eIRXIV EMERGING INVESTIGATORS PREPRINT SERVER

December 18, 2025

Large-Scale Screening of *E. coli* Promoters for Small Molecule Biosensor Development

Saanvi Dogra*[1], Jason Gao*[1], Dishti Wadhwani*[1], Risha Guha†[1], Nithika Vivek†[1], Lauren Chen[1], Anwita Bandaru‡[1], Shawn Kim‡[1], David Lanster[2, 3]

* Authorship arranged in alphabetical order. These authors contributed equally to this work.
† Authorship arranged in alphabetical order. These authors contributed equally to this work.
‡ Authorship arranged in alphabetical order. These authors contributed equally to this work.
[1] Del Norte High School, San Diego, California
[2] Department of Chemistry, The Scripps Research Institute, La Jolla, CA, 92037
[3] Skaggs-Oxford Doctoral Program in Chemical and Biological Sciences, The Scripps Research Institute, La Jolla, CA, 92037

SUMMARY

The field of synthetic biology makes significant contributions to healthcare, environmental engineering, and technology through the manipulation of cellular macromolecules and whole organisms. Oftentimes, these advancements are dependent upon biosensors to report on an activity of interest within a cell or to detect extracellular cues and report on them in a measurable way. This project was undertaken as part of iGEM 2024 (Internationally Genetic Engineered Machines Competition 2024) centered around the choice of 10 small molecules related to environmental and human health with the goal of developing transcriptional biosensors to report on their concentrations. Each molecule was screened against a library of over 2000 promoter-GFP constructs in search of promoters responsive to each molecule. Further, a Deep Learning model was used to predict active promoter-molecule pairs and in silico putative hits from the screen were analyzed with molecular docking. While no robust biosensor hits were found for the molecules of interest, our work demonstrates a useful pipeline for further small molecule biosensor development.

INTRODUCTION

Biosensors are analytical devices that elicit a measurable signal in response to specific biological processes or the presence of a molecule. Using these tools, scientists have made crucial breakthroughs in disease diagnosis and drug detection (such as measuring the remaining medicine in a bloodstream), environmental monitoring, food control, forensics, and biotechnology [1]. Biosensor development is often the first step in the detection and biodegradation of toxic chemical compounds, such as residual pesticides on fruits and

vegetables, microbial degradation products, and the presence of Persistent Organic Pollutants (POP). In a laboratory setting, an ideal biosensor elicits a response proportional to the amount of analyte present through the action of a “transducer” that recognizes the analyte and generates a measurable output.

This study was specifically focused on developing allosteric transcription factor (TF)-based biosensors in E. coli. Currently, there is a lack of well-established biosensors for numerous health and environmentally relevant molecules. Therefore, the goal of this project is to identify TF-promoter pairs that respond to several different molecules as a starting point for future biosensor development. TFs are able to bind small molecule ligands and, as a result of this binding, can impact the transcription of target genes. TFs rely on promoter sequences to control their responses. Promoters are DNA sequences that regulate when and how strongly a gene is expressed. In bacteria, biosensors often produce Green Fluorescent Protein (GFP) under the control of a specific promoter as the measurable output. GFP produces a fluorescent signal that can be measured to determine if the promoter driving its expression is responsive to small molecule induction [2]. A 2006 paper from the Weizmann Institute of Science presents a comprehensive library of ~2000 transcriptional promoters from Escherichia coli K12 called the Horizon Promoter Collection (HPC). E. coli is a well-studied gram-negative bacterium commonly used for biosensor development since the strain is easily cultured and manipulated, and nonpathogenic, making it an ideal biosensor chassis [3]. Each promoter in the HPC carries a unique promoter (covering most well-known E. coli promoters) cloned to drive GFP from a low-copy number plasmid. This collection presents an exciting opportunity to explore a multitude of different molecules and promoter combinations to aid in developing biosensors.

Ten molecules were chosen to test with each of the promoters in the HPC because of their relevance to human and environmental health. Development of biosensors for these molecules would prove useful in commercial and clinical settings. The molecules chosen were: carbaryl (CAR), 3-phenoxybenzoic acid (PBA), lovastatin (LOV), butanoyl-homoserine (BHL), phenylglyoxylic acid (PGA), propoxur (PRO), perfluorooctane sulfonate (PFS), cis-naphthalene dihydrodiol (CHD), diethyl phthalate (DEP), and tartaric acid (TAR). Carbaryl is a man-made pesticide toxic to insects and over 190 registered pesticide products contain it [4]. 3-phenoxybenzoic acid is a prominent environmental contaminant because it is a degradation product of pyrethroid insecticides [5]. Lovastatin is a medication that is used to lower cholesterol levels in blood and help prevent cardiovascular diseases such as heart attacks or strokes [6]. Butanoyl-homoserine lactone is a signaling molecule used in quorum sensing and is used in synthetic biology to engineer bacterial systems with customized gene expression profiles [7].

