ABSTRACT

Title of Dissertation: DEVELOPMENT OF COMPUTATIONALLY AIDED METHODS FOR INTERPRETING MOLECULAR INFORMATION

Kayla Erin Chun
Doctor of Philosophy, 2025

Dissertation directed by: Professor William E. Bentley

Fischell Department of Bioengineering

Modern electronics have enabled a highly connected world due to the immense amount of information and communication that is readily available. Comparatively, biological information lies within biomolecular states and their transitions rather than in electrons and photons. Thus, rendering the process of understanding and extracting biological information difficult and time consuming; often requiring labor intensive and complex analytical methodologies. The development of rapid approaches for designing, controlling, and monitoring biologics are critical to both clinical and industrial applications of biotechnology. Redox reactions are ubiquitous across biology and offer a bridge to electronically connect into biologics through their electron exchanges. By leveraging redox molecules to interface between biomolecular and electronic communication, we can implement electronic signaling in biological systems

to rapidly monitor and induce functional changes. Further, we can incorporate computational methods to model biological dynamics and parse electronic outputs from these systems, providing approaches for improved biological design and control.

This dissertation focused on developing synthetic means for information propagation within biological bacterial co-cultures that made use of multi-modal molecular signals. We showed how molecular information could be initiated and propagated through these co-cultures, and controlled through modulating species composition. A graph network model was then developed for simulating multi-modal signaling in synthetic bacterial consortia to study the effects of spatial conformation within co-cultures on information transmission. Finally, we developed electrochemical methods for rapidly extracting biological information using redox-mediated "probing". A novel machine learning methodology was implemented to extract biological information from this electronic data. These methodologies were developed for specific industrial and clinically relevant use cases such as: 1) Biomanufacturing of antibody therapeutics, to quantify a critical nutrient, L-cysteine, and antibody fragmentation as a marker for product quality; and 2) Female reproductive health monitoring, to classify vaginal microbiome states, where dysbiosis is linked to higher risk of health complications such as pelvic inflammatory disease and preterm labor. In sum, this work contributes molecular and electronic tools for designing and monitoring complex biological systems to predict responses and enhance clinical and industrial manufacturing applications.

DEVELOPMENT OF COMPUTATIONALLY AIDED METHODS FOR INTERPRETING MOLECULAR INFORMATION

by

Kayla Erin Chun

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:
Professor William E. Bentley, Chair
Professor Amy Karlsson, Dean's Representative
Professor Elebeoba May
Professor Sara Molinari
Professor Gregory F. Payne

Dedication

In loving memory of Doris Fong and Ying Kwan Chun.

Acknowledgements

Thank you to my advisor Dr. William Bentley, for your mentorship and never-ending kindness. Your support and guidance have pushed me to grow academically and as a person. Thank you to my dissertation committee members: Dr. Amy Karlsson, Dr. Elebeoba May, Dr. Sara Molinari, and Dr. Gregory F. Payne. Specifically, Dr. Payne- thank you for serving as an unofficial co-advisor to our lab, your curiosity and insight has been incredibly valuable to my journey. Thank you to Dr. Jenna Mueller for your feedback and time as a committee member for my proposal exam.

I am thankful for having the honor of working amongst so many wonderful lab mates throughout the years. To Kristina, Jinyang, Sohyun, Eric, Dana, Sally, and Chen-Yu Chen, thank you for mentorship and collaboration as my lab seniors. To Rahma, thank you for supporting me both in and out of lab. I couldn't have asked for a better person to have gone through the last six years with; from becoming cat parents to climbing it has been a blast. To Futoon, thank you for bringing such bright energy and delicious baking skills to our group, I will never forget your laoganma gruyere sourdough. To Divya, having you in our cohort has always been a joy but you joining our lab truly changed things for the better, thank you for your support and company on many adventures. To Varun, it was truly a blessing when you joined the lab, thank you so much for your help in these final months. To Evelyn, thank you for all your time and dedication to our projects, I always appreciated the energy you brought to the lab. To Monica, Caitlyn, Nick, Briahna, Ben, Alex and Chen-Yu Tsao, thank you for all the feedback over the years, it was a pleasure working with you all. I would also like to thank Dr. Eunkyoung Kim for all of her electrochemistry help and kind advice throughout the years.

I owe much of my success and truly the completion of this endeavor to my family. To my mother, Elizabeth, and father, Glenn, thank you for supporting my pursuits despite them moving me to a different coast. To Auntie Rosita, thank you for your encouragements throughout my lifelong academic journey. To my many siblings and their partners, Vincent and Jenny, Gregory and Holly, Curtis and Christina, Janine and Ryan, Ryan and Jenn, thank you for cheering me on and bringing me the joy of being the "fun" childless Aunt to my many niblings (Coconut and Foxy included). To my current human niblings, Elisa, Jace, Vahn, Ethan, William, Sophia, and Tyler, thanks for all of the joy you bring to my life, I hope to inspire you with my efforts put forth here to pursue whatever in life you choose. I will always support y'all the best I can.

I am incredibly grateful to have had family close by during my time in Maryland. Lauren, Rachael, and Haley—it has been a pleasure to connect with y'all. To Auntie Margaret, thank you for your kindness and generosity. I will always remember the support y'all have given me while at this home away from home.

To my lovely friends back on the west coast: Emily, Debby, Gabe, Arvin, Cesar, J, Marissa, Kayleb, Cynthia, Jason, Samira, Peter, and Kay—thank you for so much.

The completion of this dissertation could not be done without the community I have gained during my time in Maryland. To the original 9613 house occupants: Anthony, Suhas and Neha—thank you for the many days and nights of shenanigans and the undying love you all have for Gordo. To Shruti, thank you for your ride-or-die support throughout the years. To my queer sports besties: James, Gabby, and Sam—thank you for all of the fun times and encouragement. To my not queer sports bestie Seun, thanks for all of the spots. To Ben Shaw, thank you for your camaraderie on the dance floor of As you are. To Keshav, I thank COMBINE for bringing us together and you for your endless kindness.

Lastly, I would like to thank my partner Michael Brown and my cat son Gordo for being the best boys on the block. Thank you for your love and care.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

List of Appendix Figures

List of Abbreviations

AI-1: Autoinducer-1

BV: Bacterial vaginosis

CM: Conditioned Media

CST: Community State Type

CV: Cyclic voltammogram

CVM: Human cervicovaginal mucus

DMEM: Dulbecco's Modified Eagles Medium

LDA: Linear Discriminant Analysis

NCA: Neighborhood Component Analysis

MAE: Mean absolute error

MEP: Mediated electrochemical probing

MLP: Multi-Layer Perceptron

PBS: Phosphate buffered saline

PCA: Phenazine-1-Carboxylic Acid

PCA: Principal Component Analysis

PYO: Pyocyanin

RF: Random Forest

RFU: Relative Fluorescence Units

QS: Quorum sensing

Chapter 1 Introduction

Communication exists anywhere information resides; this is most blatantly observed in the electronic communications of recent history. We currently operate in a world where most devices can be monitored and controlled, enabling the "Internet of Things" where everything is connected via electronic communications [1, 2]. However, the ability to electronically plug into biological communications to access information is limited. In biological contexts, information often resides in biomolecular form, such that communication occurs through cellular transmission and reception of signaling molecules [3-5]. These signaling molecules come in a variety of structures and functions. In this dissertation, we employed an engineered bacterial signaling toolbox comprised of hormone-like small molecules alongside redox active phenazines to investigate dynamics of multiple signaling modalities in microbial consortia. We also turned to redox active molecules for electronic monitoring and control of biologics as they can mediate electron transfer between electrodes and cells. These capabilities allow us to electrically induce gene expression as well as measure redox activity within a sample [6-9]. We build upon these electrogenetic systems as well as leverage the native redox capacity of biologics to extract electrical information from relevant biological systems.

