop a mathematical model for Green Fluorescent Protein (GFP) expression based on acyl-homoserine-lactone (AHL) signalling in three strains of genetically engineered bacterial cells. Parameters of the computational model were tuned so that the steady state root mean square error (RMSE) was less than 0.125 and the dynamic 2D finite difference model normalized RMSE (NRMSE) was less than 0.2 for all three receiver strains. This study was able to meet these design specifications while maintaining biological consistency of the parameters used across all three receiver strains. We found error in this model to be associated with non-radial/circular spread of the edge distance of GFP in experimental images. However, we were able to quantify these errors by performing multiple edge estimates to generate a standard error of the mean associated with every measurement. Synthetic biology may revolutionize the field of personalized medicine, by allowing for biological generation of a therapeutic molecule in response to a signalling molecule. The model generated in this study lays the groundwork for future work in this area and may prove useful in developing synthetic systems to produce a desired quantity of therapeutic molecule in response to a stimulus.

## Introduction:

Over the past two decades the field of synthetic biology has expanded significantly, and a particular area of promise is the development of designer bacteria for therapeutics.[1] These designer bacteria can be genetically modified to sense and/or produce biologically relevant chemicals.[2] Several products have already been developed to this end, such as a microbe in the intestine that can detect glucose and in return produce insulinotropic proteins and a genetically modified E. coli which produces the bacteriocin pyocin p5 in response to sensing the pathogen P. aeruginosa.[3,4] The key to the production of these designer bacteria is the creation of genetic switches. These genetic switches harness bacteria's operon model of transcription to regulate the transcription of desired genes.

While several of these designer microbes are being studied and developed, standardized methods of therapeutic delivery that perform for various detection molecules and response molecules have yet to be fully developed. In this work, we examine the possibility of creating an adaptable transcriptional circuit that can detect biochemical levels and respond, at the correct threshold of detection, with the production of a therapeutic molecule in the proper dose and time-course.

In order to test the feasibility of this new cell therapy platform, we mathematically modelled the response of an example circuit, the LuxR pathway, and validated this model with wet-lab experiments. In this LuxR testing pathway we developed, the LuxR protein is expressed via the constitutive promotion of the promoter $Promoter_{Con}$ and a Ribosome Binding Site. The LuxR protein contains a binding site for Acyl-Homoserine Lactone (AHL) and dimerizes after binding with this ligand to form the protein R. R then binds to the pLux promoter, which controls the transcription of the mRNA encoding Green Fluorescent Protein (GFP). Thus, AHL models a chemical inducer, while the expression of the detectable GFP represents the production of a therapeutic response.[5]

## Results:

The transfer function of receiver cells was first estimated using a simple Hill equation. These results are, however, not shown as the Hill model presented a few disadvantages. The Hill equation

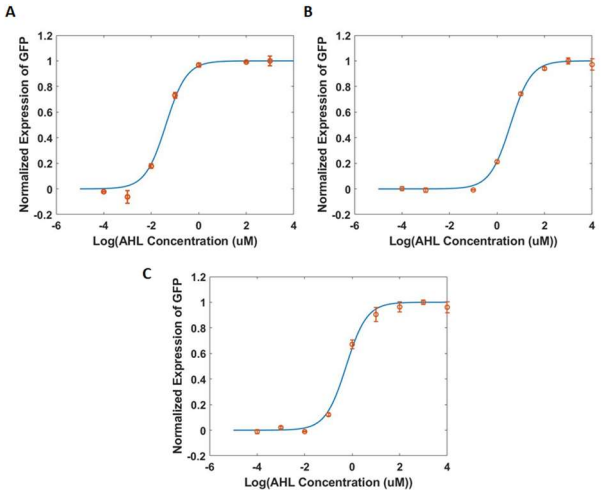


**Figure 1:** Shown is the data of normalized expression of GFP for the three strains. Strain 1 **(A)**, Strain 2 **(B)**, Strain 3 **(C )**. Note that the three plots appear to be shifted on the x-axis, with strain 1 increasing from zero first followed by strain 3 then strain 2. The data at each point represent duplicate measures of normalized GFP expression (N=2) and error is shown as plus/minus standard error of the mean. The RMSE was 0.019 for Strain 1, 0.058 for Strain 2, and 0.015 for Strain 3.