Phenylglyoxylic acid is a breakdown product of styrene, a material used to make plastics and rubber, making it an exposure biomarker used in environmental and occupational health to monitor the level of styrene and other related compounds [8]. Propoxur is a carbamate insecticide used to control a variety of pests; however, long term exposure can cause decreased cholinesterase levels in humans [9]. Perfluoro octane sulfonate is a man-made surfactant and global pollutant as it can cause cancer and developmental toxicity [10]. Cis-Naphthalene dihydrodiol is an intermediate in the microbial degradation of naphthalene, an insecticide, making it a useful biomarker [11]. Diethyl Phthalate is a synthetic substance used to make plastic more flexible, and it easily contaminates the environment when diluted with other liquids [12]. Tartaric acid is a type of alpha hydroxy acid naturally found in many plants, and it is commonly used to generate carbon dioxide. It acts as a muscle toxin by inhibiting the production of malic acid [13]. Cumulatively, these representative compounds span diverse molecular and functional classes and make for compelling screening candidates.

RESULTS

High-Throughput HPC Screen

To establish a basis for this project, an initial high-throughput screen was performed to nominate any potential promoters responsive to each inducer molecule. This screen was conducted by testing each of the ten molecules against the entire promoter library. The testing was carried out in 96-well plate format, with cells grown in LB and a single inducer at 500 mM. After incubation for 24 hours, fluorescence was measured with a plate reader and wells with the highest fluorescence were nominated for downstream studies. The 8-12 promoters that produced the most fluorescence greater than 1.35 AU (average sfGFP/OD600 divided by average promoter value for each molecule) for each of the molecules were identified (**Table 1**).

Because of the scale of the initial screen, it was undertaken at only one concentration of inducer and with only one replicate of each strain. Therefore, any putative hits were followed up using a titration of the inducer molecules and in triplicate. The strains were titrated across eight concentrations from 0 to 10 mM of their corresponding molecules to examine how the fluorescence varies with concentration. We expected to see a dose-dependent response in fluorescence to increasing inducer molecules. While many promoters did not follow the expected trend, some were indeed titratable: BHL with the P_{ydeI} promoter produced a 1.7 times fold increase, CND with the P_{ybcK} promoter produced a 2.2 times fold increase, CND with the P_{aegA} promoter produced a 2.8 times fold increase, and DEP with the P_{yfiF} promoter produced a 2.0 times fold increase from the lowest to highest measurement (**Figure 2**).

To test the specificity of our results and take the background activity of each promoter into account, we compared the molecule-specific sfGFP value to the average sfGFP value across the other 9 screened molecules. In one example, a CAR-induced 260% increase of grxA sfGFP/OD600 over the average sfGFP/OD600 indicates molecule-specific activation of this promoter (**Figure 3**).

We next sought to pursue optimization of the biosensors by exploring alternative plasmid architectures including changing the origin of replication to vary copy number, the fluorescent protein, and its ribosomal binding site. We generated and annotated a plasmid map with gadB fused to sfGFP (responsive to Butanoyl-Homoserine Lactone) and indicated the plasmid structure, promoter regions, and replication origins for use (**Figure 4**). While we did not have time to clone new variants of the plasmid, the map may be a useful starting point for others continuing our work.

Machine Learning

A deep learning multi-stream neural network was used to predict whether a given molecule-promoter pair produces a fluorescent response. After training, the model was able to correctly classify a molecule-promoter pair as fluorescent or non-fluorescent 96% of the time, indicating precise and consistent results.

Training, validation, and test data for this model were all molecule-promoter combinations that belong to one of two classes: fluorescent response or non-fluorescent response. Class distributions of the inputs to the machine learning model show an even distribution of data belonging to each class — ~48 and ~52 for non-fluorescence and fluorescence, respectively (**Figure 6C)**. This indicates that the data was well-split and the model remained unbiased towards either class.

To understand the impact of adding DNA promoter sequences to the multi-stream neural network, the two machine learning networks (one with the sequences and one without) were compared. Firstly, the capabilities of each network were compared to understand their accuracy. Both loss graphs decrease over multiple epochs, indicating that both models achieved a higher accuracy as they iterate over the data. Both graphs eventually plateau, demonstrating that each model achieves its best accuracy by the end of its training. The validation data line continues to remain close to the training, indicating that the model was able to achieve a balance between learning and generalizing without overfitting. An accuracy of 96% (>80%) and low loss of both models also rule out the possibility of the model underfitting.

Our next goal was to compare the two accuracies and capabilities of the models with and without the DNA sequences. In the model without the DNA data, training and validation loss plateaued at a higher value than in the model with DNA data (**Figure 6**). Loss and accuracy are inversely proportional. Plots of the loss curve without DNA data (**Figure 7A)** and the loss curve with DNA data (**Figure 7B**) plateau at ~5 and ~0.05, respectively, suggesting a lower accuracy for the model without nucleotide sequences.

A Receiver Operating Characteristic (ROC) plots the False Positive Rate vs the True Positive Rate. The ROC curve of the predictions generated by the model (**Figure 6D)** has an area under the curve of 0.96, which represents the testing accuracy for the model. The high area under the curve also showcases the ability of the model to maximize the True Positive Rate while minimizing the False Positive Rate.