In combination with experimental methods, we utilized computation to investigate both the information flow within biological systems as well as parse the electrical information extracted from them. By simulating biological communication systems, we gathered insights to the signal dynamics not measurable by current experimental approaches. Additionally, we applied machine learning to discern electrical signals derived from relevant biological systems for industrial and clinical diagnostics. Both of these implementations of computation aided characterization of

biological communication contribute to the development of high throughput methodologies for design, prediction, and interpretation of biomolecular information.

Specifically, we developed a coarse-grain graph-based network model derived from existing electrogenetic systems. We then applied this model to simulate alternative spatial conformations of these multi-modal co-cultures to analyze communication within bacterial systems. To gain information on the communication within biological samples, we gathered biomolecular information via electrochemical probing and employed machine learning to translate these electrical measures into comprehensible biological information.

Chapter 2 Background

2.1 Quorum sensing and phenazine synthesis by *Pseudomonas aeruginosa*

Quorum sensing (QS) is the mode by which bacteria coordinate gene activation across populations through small molecule expression; these small hormone-like molecules used for signaling are referred to as autoinducers. Autoinducers are typically produced at a basal level by all members of the microbial population, such that as population density increases so does the concentration of autoinducer. At a threshold concentration, gene activation occurs throughout the population allowing for a mass triggering of induced behavior. This behavior was first observed in gram-negative bioluminescent marine bacterium *Vibrio fischeri*, which colonizes the light organ of squid.[4] At high densities, the bacteria accumulate autoinducer 3OC6-HSL produced by LuxI, an autoinducer synthase, as depicted in **Figure 2.1A**. The autoinducer then binds LuxR, a DNA binding transcriptional activator, and initiates expression of the *lux* operon resulting in bioluminescence and further upregulation of autoinducer expression initiating a positive feedback loop [4, 5]. Many other gram-negative microbes contain LuxI/LuxR-type proteins and communications using autoinducers. Although most communication is specific to intraspecies signaling (**Figure 2.1B**), crosstalk between autoinducers and transcriptional activators across species does exist in addition to utilization of multiple QS circuits within one species [4, 10, 11].

In *Pseudomonas aeruginosa*, there are four autoinducer quorum sensing circuits, LasI/LasR, RhlI/RhlR, IQS/IqsR, and PQS/PqsR as depicted in **Figure 2.2** [12]. In this dissertation we leverage *P. aeruginosa's* LasI/LasRLasI system for engineering *E. coli* co-cultures. LasI produces 3OC12-HSL, for which in this work we will refer to as autoinducer-1 (AI-1), which binds LasR and enables a positive feedback loop of increased LasI production in addition to upregulating

RhlI and RhlR expression [5, 13]. Although it is noted that when analyzed independently, AI-1 does not activate RhlR and C4-HSL does not activate LasR in *E. coli*, demonstrating that the system components are not interchangeable [3]. These QS systems in *P. aeruginosa* control more than 300 genes in the genome, including factors related to pathogenicity such as adhesion, exotoxin and virulence factor production, and biofilm formation.

One product of interest endogenously produced by *P. aeruginosa* includes the phenazine pyocyanin (PYO), which is an aromatic compound that causes increases in intracellular superoxide [13]. In addition to PYO, a host of intermediate phenazine compounds are produced as part of the PYO pathway which host a range of biological activities due to their ability to redox cycle with various reducing agents and oxygen, these compounds are depicted in **Figure 2.3** [14-16]. In this dissertation, we employed both quorum sensing modules (LasI/LasR) and phenazine products as tools for implementing orthogonal signaling within synthetically assembled consortia.

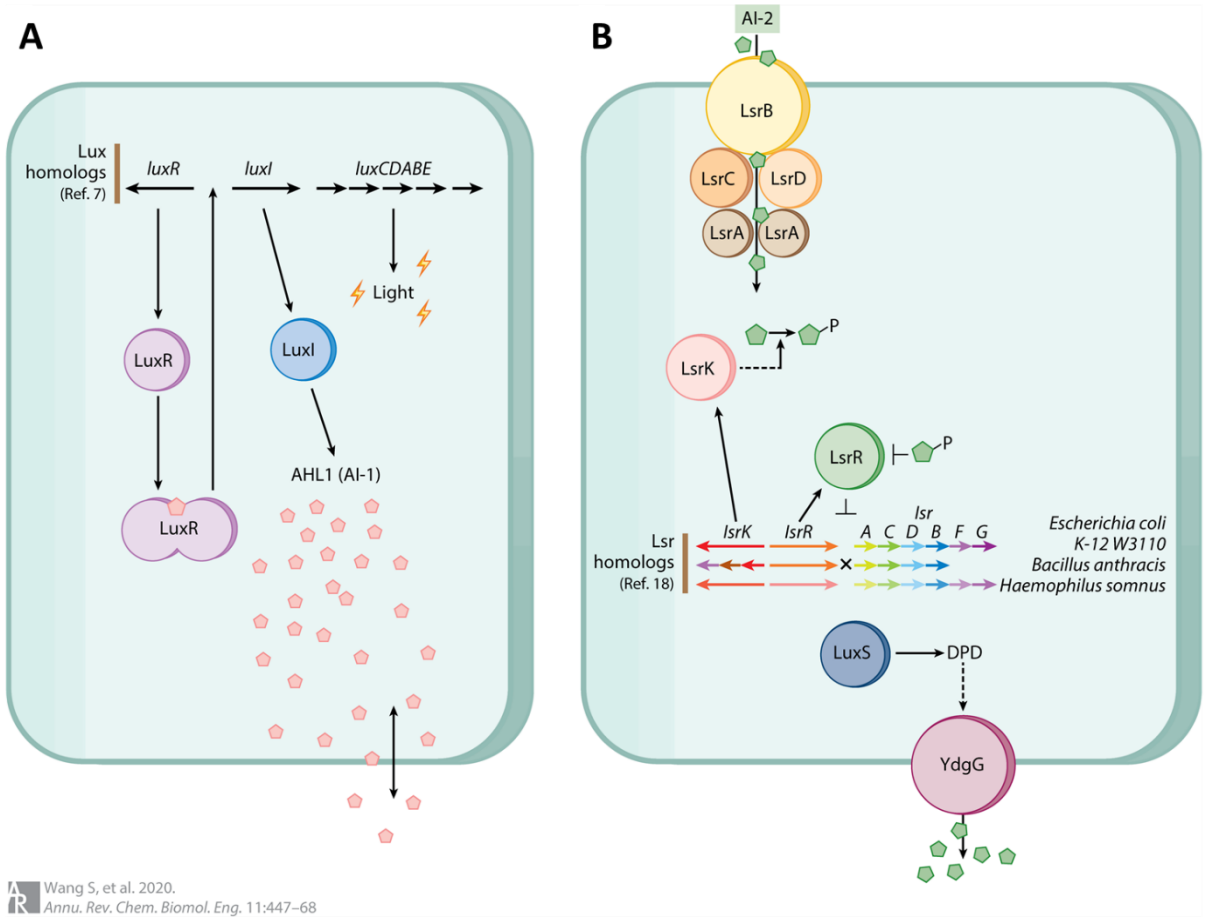


Figure 2.1 Canonical quorum sensing systems.