considered a single transcriptional activation threshold of dimeric R binding to the plux promoter. This does not take into consideration the full complexity of the circuit. It fails to include terms such as degradation rates, formation rates, and synthesis rates of GFP, TXGFP, and AHL that have a profound impact on the dynamics AHL-GFP

**Table 1:** Constants that were not varied between strains

| | | | | | |
|---|---|---|---|---|---|
| **Formation rate constant of dimeric R protein from LuxR protein binding to AHL** | $\rho_R$= 0.5μM-3 min-1 | **Synthesis Rate of $TX_{GFP}$** | αTXGFP = 0.05 μM/min | **Degradation rate constant of TXGFP** | δTXGFP = 0.2 $min^{-1}$ |
| **Transcriptional activation threshold of dimeric R binding to the pLux promoter** | $K_R$ = 1.3e-5 μM | **Degradation rate constant of dimeric R protein** | $\delta_R$ = 0.0231 $min^{-1}$ | **Synthesis rate constant of GFP protein** | $\alpha_{GFP}$ = 2 $min^{-1}$ |
| **Degradation rate of AHL** | 4.8135 x $10^{-4}$ $min^{-1}$ | **Degradation rate constant of GFP protein** | $\delta_{GFP}$ = 4e-4 min-1 | **Hill coefficient** | n = 0.5 |
| **Diffusion Rate of AHL** | D = .15 $mm^2$/min | **Simulation Time** | 23 hours | | |

**Table 2:** Constants that were varied between strains

| Strain | Formation rate constant of dimeric R protein from LuxR protein binding to AHL ($uM^{-3}$ $min^{-1}$) | LuxR protein concentration (uM) | $Promoter_{con}$ affinity | AHL affinity |
|---|---|---|---|---|
| 1 | 4.68 | 0.016 | High | Low |
| 2 | 4.68 | 1.82 x $10^{-4}$ | Low | Low |
| 3 | 29.11 | 1.82 x $10^{-4}$ | Low | High |