Additional statistics depict the effect of adding DNA sequences on the accuracy of the model and a dropout is used to mitigate any possible overfitting. Dropouts remove random subsets of neurons in the network during each iteration throughout the training process. This ensured that the model did not memorize the training data. As demonstrated by statistical analysis of model outputs, though training accuracy drops slightly, the validation and testing accuracies improve in both iterations of the model (with and without DNA data), suggesting that the use of dropout and regularization increases the generalization of the data, creating a more useful and wide-scale model. Finally, we present the accuracy metrics as a percentage (ranging from 75-95%). The data supports the conclusions made above: the model is more accurate when DNA data is used and the data is regularized due to the incorporation of more data that better aids the model in making a more informed decision of the fluorescence of a molecular-promoter pair (**Figure 7**).

Modeling

A molecular docking process was utilized to better understand the physical interactions behind potential DNA-TF pairs by visualizing the DNA-protein interactions between transcription factors and their cognate promoters. P_{gadE}, P_{gadB}, and P_{YdeO} were some of the promoters identified to have the largest fold increase with increasing concentrations of the respective molecule they detect (**Table 1**). The YdeO promoter controls expression of gadB, part of the acid-resistant GAD pathway. Structural modeling of P_{ydeO} binding to its transcription factor, GadB, reveals the nucleotides critical for stable binding and the position of the alpha helix oriented perpendicular to the DNA backbone (**Figure 5**). In the future, this information could allow for targeted mutagenesis of the transcription factor DNA binding domain or the promoter

sequence to enhance binding and improve the dynamic range of the sensor. Additionally, the modeling helps our understanding of small molecule-protein interactions, which can guide future biosensor development by identifying which molecular structures are more compatible with which compounds.

Simulation of the Nac transcription factor showed binding to a double helix present on the structure of P_{ybcK} (**Figure 5**). Nac controls biotic and abiotic stress tolerance, and overexpression of Nac through cell engineering approaches can improve stress tolerance. Nac regulates P_{ybcK} by activating it, increasing stress response and tolerance. Similarly, the DNA-binding regulator YidZ binds to P_{rof} (**Figure 5**). When YidZ is activated under stressed conditions, P_{rof} is activated, which ensures that the genes responding to environmental stress are transcribed correctly in order to properly limit the biological impact of the stress.

DISCUSSION

This study used a library of E. coli promoters to develop biosensors capable of detecting various molecules. By screening approximately 6,720 promoter-molecule combinations, we identified promoters that may respond to exposure to some of our molecules of interest. Notable results include the PgadB-YdeO promoter-transcription factor interaction, which achieved a 2.8-fold increase in fluorescence over background fluorescence calculated in the screen and the PgrxA-CAR pairing, which showed a 260% specificity increase over competing molecules. These findings establish a foundation for further biosensor development.

A key contribution of this study lies in validating the fluorescence trends across varying molecule concentrations. The consistently increasing fluorescence supported the hypothesis that these molecules act as inducers for GFP transcription. However, certain inconsistencies in fluorescence levels at intermediate concentrations indicate the need for replicates to enhance the robustness of these trends. Environmental conditions, including media composition and incubation duration, could also influence fluorescence and warrant further standardization.

The incorporation of a multi-stream neural network significantly advanced our ability to predict fluorescence outcomes for promoter-molecule pairings. The inclusion of promoter DNA sequences notably enhanced model accuracy, yielding a high ROC AUC of 0.96, which underscores the model's reliability. This approach provides a scalable framework for predicting biosensor viability beyond the molecules tested here.

Despite these successes, several limitations are present in this study. First, while the study utilized a diverse range of molecules, expanding the scope to include additional compounds could refine the model’s generalizability. Second, while dropout regularization

improves model performance, further optimization of hyperparameters may enhance accuracy and reduce potential overfitting. Additionally, the docking simulations, although informative, were constrained to specific DNA-protein interactions. Broader simulations could provide deeper insights into the mechanistic interactions underlying the observed fluorescence.

The identified promoter-molecule pairs provide a starting point for developing tailored biosensors applicable in fields ranging from environmental monitoring to healthcare diagnostics. For instance, the PgrxA promoter's specificity to carbaryl suggests its potential in detecting pesticide residues. Similarly, the robust performance of the PgadB-ydeO interaction in acidic conditions highlights its utility in studying stress-response pathways in bacteria.

Future research should focus on validating these findings through in vivo studies and exploring their applications in real-world scenarios. Additionally, expanding the dataset and refining machine learning algorithms will enhance predictive accuracy, enabling more efficient biosensor design. Ultimately, this study underscores the potential of synthetic biology in addressing pressing environmental and healthcare challenges through innovative biosensor technologies.