Canonical quorum sensing (QS) systems. (a) QS in Vibrio fischeri. Pink pentagons denote the acyl-homoserine lactone (AHL) autoinducers (AIs) that LuxI produces (3OC6-homoserine lactone). Transcriptional regulator LuxR modulates expression of AHL synthase, LuxI, and the lux operon, leading to luciferase-mediated light emission. (b) Regulatory mechanisms of the LuxS/AI-2 circuit in Escherichia coli. AI-2 (green pentagons) is imported into the cell by the Lsr transporter (LsrACDB) and is then phosphorylated by LsrK. When AI-2P binds LsrR and releases LsrR from the promoter region (thus modulating Lsr gene expression), AI-2 uptake is increased. LuxS produces DPD, the precursor to AI-2. The AI is exported by YdgG (TqsA). Lsr homologs and their organizational structures are found in both Gram-positive and Gram-negative bacteria [5].

Figure and caption was adapted from Wang, S., et al. (2020). "Quorum Sensing Communication: Molecularly Connecting Cells, Their Neighbors, and Even Devices." Annu Rev Chem Biomol Eng **11**: 447-468.

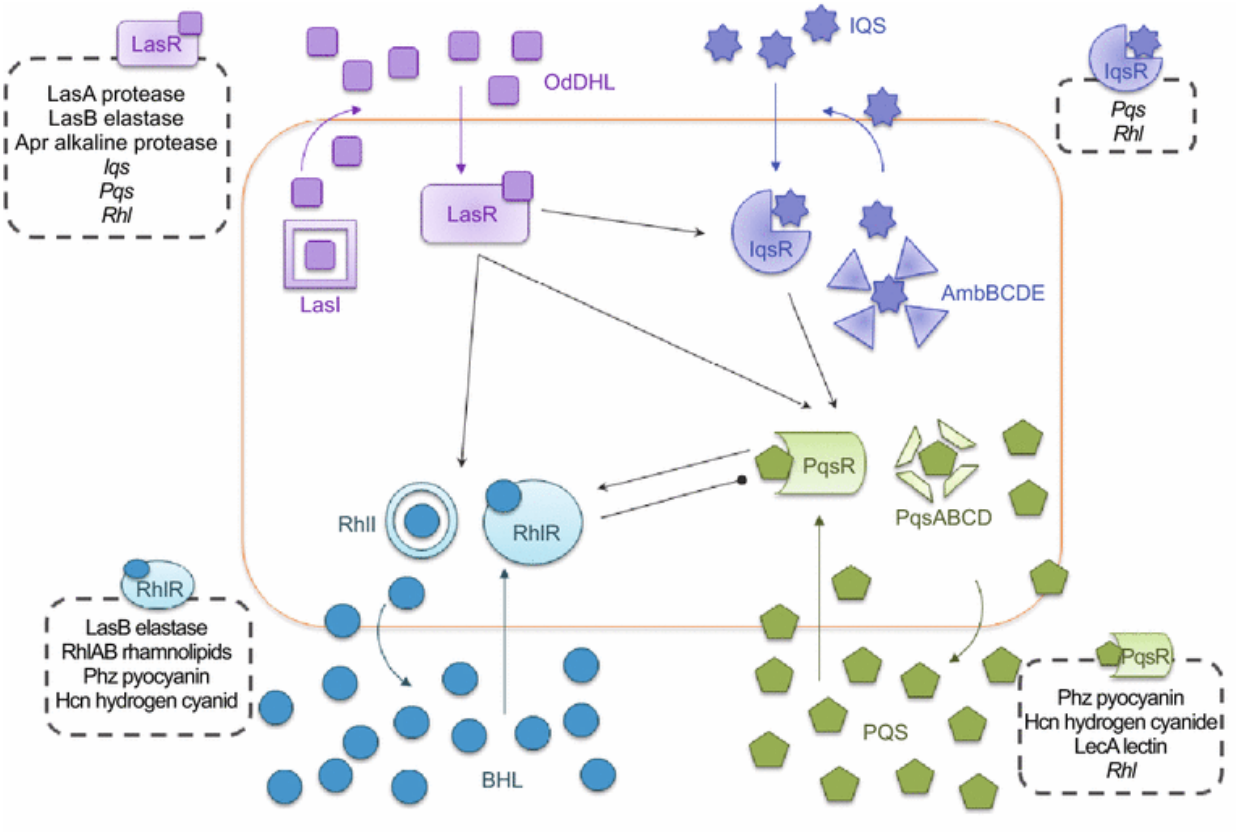


Figure 2.2 Pseudomonas aeruginosa quorum sensing.

P. aeruginosa and their respective regulons**.** Arrows indicate a stimulatory effect. Perpendicular lines indicate an inhibitory effect [12].

Figure and caption was adapted from Lee, J. and L. Zhang (2015). "The hierarchy quorum sensing network in Pseudomonas aeruginosa." Protein Cell **6**(1): 26-41.

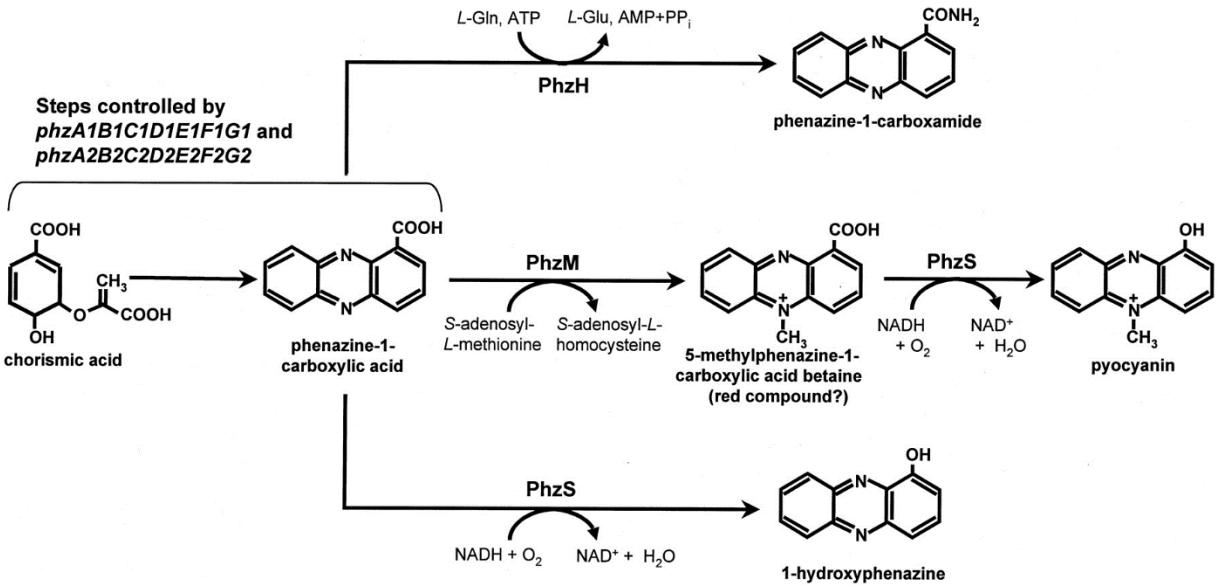


Figure 2.3 Proposed mechanism for phenazine synthesis in P. aeruginosa [14].

Figure and caption was adapted from Mavrodi, D. V., et al. (2001). "Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from Pseudomonas aeruginosa PAO1." J Bacteriol **183**(21): 6454-6465.

2.2 Electrogenetics and signaling via oxidative stress regulons OxyRS and SoxRS

In bacteria, two oxidative stress management pathways that allow for defense against oxygen damage are the OxyR and SoxRS regulons, these mechanisms are shown in **Figure 2.4**. Specifically, the OxyR protein acts as a repressor until its cysteine groups are oxidized into disulfide bonds by hydrogen peroxide [17]. This oxidation leads to a change in conformation, which allows the protein to stabilize the transcription complex [17, 18]. In *E. coli*, OxyR strongly induces NADH peroxidase (AhpCF) and a catalase (KatG) which act to reduce the hydrogen peroxide level thus lowering DNA damage [17, 19].