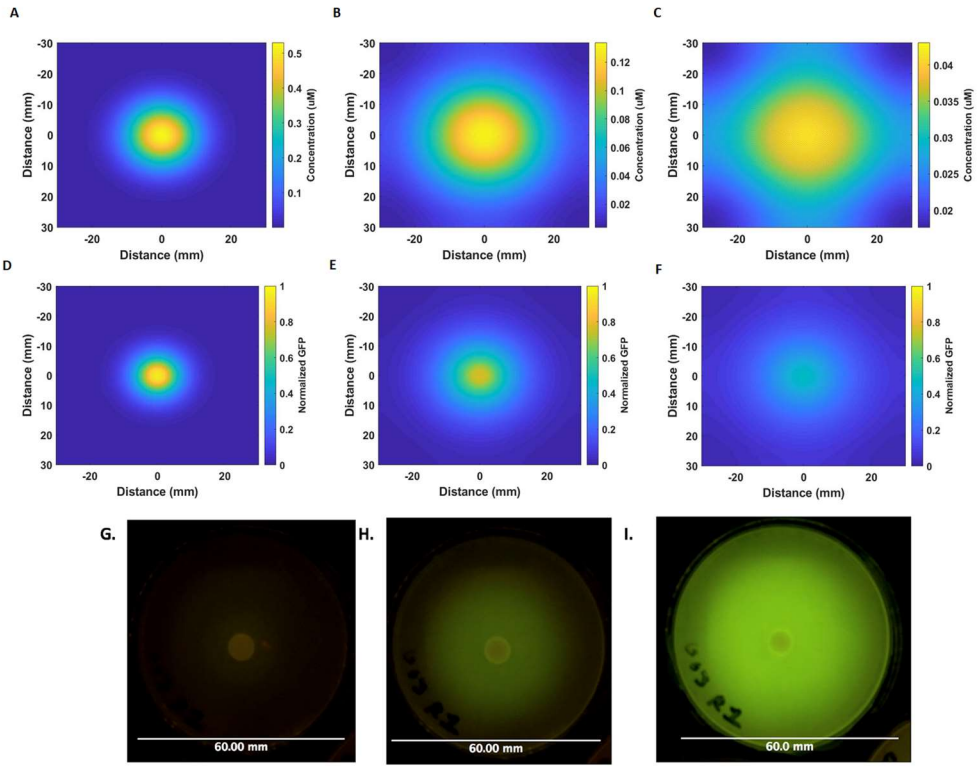


**Figure 2**. Shown are representative heat maps and images from receiver 1 used to develop the model GFP and AHL edge distance. Heat map of predicted AHL concentration at **(A)** 3 hr **(B)** 10 hr **(C)** 21 hr. Heat map of normalized GFP expression at **(D)** 3 hr **(E)** 10 hr **(F)** 21 hr. Experimental images of GFP expression at **(G)** 3 hr **(H)** 10 hr **(I)** 21 hr. Note that GFP expression spread follows the diffusion of AHL. The edge distance of the AHL (**A,B,C)** was calculated as the point where AHL concentration exceeded switching concentration for the given receiver strain. Edge distance of GFP (**D,E,F)** was calculated as the point where normalized GFP concentration exceeded a specified threshold. Edge distance of experimental data (**G,H,I)** was computed using a custom MATLAB program.

 

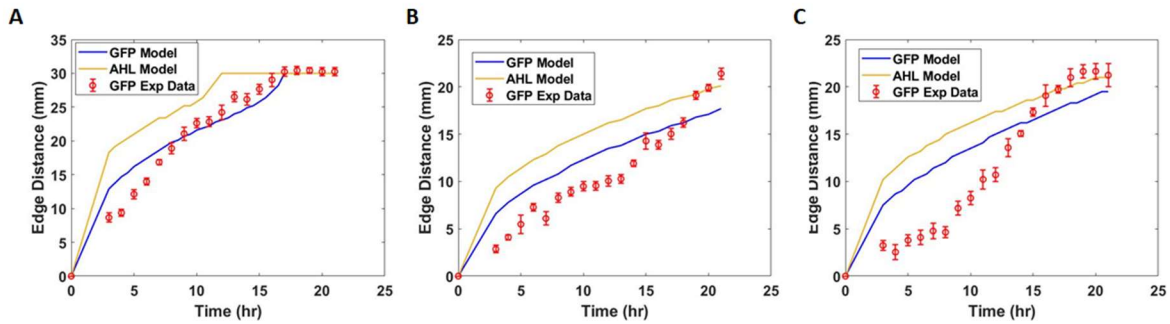


**Figure 3:** Shown is the experimental data for GFP edge distance plotted with predicted AHL and GFP edge distances from the model for each strain: Strain 1 **(A)**, Strain 2 **(B)**, Strain 3 **(C).** Experimental GFP data (N = 8) is plotted as mean ± standard error of the mean. The NRMSE was 0.0806 for Strain 1, 0.1230 for Strain 2, and 0.1286 for Strain 3.

The results of fitting a more detailed steady state transfer function (as described in methods) for the three receiver strains are shown in Figure 1. Here, a model was developed to relate normalized expression of GFP to log concentration of AHL. At steady state, the transfer functions of all three strains appeared sigmoidal and were able to be fit with parametric transfer functions. Note that the RMSE for each strain is less than 0.15 (Fig. 1). Strain 1 had the earliest switch point, at a log(AHL) concentration of approximately 0.02 uM, strain 2 at approximately 3.1 uM, and strain 3 at approximately 0.79 uM. Furthermore, when looking at non-normalized data (not-shown) strain 1 had a higher baseline GFP expression than strains 2 and 3. Steepness remained relatively constant through the three strains, and was well modeled by a hill coefficient of 0.5. The steady state models generally predicted the features of the curve well, as indicated by the low RMSE values. The model does not pick up on fluctuations in the high and low region of the curve. This, however, may be an advantage as in theory there is no reason for these fluctuations.