MATERIALS AND METHODS

Media and Chemicals

LB Broth medium was created using 50 g yeast extract, 50 g peptone, and 25 g sodium chloride to 5 L of water in a volumetric flask and mixed with a magnetic stir bar. The medium was divided into several bottles and autoclaved. One milliliter of 1000X Kanamycin (25 mg/mL) was added to each autoclaved 1 L bottle before use. The minimal M9 medium was prepared by adding 52 g M9 salts (KH_2PO_4 at 15 g/L, NaCl at 2.5 g/L, Na_2HPO_4 at 33.9 g/L, NH_4Cl at 5 g/L) into 4.9 L of DI water. The medium was autoclaved and stored at room temperature. Then, 20 g filter-sterilized glucose was added to the M9 salt solution to a final concentration of 20 mM. A 10mL $MgSO4_4$ (2 mM) and 0.5ml $CaCl_2$ (0.1 mM) was added to the 5 L solution before use.

Based on their available quantities, the molecules were dissolved into individual solutions. A 500mM solution for 3-Phenoxybenzoic Acid (128.4 mg in 1 mL DMSO), Lovastatin (242.7 mg in 1 mL DMSO), Propoxur (125.4 mg in 1 mL ethanol), and Perfluorooctane Sulfonate (247.6 mg in 1 mL water) was created. Diethyl Phthalate was already in a solution of 121.2 µL. A 1 M solution of Tartaric Acid (150 mg in 1 mL water), Carbaryl (201.22 mg in 1 mL DMSO), Butanoyl-Homoserine Lactone (171.19 mg in 1 mL DMSO), Phenylglyoxylic Acid (150.13 mg in 1 mL DMSO), and Cis-Naphthalene Dihydrodiol (162.16 mg in 1 mL water).

Strains

All strain handling was done in 96 well plates. To each well of the plates, 250µL of LB Broth with kanamycin was added using a multichannel pipette. The promoter collection is supplied in twenty-one 96 well plates. To inoculate the strains, one of the twenty-one plates was removed from the -80° freezer. Sterile tips attached to a Gilson PlateMaster machine were lowered onto the 96-well strain plate, moved side to side to collect some bacteria, then lifted and lowered onto one of the 96-deep-well plates with the media. The pipettes were mixed in the broth to distribute the bacteria. A gas permeable covering was rolled onto the top of each of these plates and they were grown overnight in an incubator at 37°C with 900 RPM orbital shaking to encourage *E. coli* growth.

Fluorescence Assays

One mL of 1 mM concentrated molecule stocks were added to 200mL of M9 media. Each of the stocks were then thoroughly mixed with the M9 media through using stir bars and heating up to dissolve the concentrated molecule stocks. Using a multichannel pipette, 275 µL of each molecule M9 solution was added to each well of the different 96-deep-well plates. For each plate, a Gilson PlateMaster was used to add 2.75 µL of cells with promoters into each well containing molecules. Each promoter was matched to each different molecule for one iteration. The PlateMaster was also used to transfer 150 µL from the assay deep well plates into a 96-well black walled, clear bottom plate. The 96-well plates were inputted into the Plate Reader at 37°C with shaking for 10 seconds to measure absorbance in each well using a 600nm wavelength and fluorescence with 485 nm excitation wavelength and 510 nm emission wavelength. The resulting data was analyzed to select the 8-12 promoters with at least a 1.35 average sfGFP/OD600 divided by average promoter value for each molecule.

Dose Responses

500µL of 1x Kanamycin LB broth was added to each well of 96-deep-well plates using a multichannel pipette. The startup procedure was repeated on only the most fluorescent promoters selected, with autoclaved toothpicks used to transfer cells into the wells. After the strains grew in the incubator overnight, 2.75µL cells/well were added to another 96-deep-well plate. For titration, serial dilutions of eight concentrations of the molecule (0 nM, 10 nM, 100 nM, 1 µM, 10 µM, 100 µM, 1 mM, and 10 mM) with three replicates for each were prepared for each molecule-promoter pair. After 24 hr growth the plates were analyzed the same as above. The

trends of fluorescence with increasing concentration of the respective molecules were analyzed to identify the best promoter to detect each.

Machine Learning

Using data produced from wet lab experiments, we created and trained a multi-stream neural network machine learning model that predicts if a molecule-*E. coli* promoter combination generates a fluorescent response.

Our input variables consisted of molecule name, promoter protein sequence, promoter DNA sequence, and molecular structure. Our output variable was the significance (fluorescent response) of the molecule promoter pair. Class distributions were analyzed to ensure that the class split was close to 50-50 to ensure equal data feeding of each class. Each input variable in text form was transformed into numerical values. Molecule name was assigned a value between 1 to 10. Promoter protein sequence was encoded into a numerical value using label encoded and promoter DNA sequence was encoded using a hash function which condensed 10,000 base-pair long DNA sequences into a value less than 1000. Molecular structure images were stored as NPZs for the model to analyze. All input variables were normalized and divided by a common number to contain ranges from 0 to 1. Train, test, and validation datasets were generated by grouping the original data into 70-20-10 split randomly.

The model was first tested without DNA data of each promoter sequence. Loss results were obtained from this model. Then the DNA data (from EcoCyc) was added to obtain new loss results. The model was designed to combine both the image and tabular inputs by building a multi modal pipeline. The image was analyzed using a convolutional neural network and the tabular data was analyzed using a dense neural network. The two modes were trained independently on the surface level and then combined to analyze trends in both simultaneously. The model outputs a probability between 0 and 1 for the pair to produce a fluorescent response. Thresholds >0.5 for class 1 and <0.5 for class 0 were defined to ensure a binary output. The model was then tested on the generated test data to ensure little to no overfitting. An ROC curve (a plot of the number of false positive results generated by the model versus the number of true positive results when testing data was used) was created to visualize the results.