SoxRS responds to super oxide, in which SoxR contains an iron sulfide cluster which are sensitive to oxidization. Once SoxR is exposed to superoxide, its iron sulfide clusters are oxidized leading to a change in conformation. The oxidized SoxR induces the expression of a second

transcription factor, SoxS, which activates a host of genes which help to reduce superoxide damage [17, 18]. Compared to OxyR, SoxRS is not inducible by hydrogen peroxide but can be induced by nitric oxide and other redox cycling compounds, such as the previously mentioned PYO [8].

These regulons have previously been employed by our lab to enable electrogenetic control in *E. coli*. This has been done both directly through the production of hydrogen peroxide at an electrode surface as well as indirectly through the redox cycling of mediators to induce superoxide stress on cells in culture [7-9, 20]. In this dissertation, we leveraged these oxidative stress regulons to enable redox communication across subpopulations in co-culture and develop a model that describes electrogenetic induction.

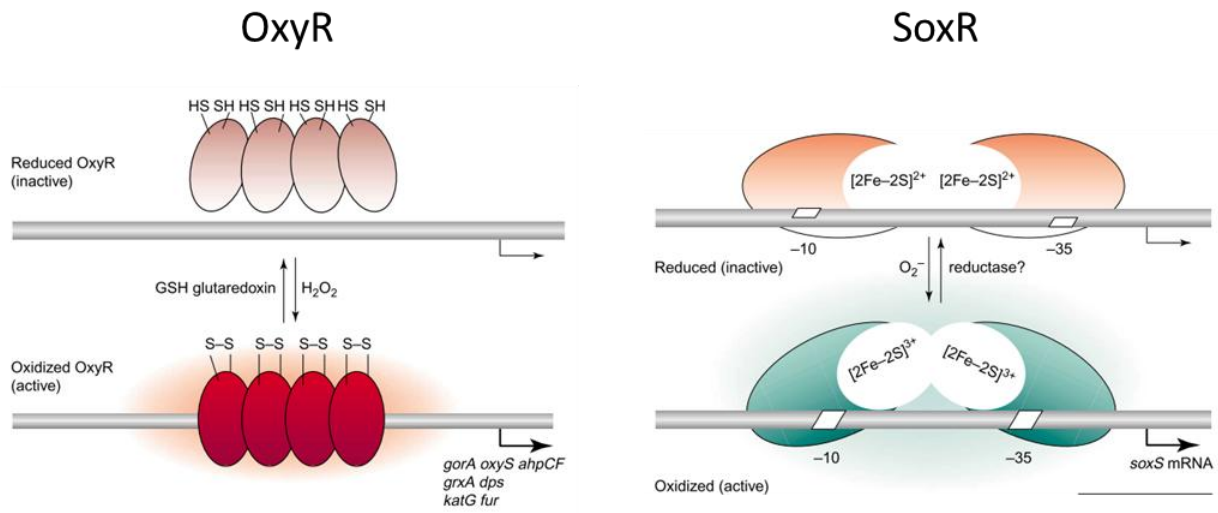


Figure 2.4 OxyR and SoxR activation mechanisms [17].

Figure and caption was adapted from Demple, P. J. P. a. B. (2001). "Redox-operated genetic switches: the SoxR and OxyR transcription factors." Trends in Biotechnology.

2.3 Electrochemical techniques for biosensing

Electrochemical methods use imposed voltages and currents to infer chemical information mediated by the flow of electrons within a sample. The development of rapid, sensitive, and low-cost biosensors is in demand for applications in point of care diagnostics as well as manufacturing quality control. Implementation of electrochemical methods for diagnostics in these biomanufacturing and clinical settings are favorable for the needs listed prior in addition to their ability to provide a quantitative readout. Electrochemical sensors fall under indirect or direct detection methods. Indirect methods rely on electroactive metabolites associated with the desired target or enzymatic activity with the desired substrate to yield electroactive products detectable by their redox cycling at the surface of the electrode [21-23]. Direct methods require a biorecognition probe on the electrode surface which specifically binds to the target, allowing for detection in changes in impedance or capacitance [21, 22].

Mediated electrochemical probing (MEP) is an indirect technique that uses redox-active chemical species as electron shuttles that can exchange electrochemical information from a biologic sample in the local environment to an electrode surface [24]. Mediators carry information by first initializing to their reduced or oxidized states by electrical input, then diffusing out and interacting with redox-active biological elements via electron exchange (**Figure 2.5**). As the mediators diffuse back to the electrode surface their redox state encodes information about the biological sample they encountered. The electrical signals measured during this process can then be analyzed to discern various samples based on redox activity. Here, we utilized MEP to gather data on molecular products and cell cultures to further classify and quantify using machine learning techniques.

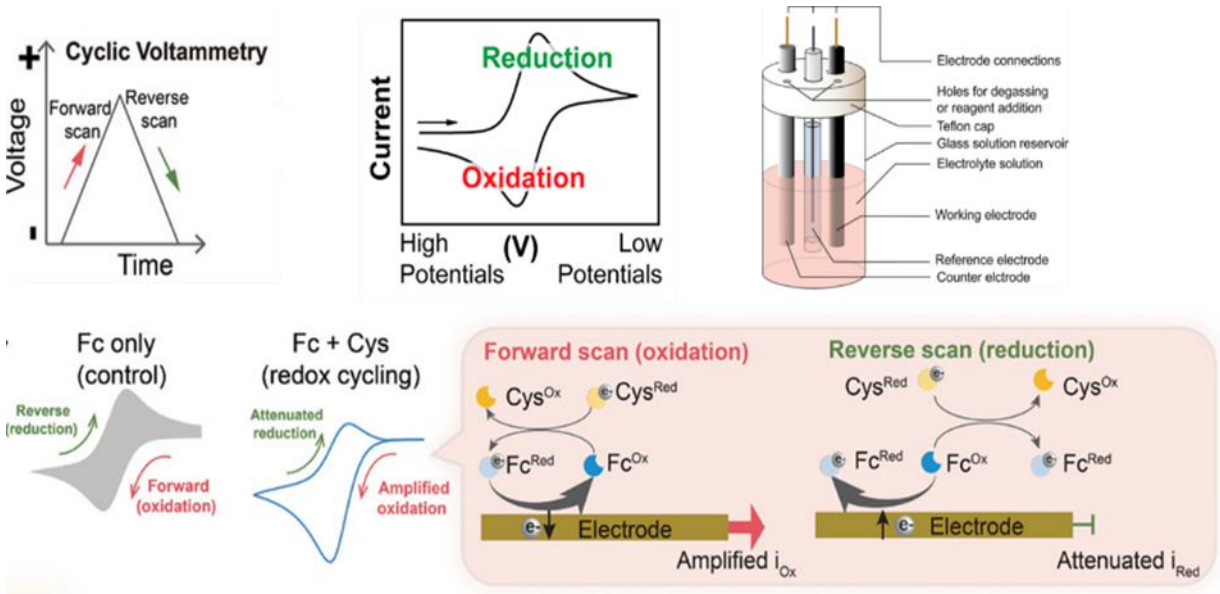


Figure 2.5 Cyclic voltammetry and mediated electrochemical probing.

Top: Voltage input and current output expected in cyclic voltammetry alongside working electrochemical setup [25]. Bottom: Cyclic voltammogram with mediator Ferrocene (Fc) alone or with cysteine alongside schematic of Fc mediated redox cycling at the electrode surface [26].