When fitting parametric transfer functions to receivers 1,2, and 3, some parameters were held constant across all three strains while others were varied. Table 1 shows constants that were held the same across strains. Between strains, the formation rate constant of dimeric R protein from LuxR protein binding to AHL ($p_R$) and LuxR protein concentration [LuxR] were varied given each individual bacterial strain's AHL affinity and Promoter(con) affinity respectively. Table 2 shows the values that were found to be optimal for steady state transfer functions. These results and biological differences between strains also support the above mentioned differences in steady state switching points and baseline activity. Additionally, these parameters were ensured to be consistent with the underlying biology of each strain: strain 1 being low in promoter(con) high in AHL affinity, strain 2 being low in both promoter(con) and AHL affinity, and strain 3 being high in Promoter(con) affinity but low in AHL affinity.

Using the same constants found optimal in the steady state analysis, a model was then developed to predict 2-dimensional spatial distribution of GFP over 21 hours in agarose plates (seeded with AHL soaked disc) for each receiver strain (Fig. 2). Table 1 shows the optimal values that were found for additional constants of degradation of AHL and diffusion of AHL. These constants were held the same across all three receivers. From the representative heat maps from strain 1 (Fig. 2), it can be seen that GFP expression follows AHL diffusion. Once AHL diffuses into an area and increases in concentration past the switching point of a receivers transfer function, GFP can begin to propagate into that area. For strain 1, this switching value was estimated to be approximately 0.02 uM, which is supported by the correlation between GFP and AHL heat maps (Fig. 2A-F). It can also be seen that over time GFP is activated in more areas; however, due to the degradation term, GFP degrades in areas it has already diffused to (Fig 2F).

The final readout of the model created for each receiver strain was edge distance of GFP and AHL versus time (Fig. 3). It can be seen that, as shown by the heat maps, GFP edge distance closely follows AHL edge distance. The experimental data of GFP edge distance, quantified in MATLAB, was compared to the model theorized GFP edge distance at time points 3 to 21 hours every hour. Timepoints up to 21 hours were used in order to utilize all available experimental data. For all three receivers the Normalized RMSE (NRMSE) was less than 0.2. Here inaccuracy in the experimental data was quantified with the standard error of the mean. Two individual plating experiments were conducted and four measurements per time period were taken (in the four cardinal directions) of the resulting GFP

diffusion. The results from the two individual experiments were combined to yield 8 replicates per time point. The average standard error of the mean across all time points was 0.5151, 0.7256, 0.8485 for strains 1,2, and 3 respectively.

## Discussion:

This study demonstrated the successful modeling of a genetic circuit based on AHL induced expression of GFP. The results of this study show strong efficacy, NRMSE <0.2, with the limitation of only two plating trials for time modeling and one plating for steady state estimation. Furthermore, the parametric defining of this model allows it to be utilized to predict the output of a potentially therapeutic molecule, modeled by GFP, to an input signaling molecule, modeled by AHL. In the design of our model, biological consistency of parameters was ensured. As stated in results, the only two parameters changed between strains were the formation rate constant of dimeric R protein from LuxR protein binding to AHL ($\rho_R$) and LuxR protein concentration [LuxR]. These constants represent AHL affinity and Promoter(con) affinity respectively. Since strains that have the same affinity level (low or high) have the same DNA, the same values for low affinity ($\rho_R$) or [LuxR] were used across strains. This ensured biological consistency. Furthermore, if any other constant was changed, they were changed across all 3 strains given that there was no biological reason to expect differences with regard to non ($\rho_R$) or [LuxR] constants. Within these biological constraints, parameters were optimized by first optimizing all constants for strain 2 (which had low affinity for LuxR and r) by minimizing RMSE. From there, the study chose to optimize the high affinity parameter for Strains 1 and 3 while holding the low affinity parameter fixed from the Strain 2 optimization.