Modeling

To model DNA-protein interactions, three primary promoters were selected for their high titration levels with BHL in our research: P_{ybcK}, P_{gadB}, and P_{rof}. Using the Ecocyc database, we identified transcription factors that bind to these promoters. Then, the nucleotide sequences of

the promoters and the protein sequences of the transcription factors were input into AlphaFold to generate predictions of DNA-protein complexes. The resulting models were visualized in PyMol to identify and analyze the binding interfaces.

REFERENCES

[1] Bhatia, D., Paul, S., Acharjee, T., & Sundar Ramachairy, S. (2023, July 23). Biosensors and their widespread impact on human health. Science Direct. Retrieved September 14, 2024

[2] Feng Y, Xie Z, Jiang X, Li Z, Shen Y, Wang B, Liu J. (2018) The Applications of Promoter-gene-Engineered Biosensor, Sensors, 18(9)

[3] Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nature methods, 3(8), 623–628.

[4] Carbaryl | C12H11NO2 | CID 6129. (n.d.). PubChem. Retrieved September 14, 2024

[5] 3-Phenoxybenzoic acid | C13H10O3 | CID 19539. (n.d.). PubChem. Retrieved September 14, 2024

[6] Lovastatin. (n.d.). MedlinePlus. Retrieved September 14, 2024

[7] PubChem. (n.d.). N-Butyrylhomoserine lactone. PubChem.

[8] Benzoylformic acid | C8H6O3 | CID 11915. (n.d.). PubChem. Retrieved September 14, 2024

[9] Propoxur | C11H15NO3 | CID 4944. (n.d.). PubChem. Retrieved September 14, 2024

[10] PFOS (Perfluorooctane Sulfonate or Perfluorooctane Sulfonic Acid) - Proposition 65 Warnings Website. (n.d.). P65Warnings.ca.gov. Retrieved September 14, 2024

[11] (1R, 2S)-cis 1,2 dihydroxy-1,2-dihydronaphthalene: Uses, Interactions, Mechanism of Action | DrugBank Online. (n.d.). DrugBank.

[12] Diethyl Phthalate | C12H14O4 | CID 6781. (n.d.). PubChem. Retrieved September 14, 2024

[13] L-Tartaric acid | C4H6O6 | CID 444305. (n.d.). PubChem. Retrieved September 14, 2024