Figures and captions were adapted from Elgrishi, N., et al. (2018). "A Practical Beginner's Guide to Cyclic Voltammetry." Journal of Chemical Education **95**(2) and
Li, J., et al. (2020). "Mediated Electrochemistry to Mimic Biology's Oxidative Assembly of Functional Matrices." Advanced Functional Materials **30**(30), respectively.

2.4 Machine Learning for industrial and clinical diagnostics

Machine learning is a discipline which focuses on algorithms capable of adapting their structure based on a set of observed data by optimizing over an objective or cost function [27]. These algorithms fall under a wide array of technical approaches, suited to a host of biological contexts from aiding in tumor identification to predicting titer yield in biomanufacturing settings [28, 29]. When applying machine learning to a task, deciding on the types of input data and their preprocessing as well as which model to use is an important consideration. Previous work has shown that MEP alongside optical metrics can be used in multivariant analysis to discern between

serum samples derived from healthy and schizophrenic patients [6]. In addition to previously established electrochemical signal metrics, existing dimensionality reduction techniques were applied to electrochemical measurements to extract features. In this dissertation, we tested various dimensionality reduction techniques alongside electrochemical signal metrics as input features for training existing models to quantify antibody quality and classify disease states.

2.5 Vaginal microbiome in female reproductive health

The vaginal microbiome has been shown to be a key player in female reproductive health. Beneficial vaginal microbiota exist mutualistically with the host, providing protection from pathogenic species in exchange for nutrients and shelter. These healthy vaginal microbiome states are commonly dominated with *Lactobacillus* species, which contribute maintaining a healthy pH environment (~pH 3.5-4.5) with the production of lactic acid [30]. Bacterial vaginosis (BV) is a disease that plagues the female lower genital tract; it is characterized by dysbiosis, a shift in the vaginal microbiota from beneficial *Lactobacillus* commensal bacteria to a pathogenic polymicrobial microbiota that is associated with an increase in vaginal pH to above the optimal range of pH 3.5-4.5 [31, 32]. BV is linked to increased risk of gynecological complications such as preterm labor, spontaneous abortion, cervical cancer, pelvic inflammatory disease, and acquisition and transmission of sexually transmitted agents, including HIV [31, 33, 34]. In addition, it is also notable that *Lactobacillus iners* dominated microbiomes are implicated in low stability and transition into dysbiosis compared to other dominating *Lactobacillus* species (*L. crispatus, L. gasseri, L. jensenii*) [35]. Current clinical diagnostic methods do not screen for *L. iners*, such that preventative interventions are not possible.

BV is typically diagnosed using clinical assessment or laboratory-based testing. In clinical assessment, a widely accepted rubric referred to as Amsel's criteria is used to diagnosis. Four of the following Amsel's criteria must be met to clinically diagnose BV: a vaginal pH higher than 4.5, a presence of clue cells in vaginal fluid (via saline microscopy), a milky homogeneous vaginal discharge, and the release of a fishy odor after addition of 10% potassium hydroxide to the vaginal fluid as an indicator of amines [36]. Comparatively, the gold standard laboratory-based testing consists of conducting a gram stain on vaginal swab samples, then utilizing a standardized scoring system, called the Nugent score, to examine quantities of Gram-positive rods and Gram-negative or Gram-variable present [36, 37]. Amsel's method is considered point-of-care but has moderate reproducibility and sensitivity compared to Nugent scoring [37]. However, Nugent scoring requires sample processing by those with the skill set to prepare and read slides, results are not ready at the time of visit and turnaround times are dependent on the capacity at clinical laboratories. Due to these constraints, Nugent scoring is often used for research purposes rather than common clinical diagnosis [37, 38].

In this work, we developed an electrochemical method as a novel rapid approach to monitoring the vaginal microbiome. We utilized redox mediated electrical probing along with machine learning tools to quantify global redox measures that can be further parsed to determine the dominant species of vaginal microbes present in culture. The ability to probe the specific microbiome state would extend the capabilities for preventative treatments through the identification of transitional communities (*L. iners*). Further, development of these tools may be useful for in vivo monitoring of the microbiome state when combined with advances in "smart" intervaginal devices [39-41].

Chapter 3 Development of a redox and quorum sensing toolkit for co-cultures

This chapter is adapted with permission from the following published work:

Chun, K., Stephens, K., Wang, S., Tsao, C., Payne, G., and Bentley, W. Parsed synthesis of pyocyanin via co-culture enables context-dependent intercellular redox communication. *Microb Cell Fact* 20, 215 (2021).

3.1 Introduction

Metabolic engineers and synthetic biologists are currently investigating novel strategies to both expand the repertoire of molecular products synthesized in microbial systems and to enhance the production of established products [42]. One strategy that continues to gain traction is the use of microbial co-cultures or consortia, where the tasks required to produce a molecular product are divided amongst multiple populations, rather than being carried out by a single population [43, 44]. Implementation of this strategy allows for division of labor between populations and modularity, where host strains can be optimized for specific tasks [45]. Partitioning a product pathway among different populations also enables culture composition as a control parameter beyond traditional targeting of transcription or translation. To utilize this advantage, development of new methods that allow for robust regulation and coordination of subpopulations for functional control of the whole consortia is necessary.

Previously, we developed a genetic "growth control" module that allowed for autonomous modulation of cell growth rate and culture composition in response to a quorum sensing signal [46]. In our system, upregulation of HPr, a protein involved in sugar transport, in a *ptsH* (encoding HPr) mutant host strain resulted in increased cell growth rate. In a recent example, Dinh et al. developed

a co-culture for the production of naringenin, where the growth rate of the population responsible for the earlier part of the pathway autonomously decreased after reaching a certain cell density [47]. Similarly, others have regulated the composition of subpopulations in a microbial community by regulating production of lysis proteins or toxins [48, 49]. These strategies have only just begun to be applied to co-culture systems that are cooperatively synthesizing a molecular product. That is, while there are now many examples of co-cultures being used to synthesize products [50-55], there are few examples with modulated coordination of culture composition and activity.

In our previous work, bacterial quorum sensing signaling has provided the molecular basis for enabling intercellular communication and control. While quorum sensing methodologies can offer cell-specific targeting, particularly through acylated homoserine lactone (AI-1) signaling, this is not always the case. For example, in autoinducer-2 (AI-2) QS systems, signal perception is context dependent; both the specific signal transduction mechanisms and the concentrations that permit collective behavior, can vary based on species [56]. In general, while there is great interest in developing orthogonality in molecular signaling, signal perception and cellular responses are often context dependent. As noted, a cell's response to the same molecule, even at the same concentration, might result in completely different cell trajectories in different environments. The creation, propagation, and perception of signals for modulating co-cultures and more complex consortia is non-trivial, context dependent, and of great interest.

Here we demonstrate the signaling capabilities enabled by the redox-active phenazine, pyocyanin (PYO). We developed a robust methodology to produce pyocyanin by separating its synthesis pathway into two strains; one which produces its intermediate phenazine-1 carboxylic acid (PCA) and another which converts this intermediate into PYO (**Figure 3.1)**. We characterized this two-strain phenazine producing system's ability to produce signals, the perception of which

are dependent on the prevailing environment. We assessed this co-culture's ability to propagate signal within a complex microbial community (> two strains), where a third strain's activity is guided by each strain's relative molecular contributions. To provide further guidance within multipopulational cultures and to explore coordination between QS and redox based signaling, we overlayed QS-directed compositional control (via AI-1 signal-dependent expression of HPr) by stimulating the growth rate of Population A. When used within the phenazine signaling system, we showed it can guide outcomes of a redox responsive third strain within a tri-culture. We tested this growth control strategy in both rich and minimal media, demonstrating its ability to work with different environments to drive various outcomes.