A few sources of error contributed to the non-trivial NRMSE values found in the experiments. Error in measuring edge distance for the experimental data from the images collected amounted to an average standard error of the mean of 0.5151 mm, 0.7256 mm, 0.8485 mm for strains 1,2, and 3 respectively. This is in line with the visual observation that the strains did not follow a perfectly circular outgrowth of GFP. In addition, when optimizing the additional parameters for the 2-D time model, the study was limited by the computational power available. Given that MATLAB code took a great deal of time to run (approximately 2 min per simulation), only a limited amount of combinations of parameters were explored (computation power was a limiting factor). Thus, a more optimal combination may exist.

Future studies should examine this genetic circuit in more detail. It would be desirable to have the capability to conduct more than two 2-D time analysis plating trials and more than 1 steady state plating trials. If N was increased above ten for both of these conditions, the relative variation and magnitude of error would decrease. Future studies should also utilize more sophisticated MATLAB imaging processing, such as smoothing and automated edge detection, to better mitigate the error of experimentally quantifying GFP edge. Lastly, more time and computational power should be devoted towards precisely optimizing parameters, within the accepted biology, to minimize NRMSE.

However, the implications of this study are incredibly robust. Such a model has many important applications and provides a foundation upon which to design new programmable cell therapeutics and predict their dynamic response. Potentially, similar mathematics can be used to tune the production of therapeutic molecules to various signals in the cell. This allows for the potential of personalized medicine based on needs of individuals for biologically created therapeutics in response to a biological signal. One example of how this study fits into the broader literature is the application of gene therapy for the treatment of cancer.[6]

## Materials and Methods:

### AHL Induction of LuxR in Receiver Strains:

A 10 mM stock N-(3-Oxohexanoyl)-L-homoserine lactone (AHL) solution was serially diluted with sterile water to concentration in log 10 increments between 10 mM and 100 pM (9 increments). 2 mL of each receiver strain (R1, R2 and R3) of NEB Turbo Cells (E. coli, pre-grown to OD ~ 0.6 in M9 media + Chloramphenicol) were induced with 20 uL of each of these 9 AHL concentrations as well as a 0 mM negative control. Receiver strains were induced with AHL during the mid log phase (OD~ 0.6) of cell growth (the mid log phase of cell growth represents exponential growth of bacteria and is the idea time to ensure efficient AHL binding along with bacterial cell propagation and survival). The bacterial strains used in this study were grown in the presence of chloramphenicol to ensure that the desired genetic mutations were present in the bacteria, as these mutations were introduced with a mutation conferring chloramphenicol resistance. Additionally, the bacterial strains used in this study were grown in the presence of chloramphenicol to prevent the growth of other undesired bacteria. Each sample was then

 

dispensed in duplicate into 48-well deep well plates, which were then covered with a Breathe-Easier sealing strip. The plates were shaken at room temperature (~21 ºC, 250 RPM) for 20 hours, and then placed in a 4ºC fridge for 5 days. This minimally impacted results as GFP expression reaches steady state during shaking.

**Modeling the transfer function of the Receiver:**
The predicted steady state concentrations of GFP $TX_{GFP}$ and R at various concentrations of AHL were found by solving equations two, three, and four (shown below) at steady state (setting the derivatives equal to 0).[5]

$$Equation\ 1:\ \%\ GFP\ Expression = \frac{1}{1+\left(\frac{K_R}{[Signal]}\right)^{n1}}$$