Figures and Figure Captions

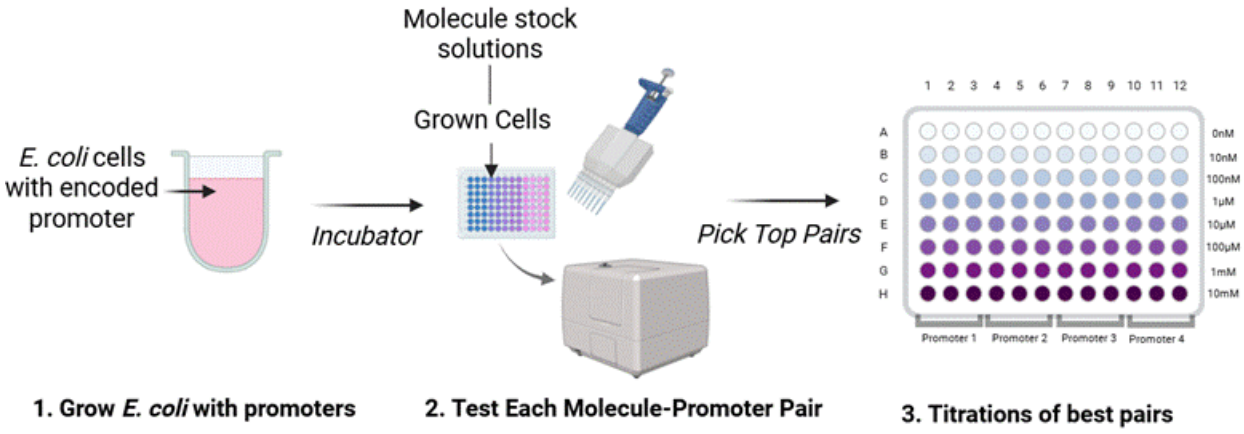

Figure 1. Schematic of the methodology used in this research

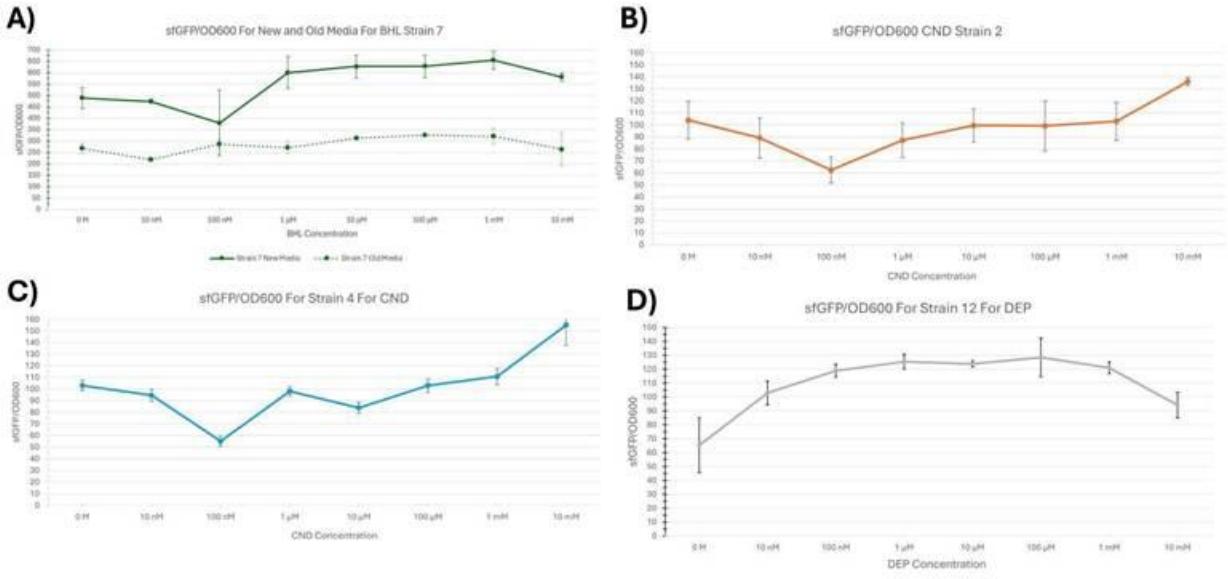

Figure 2. sfGFP/OD600 trends across increasing concentrations from 0 M to 10 mM of the molecule for the four best performing molecule-promoter combinations, **A)** BHL with the P_{ydel} promoter, **B)** CND with the P_{ybcK} promoter, **C)** CND with the P_{aegA} promoter, **D)** DEP with the P_{yfiF} promoter. The error bars represent ±1 SD.

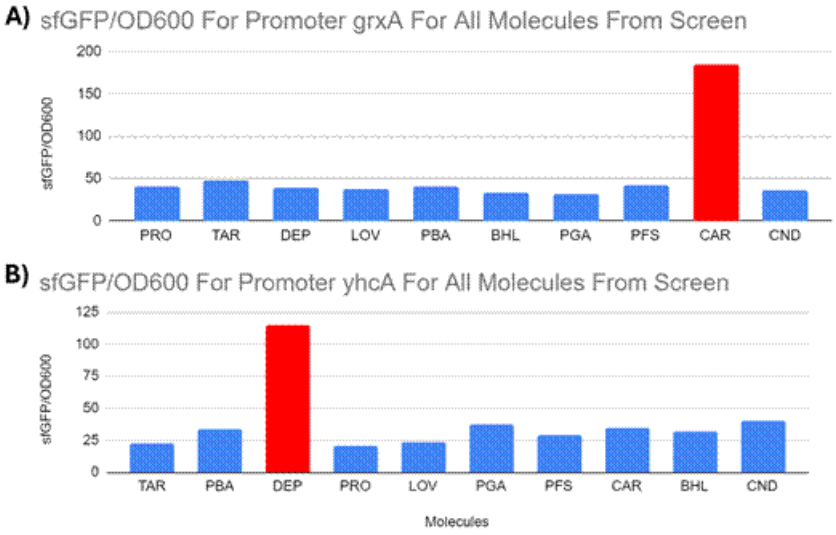

Figure 3. The sfGFP/OD600 value for the **A)** P_{grxA} promoter and the **B)** P_{yhcA} promoter across all ten molecules screened. CAR produced the one of the largest fluorescent signals with grxA and DEP produced one of the largest fluorescent signals with yhcA, with both values highlight in orange in their respective graphs.

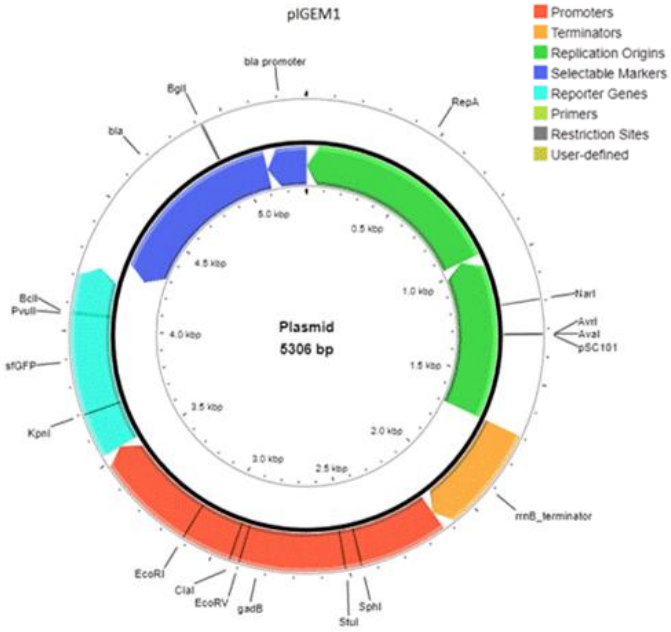

Figure 4. Plasmid map created with gadB fused to sfGFP to serve as a potential biosensor for Butanoyl-Homoserine Lactone when transformed into *E. coli.* The structure indicates the plasmid structure, promoter regions, and replication origins.