The relative paucity of orthogonal quorum sensing systems has been cited as a limiting factor in engineering synthetic microbial consortia [57]. Importantly, our design integrates redox-responsive molecules with quorum sensing signals, contributing to diversity of signaling interfaces through which synthetic biologists can exploit for controlling cell populations. We showed that the two strain phenazine producing system could induce production of two marker proteins and a quorum sensing signal synthase in a third strain within a multi-population culture, demonstrating potential for coordinating various activity within a complex culture. We also employed user-applied quorum sensing signal (AI-1) to interface with the phenazine signaling system. Thus, the QS/phenazine system can be used to generate a variety of molecular products and even additional signal molecules, opening new lanes of molecular cell-cell communication. Our signal-actuated growth rate control module was successfully ported into various systems, showing how population control can be effectively embedded into complex signaling contexts. In sum, we demonstrate implementation of parsed PYO biosynthesis for signal propagation in multi-population systems

and develop additional methods of coordination by use of a user-directed growth control module to guide redox signaling in various environments [46].

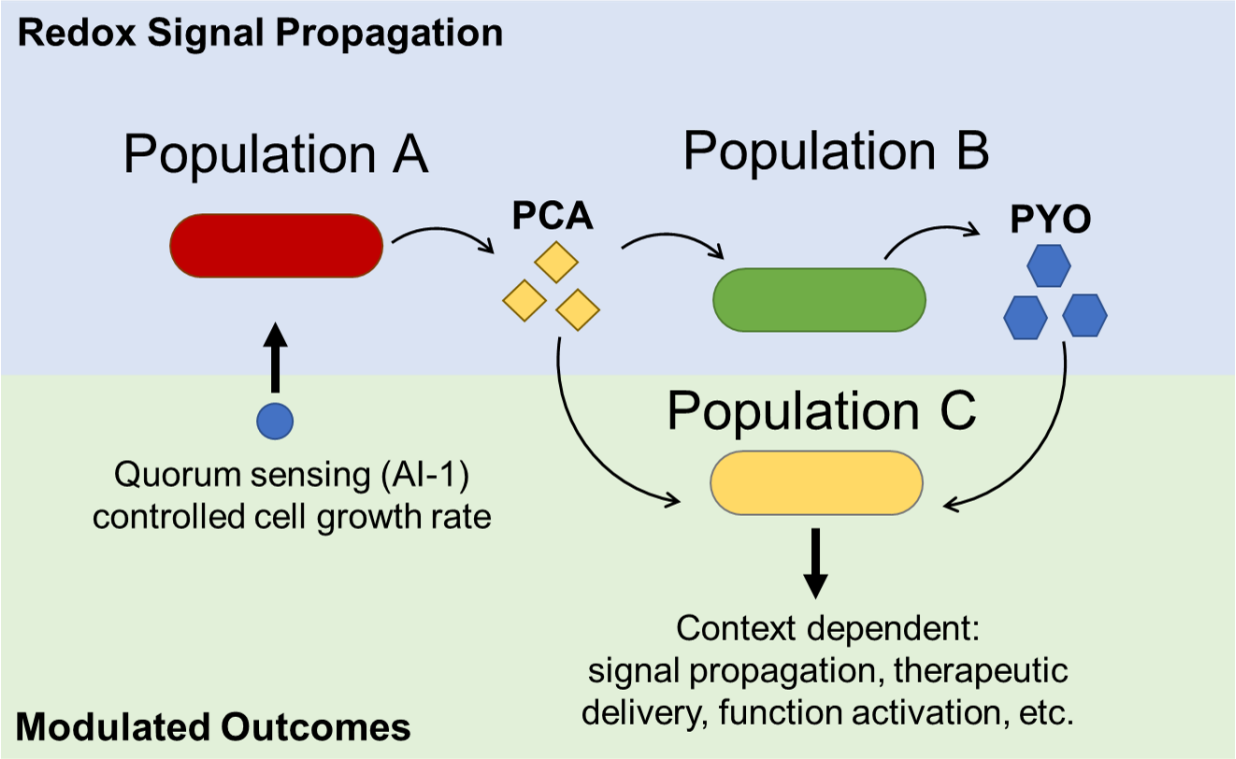


Figure 3.1 Co-culture overview.

Redox-active phenazine signaling system with quorum sensing regulated growth control module enables context dependent coordination. Parsed synthesis of PYO between Populations A and B enable context dependent activation of Population C in a triculture based on relative composition of the strains. Population A contains a quorum sensing autoinducer–activated growth control module wherein AI-1 (blue dot) activates HPr expression, allowing for tuning of relative population within the consortia for an additional control parameter for subsequent activation of Population C.

3.2 Results

3.2.1 *E. coli* strains expressing *phzA1-G1* (Population A) and *phzMS* (Population B) differentially activate *soxRS* based on environment.

To engineer a co-culture that synthesizes pathway intermediate PCA and ultimately pyocyanin (PYO), we split the overall pyocyanin synthesis pathway between two populations. *Pseudomonas aeruginosa* genes *phzA1B1C1D1E1F1G1* were expressed in Population A and genes *phzMS* were expressed in Population B. Mavrodi et al accomplished this previously in *E. coli* to assist with determining the genetic pathway for pyocyanin synthesis [14]. In that study, the authors showed that the co-culture synthesized pyocyanin (verified by HPLC), but they did not investigate the effect of culture composition on pyocyanin synthesis.

We first investigated the ability of Population B to convert PCA to PYO. Several host strains were transformed with high copy number plasmid pZE-*phzMS* containing *phzMS* under a LacO-1 promoter. 30 µM PCA was added to cultures grown in M9 media and, after overnight growth, cell-free conditioned media (CM) samples were collected (**Figure 3.2A**). We developed PYO reporter cells, *E. coli* SW101 pCT10 pET-DsRedExpress2 (SW101-DsRed), to detect the presence of PYO in experimental CM samples. These reporter cells were built on our previous work, wherein pyocyanin was used to control expression of a gene of interest under the *soxS* promoter.[8] The genetic construct in that work consisted of *soxR*, the bidirectional *soxRS* promoter, and the gene of interest. Pyocyanin can influence the oxidative state of SoxR, which in turn regulates transcription from the *soxS* promoter. Here, we built a two plasmid system that amplifies expression from the *soxS* promoter, a strategy we used previously to amplify expression from other promoters [58]. A single copy plasmid contains *soxR, the soxRS* promoter and T7 RNA polymerase downstream of the *soxS* promoter (pCT10)- a high copy plasmid contains a fluorescent reporter

under the T7 promoter (pET-DsRedExpress2). To detect the presence of PYO in our experimental CM samples, the reporter cells were grown in LB media to approximately OD_{600} 0.2 and CM sample was added. Fluorescence (indicating pyocyanin activity) was measured after approximately four hours. As indicated, all strains containing *phzMS* converted PCA to PYO (Figure 2a). Equally importantly, reporter cells in LB media did not respond to CM from strains that did not express *phzMS* but were cultured with PCA. We then tested the ability of strain PH04 pZE-phzMS to convert varying levels of PCA to PYO. We grew PH04 pZE-phzMS in M9 media with 0 to 40 µM PCA, and after overnight growth, collected CM. We added CM to the fluorescent reporter cells in LB media, to determine relative PYO activity (**Figure 3.2B**). Increasing PCA levels resulted in increasing PYO synthesis. At the cell densities and concentrations tested, there did not seem to be a factor limiting PCA conversion, even at the higher concentrations tested.