$K_R$ is transcriptional activation threshold of dimeric R binding to the pLux promoter, [Signal] is the concentration of AHL in micromolar, and n1 is the hill coefficient.[5]

$$Equation\ 2:\ \frac{d[R]}{dt} = (\rho_R[LuxR]^2[AHL]^2 - \delta_R[R])$$

d[R]/dt is the rate of change of R protein concentration with respect to time in units of micromolar per minute, $\rho_R$ is the formation rate constant of dimeric R protein from LuxR protein binding to AHL, [LuxR] is the concentration of LuxR protein in micromolar, [AHL] is the concentration of AHL in micromolar, $\delta_R$ is degradation rate of dimeric R protein, and [R] is the concentration of R protein in micromolar.[5]

$$Equation\ 3:\ \frac{d[TX_{GFP}]}{dt} = \left(\frac{\alpha_{TX_{GFP}}\left(\frac{[R]}{K_R}\right)^{n1}}{1+\left(\frac{[R]}{K_R}\right)^{n1}}\right) - \delta_{TX_{GFP}}[TX_{GFP}]$$

d [$TX_{GFP}$] / dt is the rate of change of concentration of mRNA encoding for GFP in units of micromolar per minute, $\alpha_{TX_{GFP}}$ is the synthesis rate constant of $TX_{GFP}$ (mRNA encoding for GFP), [R] is concentration of R protein in micromolar, $K_R$ is transcriptional activation threshold of dimeric R binding to the pLux promoter, n1 varies from one to four, and $[TX_{GFP}]$ is the concentration of mRNA encoding for GFP in micromolar.[5]

$$Equation\ 4:\ \frac{d[GFP]}{dt} = (\alpha_{GFP}[TX_{GFP}] - \delta_{GFP}[GFP])$$

d[GFP]/dt is the rate of change of the concentration of GFP in units of micromolar per minute, $\alpha_{GFP}$ is the synthesis rate constant of GFP protein in $min^{-1}$, $[TX_{GFP}]$ is the concentration of mRNA encoding for GFP in micromolar, [GFP] is the concentration of GFP in micromolar.[5]
Refer to tables 1 and 2 for values of constants used.

**Measuring the Steady-State Transfer Function in Liquid Culture:**

After retrieval from incubation, the samples were resuspended and their fluorescence and optical density was measured. M9 media + Chloramphenicol was used as the blank. The GFP fluorescence was measured by spectrophotometry (SpectroVis Plus, Vernier, Beaverton, OR) at $\lambda_{Excitation}$= 500 nm and $\lambda_{Emission}$= 540 nm with a sample time of 50 ms, while optical density was measured at λ= 600 nm with a sampling duration of 4s and a sampling rate of 2 samples/second. A transfer function for each strain was developed by plotting the average GFP fluorescence, normalized to maximum GFP fluorescence, for each AHL concentration (N=2). Normalizing the data to the max has the shortcoming of losing information about the absolute expression range. This may be especially pertinent to drug delivery where the absolute range has a greater impact than the relative normalization. Outliers were excluded based on the following criteria. The mean of the replicates at each AHL concentration was first calculated along with the standard error of the mean. If the standard error of the mean for any replicate was greater than three times the average standard error of the mean across all replicates, that time point was excluded and labeled an outlier. The RMSE between the experimental and model transfer functions was calculated using the formula shown below.[5]

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}$$

RMSE was minimized while maintaining biological significance of parameters (optimization further discussed in Results section above).

**M9 Plate Preparation and Receiver Transduction:**
60 mm M9 media plates supplemented with chloramphenicol were warmed in a 37°C incubator. NEB Turbo Cells of the three strains were pre-grown in M9 media and chloramphenicol to an OD~0.6. 2 mL of each strain were mixed with a 2 mL aliquot of 0.7% agarose, warmed in a 55°C water bath. Each solution was mixed and poured onto a corresponding plate, which was tilted to create an even layer of agarose. The plates were allowed to dry for ~15 minutes. 10 µL of 10 µM AHL was added to filter paper disks of 5 mm diameter, which were then placed in the center of each plate with sterile forceps. Condensation was removed from the lid with a Kimwipe and the plate was flipped upside down to prevent future condensation. The plate was taped around the rim edge with black masking tape to preserve data quality.