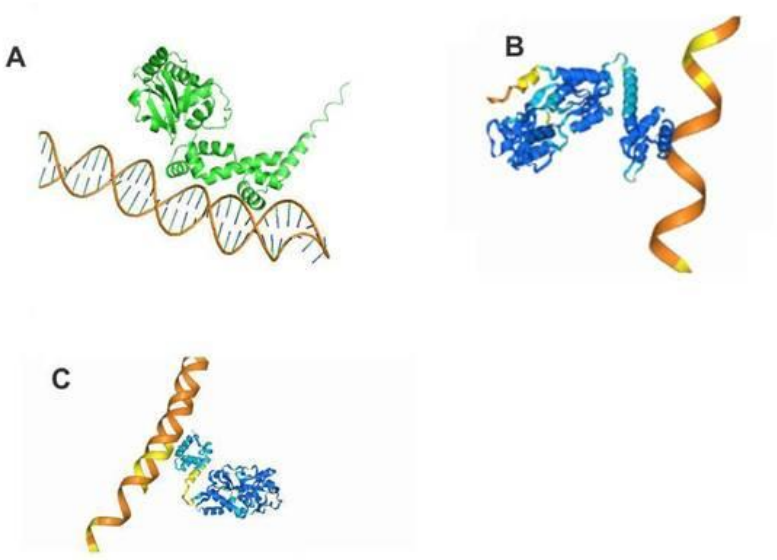

Figure 5. Modeling of the interactions between proteins and transcriptional factors. **A)** Binding of transcriptional factor YdeO(green) to gadB, where YdeO activated the transcription initiation for gadB. **B)** Binding of transcriptional factor Nac(blue) to ybcK, where Nac initiated transcription. **C)** Binding between transcriptional factor YidZ to P_{rof}, where YidZ inhibits transcription initiation

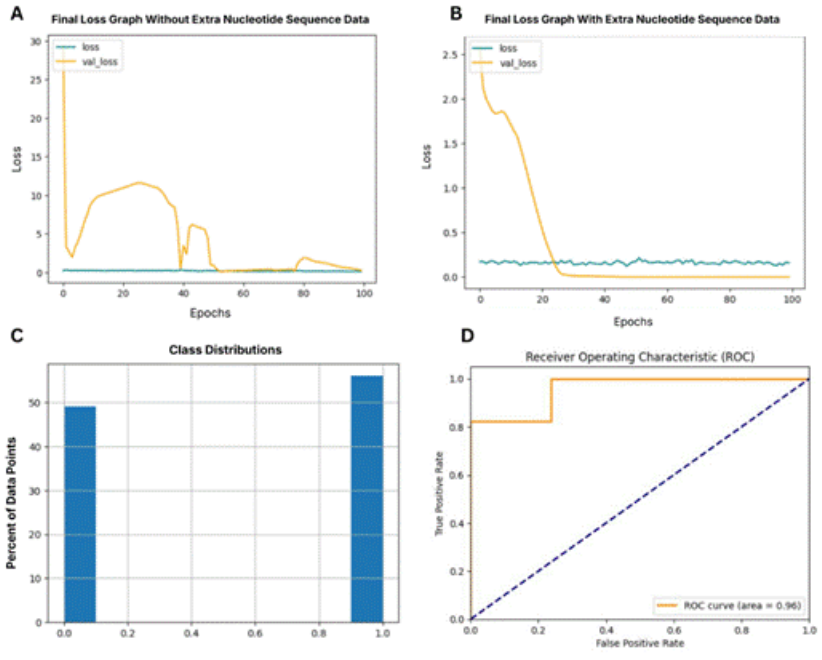

Figure 6. Training and validation loss, represented by the teal and orange lines respectively, **A)** without and **B)** with the addition of nucleotide data from EcoCyc. **C)** Class distributions (binary)

of data (fluorescent and non-fluorescent), **D)** Receiver Operating Characteristic curve and area based on the true and false positive rates