As noted above, in earlier work we discovered SoxR-responsive elements were actuated by PYO in minimal media [8, 59]. To systematically evaluate signal perception in response to media context, we carried out experiments with cells containing SoxR-regulated fluorescent protein expression in both M9 minimal media and LB media supplemented with both the final product, PYO, and the intermediate, PCA (0-1 mM, **Figure 3.2C**). Additionally, we swapped the reporter of *soxRS* activity to eGFP from DsRed (tested using *E. coli* PH04 pCT10 pET-eGFP) to preserve DsRed as a marker for indicating Population A in later experiments. As expected, the level of activation from PYO was high in M9 and was dependent on the supplemented PYO concentration in samples taken after 4 hours of incubation. Notably, the level of *soxS* activation from PYO was even higher when cells were cultivated in LB media. Similarly, the level of *soxS* activation was directly proportional to the concentration of PYO added to the LB media. However, PCA had almost no stimulatory activity when cells were grown in LB media but led to high levels of *soxS*

promoter activity in M9 media; even higher than PYO at the same levels. These results demonstrate that in M9 media, both PCA (the product of Population A) and PYO (the product of Population B) can activate eGFP expression within our sensing Population C. They also demonstrate that in LB media, only PYO (the product of Population B) activates eGFP expression in our sensing Population C. Moreover, in no case was the *soxS* response the same in any media, given the same concentration of stimulant (PYO or PCA). That is, the biological context influenced the responses of Population C to the various signaling molecules.

We then tested whether co-culturing the Population A strain with Population B resulted in PYO production, thus enabling signal propagation. Population A contained plasmid pZE-phzAG comprising *phzABCDEFG* under a LacO-1 promoter. We grew Population A (*E. coli* NEB10β pZE-phzAG) and Population B (*E. coli* NEB10β pZE-phzMS), both alone and together in M9 media. (**Figure 3.2D**). Similarly, we grew *E. coli* PH04 pZE-phzAG and *E. coli* PH04 pZE-phzMS, both alone and together in M9 media. Importantly, PH04 is a *ptsH* knockout strain, which facilitated later addition of the growth control module [60]. After overnight growth, we took cell-free CM from the cultures and tested for extracellular PYO using the fluorescent reporter cell assay described earlier (performed in LB media). Each of the co-cultures resulted in high levels of DsRed fluorescence while the monocultures did not, indicating only the co-cultures were able to produce signal in LB media via PYO production.

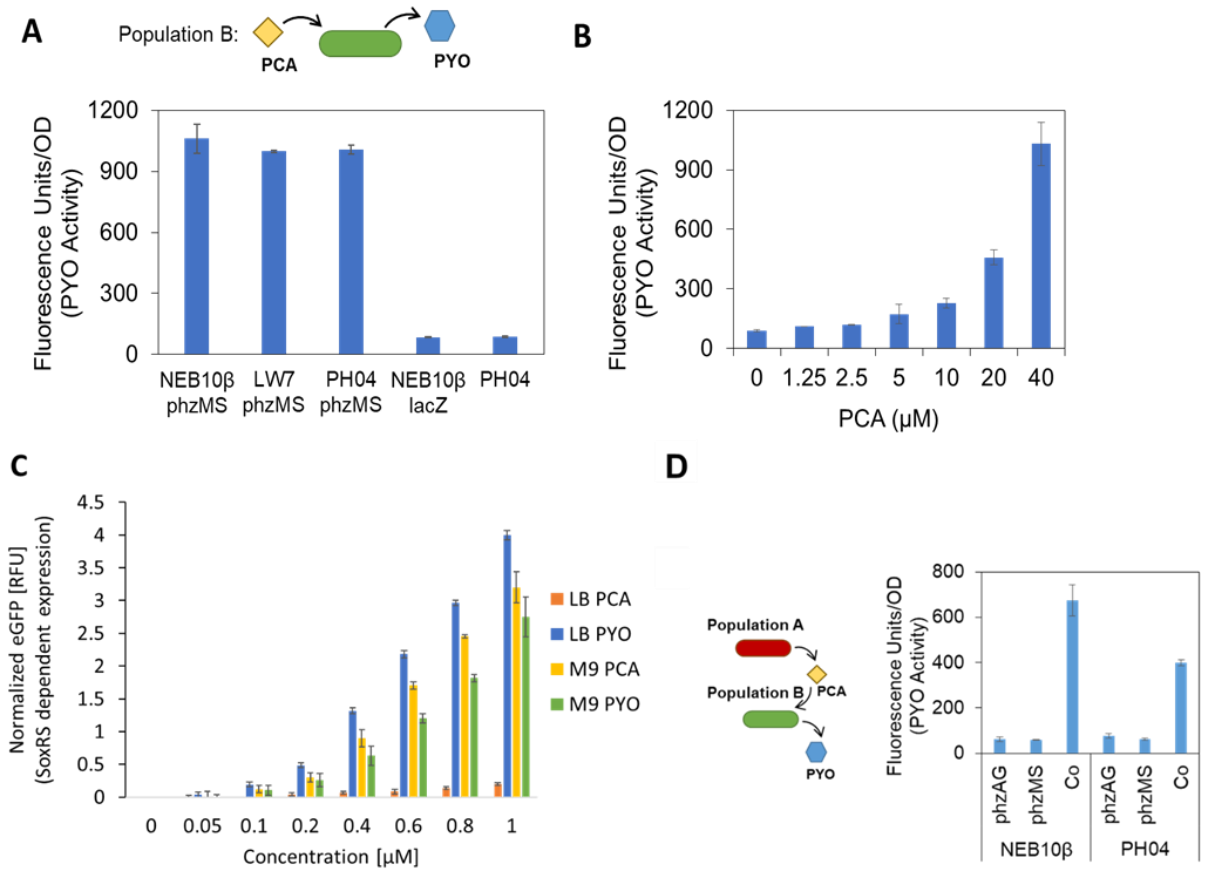


Figure 3.2 Parsed PYO synthesis gene activation.

Parsed PYO synthesis enables context dependent activation of *soxRS* induced gene expression. a) NEB10β, LW7, and PH04 host strains containing pZE-phzMS, pZE-lacZ, or no plasmid (as indicated) were grown in M9 media with 30 μM PCA. Cultures were inoculated in M9 1% (to approximately 0.05 OD_{600}) from overnight cultures and PCA was added 1.5 hours after inoculation. After overnight growth, CM was collected and tested in a PYO reporter assay using fluorescent reporter cells (SW101 pCT10 pET-DsRedExpress2). The reporter cells were grown to approximately OD_{600} 0.2 in LB media, and 20 μL CM were added to 180 μL reporter cells. Fluorescence was recorded after approximately four hours. The values reported are the relative fluorescence units divided by the cell density (OD_{600}) of the reporter cells. **b**) PH04 pZE-phzMS was inoculated 1.5% from overnight cultures. After one hour, varying amounts of PCA (ranging 0-40 μM) were added. After overnight growth, CM samples were collected. Samples were diluted 5x in LB media then measured for PYO activity using the reporter bioassay as in (a). The values reported in (a) and (b) are the RFUs divided by the reporter cells' OD_{600}. **c)** PH04 pCT10 pET-eGFP was grown for 4 hours during log-phase in either LB or M9 with PYO or PCA at the concentrations indicated (ranging 0-1 μM), then eGFP expression was measured. Fluorescence values were normalized by subtracting out background from reporter cells, then taking its fold over blank. Error bars represent s.d. of technical triplicates. **d)** NEB10β pZE-phzAG and NEB10β pZE-phzMS were cultured alone or together (9:1 ratio). Similarly, PH04 pZE-phzAG and PH04 pZE-phzMS were cultured alone or together (9:1 ratio). After overnight growth in M9 media, CM media was collected and assayed for PYO activity in LB media. All error bars represent s.d. of technical duplicates.