**Image Acquisition and Analysis:**
The plates were set lid side down on the automated fluorescence imager which consists of a Raspberry

Pi computer and camera synced to a transilluminator (University of Pennsylvania Department of Bioengineering, PA). This system allowed for hourly images at three exposure times (0.5s, 1s and 2s). The exposure time of 0.5s demonstrated the best results and was therefore used for the following analysis. The resulting images were analyzed for estimated edge distance via a custom MATLAB program. The edge distance of the GFP fluorescence at each of four cardinal directions was calculated and averaged at every hour from 3 to 21 hours.

**Computational Model Development:**
A finite difference model for GFP concentration based on AHL diffusion in 2D was created. The following equation described AHL diffusion.[1]

$$A^{t+1}{}_{x,y} = A^{t}{}_{x,y} + \frac{\Delta t}{\Delta p^2} D\left(A^{t}{}_{x-1,y} + A^{t}{}_{x+1,y} + A^{t}{}_{x,y-1} + A^{t}{}_{x,y+1} - 4A^{t}{}_{x,y}\right) - \Delta t(d_{AHL} * A^{t}{}_{x,y})$$

$A^{t}_{x,y}$ is the concentration of AHL in micromolar at a specific time point (min) and x,y location (mm) inside of the plate, dAHLis the degradation rate of AHL in $min^{-1}$, $\Delta t/\Delta p^2$ is the stability factor (equal to 0.25), and D is the diffusion rate of AHL in $mm^2/min$.[5]

The resulting theoretical GFP concentration was calculated using equations two, three, and four (shown above). The additional parameters of the 2-D model (degradation and diffusion rates of AHL) were optimized by minimizing the NRMSE between the experimentally calculated GFP edge distance and the edge distance derived from the analytical model. The model edge of GFP expression was calculated by normalizing the GFP concentration values at all locations to the max GFP value and finding the distance from the center to the location where normalized GFP concentration exceeded a specified threshold value. The edge of AHL in the model was defined as where AHL concentration was equal to the steady-state switching point of the corresponding transfer function.

## Conclusions:

This study resulted in creating a mathematical model for therapeutic designer bacteria that successfully predicted edge distance of GFP in response to AHL for all three receivers within an NMRSE of 0.2: 0.0806 for Strain 1, 0.1230 for Strain 2, and 0.1286 for Strain 3. This was done while simultaneously ensuring steady state RMSE did not exceed 0.15. In designing this model, it was ensured that biological significance of parameters was preserved between strains. Thus, our model represents a predictive method for therapeutic molecule generated in response to a biological signal. The efficacy of the model can further be improved by increasing computational power and applying advanced image processing algorithms for processing experimental data. The work presented in this study contributes toward the larger effort in the field of synthetic biology to derive models to quantify the amount of relevant therapeutic biomolecule produced for a given amount of signal.

**Supplementary Information (SI):**
Source code used in developing the finite difference model and performing edge detection is included in the supplement. Also, we include the code used to perform image analysis in MATLAB.

 

## References:

1 D. Ewen Cameron, Caleb J Bashor, James J Collins, *Nature Reviews Microbiology*, 2014, **12**, 381-390
2 Jan Claesen and Michael A. Fischbach, ACS Synthetic Biology 2015 ***4*** (4), 358-364
3 F. Duan, K. L. Curtis, J. C. March *Appl. Environ. Microbiol.* 2008, **74** (23) 7437-7438
4 Saeidi, N., Wong, C. K., Lo, T.-M., Nguyen, H. X., Ling, H., Leong, S. S. J., Poh, C. L., and Chang, M. W, *Mol. Syst. Biol.,* 2011*,* **7**
5 Lab 2: Synthetic Biology*, University of Pennsylvania Department of Bioengineering*, 2019, 1-11.
6 Alexei Tsygvintsev, Simeone Marino, Denise E. Kirschner, *Mathematical Methods and Models in Biomedicine*, 2013

 