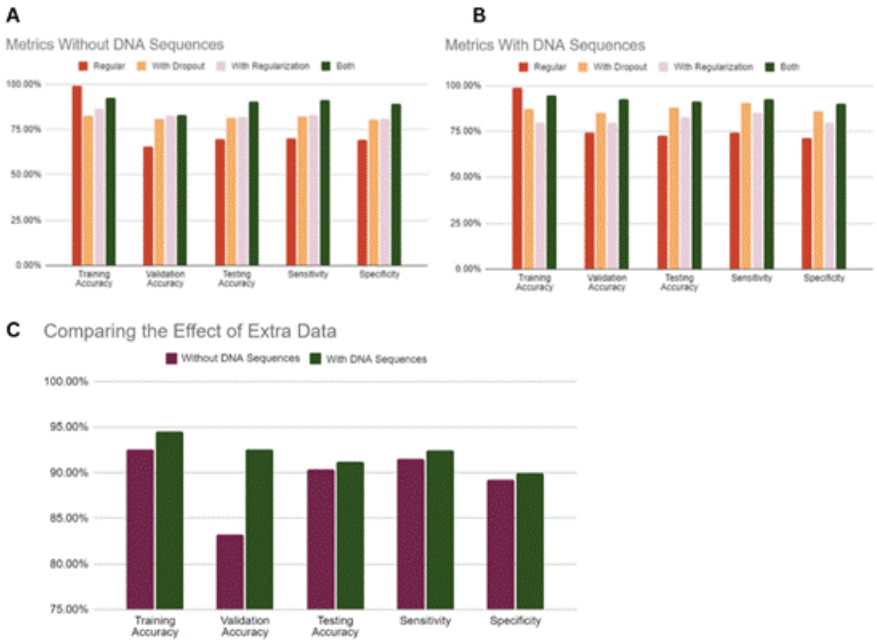

Figure 7. Training, validation, and testing accuracy + sensitivity and specificity of model with varied dropout and regularization levels (red represents regular, yellow represents inclusion of dropout, purple represents inclusion of regularization, and green represents both) **A)** without and **B)** with addition of nucleotide sequences to training data, respectively. **C)** Comparison of training, validation, and testing accuracies + sensitivity and specificity after addition of DNA data with purple presenting lack of DNA sequences and green representing the presence of them

Tables with Captions

PBA	LOV	PRO	DEP	TAR	CAR	BHL	PGA	PFS	CND
2.08 fadB	2.84 map	1.83 mngR	2.67 yhcA	2.13 deoC	4.29 ycfR	1.89 eutB	2.01 alaS	2.45 yagK	1.97 yffH
2.07 glgS	1.85 rmf	1.73 gcl	1.57 yraR	1.75 yciG	3.52 recA	1.83 gadW	1.70 glyA	1.45 yffH	1.59 ybcK
1.80 ycfR	1.81 ompX	1.57 ypdA	1.56 def	1.71 ydcJ	2.82 grxA	1.56 lpxC	1.52 yhjY	1.40 adhE	1.45 asnA
1.63 cyoA	1.73 rbsD	1.52 yacH	1.53 prpR	1.67 alsB	2.75 rrnD	1.53 gadB	1.52 ygiW	1.39 fadB	1.44 aegA

1.60 yrbL	1.73 ydcJ	1.51 metK	1.53 yaeH	1.55 yjbJ	1.62 gsk	1.51 yqjF	1.52 thrU	1.37 hslV	1.37 sieB
1.60 ycjM	1.70 alsB	1.48 pppA	1.53 rof	1.54 tnaC	1.52 yqjF	1.51 b4283	1.48 yciG	1.36 yeiP	1.37 metC
1.56 yibL	1.60 cspB	1.45 yaaJ	1.50 yhbX	1.52 uspG	1.51 aspA	1.50 ydeI	1.45 ychH	1.35 fdoG	1.37 U139
1.56 ygeY	1.58 nmpC	1.42 yaaW	1.50 slp	1.50 ygjH	1.49 hyfR	1.46 yeiP	1.44 ybhQ	1.35 ypfG	1.36 rihC
1.51 ycjG	1.55 rpmE	1.40 glnK	1.42 ygjG	1.50 pitB	1.47 aes	1.42 deoB	1.40 sodC	—-	—-
1.42 aldH	1.55 yhjY	1.40 ylbE	1.42 cca	1.49 prpR	1.45 rfe	1.42 ygiW	1.39 hupA	—-	—-
1.40 yafS	1.53 cspD	1.40 ykgM	1.39 yhfG	1.49 ygeH	1.45 eaeH	1.38 ycdZ	1.39 galR	—-	—-
1.40 hscB	1.52 uspF	1.39 yedP	1.39 yfiF	1.46 rhsD	1.41 U139	1.38 mglB	1.38 rssB	—-	—-

Table 1. The sfGFP/OD600 divided by average promoter value and promoter name of the best 8-12 promoters that produced the highest fluorescence for each of the ten molecules. PBA represents 3-Phenoxybenzoic Acid, LOV represent Lovastatin, PRO represents Propoxur, DEP represents Diethyl Phthalate, TAR represents Tartaric Acid, CAR represents Carbaryl, BHL represents Butanoyl-Homoserine Lactone, PGA represents Phenylglyoxylic Acid, PFS represents Perfluorooctane Sulfonate, and CND represents Cis-Naphthalene Dihydrodiol. To establish a basis for biosensor development, a series of experiments were conducted using the library of promoters and these ten molecules that we selected to determine which combination was most effective by observing their fluorescence. While the level of fluorescence in our final strains had some drops and inconsistencies at certain concentrations, the overall increasing trend indicated that these strains can be further researched to be eventually developed into effective biosensors.