3.2.2 Initial co-culture composition of Populations A and B alter signal propagation in tri-culture systems to modulate behavior of a phenazine sensitive strain.

To validate the phenazine system's signal propagation in a higher order multi-population environment, we tested adding phenazine sensitive strains directly to the two-strain phenazine producing culture **(Figure 3.3A)**. We grew *E. coli* NEB10β pZE-phzAG (Population A) and *E. coli* NEB10β pZE-phzMS (Population B) in LB media. We started with different initial ratios of Population A to Population B with total OD_{600} being around 0.1. After 5 hours of growth, we added 20 μL of Population C (PH04 pSox-LasI) culture that had grown overnight in LB media. This third population produces AI-1 in response to PYO. Here, AI-1 serves as a model product which could further communicate by serving as an orthogonal signal in a high order quorum sensing consortium. Only one hour after adding Population C, we took samples and measured AI-1 activity in the extracellular media (using a previously developed AI-1 reporter bioassay) [46, 61]. The results (**Figure 3.3B**) show that Population C responded to PYO in the co-cultures by producing AI-1. As anticipated, cultures without either Population A or Population B did not result in AI-1 production by Population C in LB media since these populations cannot produce PYO on their own (**Figure 3.3B)**.

We additionally tested the system's ability to influence other types of outcomes in tri-culture by swapping *E. coli* PH04 pSox-LasI Population C for other phenazine sensitive strains. That is, we wanted to verify that the redox-active signal molecules could be used to activate genetic responses for a variety of products. In Figure 3b, we showed redox-actuated synthesis of LasI and corresponding AI-1 secretion. For additional tests we set up the same co-cultures of Populations A and B as described previously for AI-1 production. Then we regrew overnight cultures of *E. coli* PH04 pCT10-pET-DsRed and *E. coli* PH04 pCT10-pET-eGFP, which produce protein markers

DsRedExpress2 and eGFP in response to *soxS* promoter activation respectively, in fresh LB media to test as Population C. After 5 hours of culture, 20 µL of the phenazine producing co-culture was added to 180 µL of *E. coli* PH04 pCT10-pET-DsRed and *E. coli* PH04 pCT10-pET-eGFP at ~0.2 OD_{600}. The tri-culture was then incubated at 30℃ for 4 hours, then measured for its respective fluorescence **(Figure 3.3C-D)**. The results indicated the fluorescence signals were produced via similar dependencies on initial culture composition as the AI-1 producing tri-culture, where a peak signal was exhibited at the 5:1 ratio. Interestingly, these protein producing cultures required longer induction times for signal propagation to yield its final outcome. This is likely due to longer maturation times for these non-enzymatic fluorescent markers, demonstrating the notable differences Population C can have on processing the phenazine signal into downstream effects dependent on its desired activity.

These results demonstrate the phenazine system's capabilities of cell-cell signaling within complex synthetic systems (> two strains). That is, in addition to frequently used quorum sensing molecules (i.e., AHLs and AI-2), phenazines can elicit specific gene expression via the SoxRS regulon dependent on its environment that can be coopted for additional functions [8, 59]. PYO serves its native host, *P. aeruginosa*, in several ways, including as a toxin and as a signaling molecule [62]. Here we show that phenazine producing messenger cells and phenazine responding receiver cells can be developed from common laboratory *E. coli* strains for signal propagation and consortium coordination, complementing previous development of *E. coli* strains engineered to produce high levels of PYO [63]. Also, like other ported signaling transduction motifs, should there be a desire to minimize off-target effects stemming from PYO and PCA signaling, hosts can be generated as SoxRS deletion mutants or use multiplexed CRISPR-mediated downregulation [8, 59]. The data also indicates that the tri-culture outcomes varied based on the initial ratio of Population A to

Population B, demonstrating that the culture composition affects the rate of biosynthesis and subsequent communication. This allows for utilization of initial A:B ratio as a tunable design parameter for altering activity of the third population.

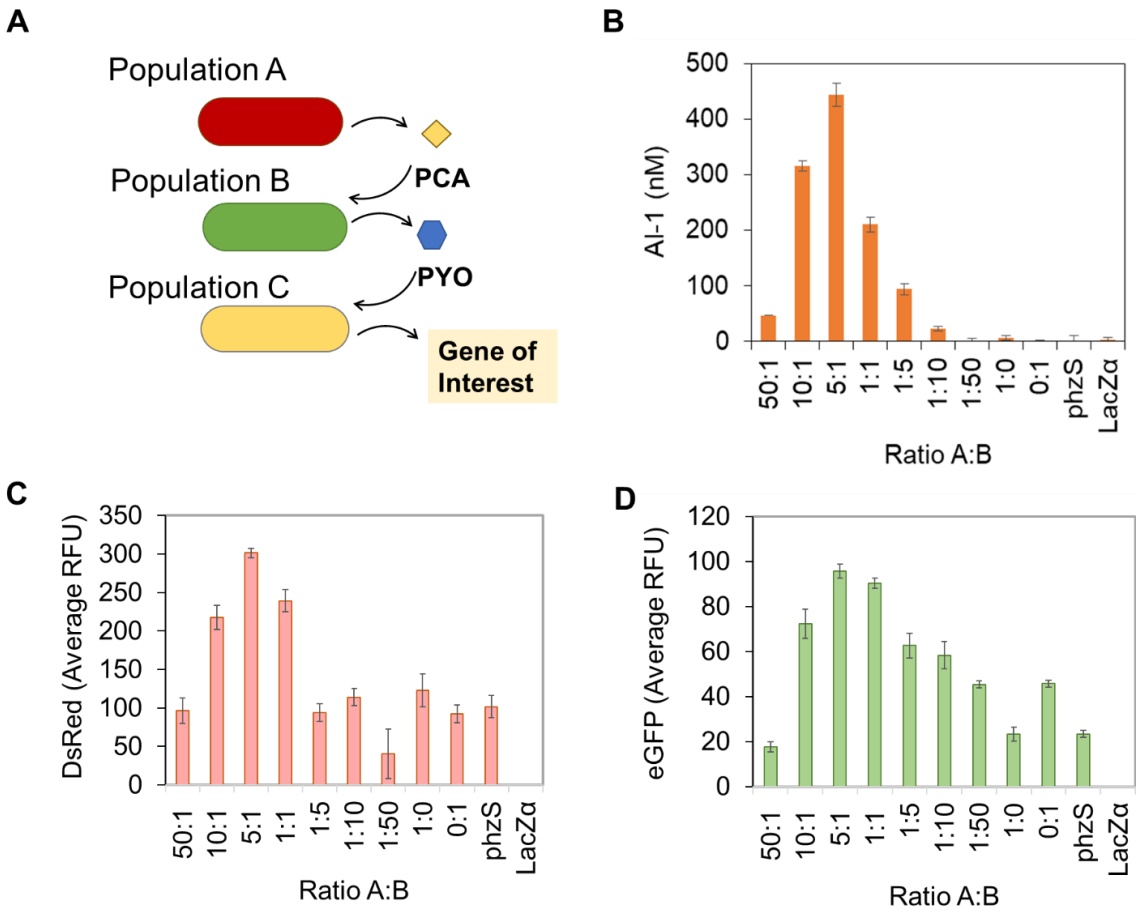


Figure 3.3 Various co-culture seed compositions within tri-cultures.

Initial co-culture composition influences activity of a phenazine sensitive strain within tri-culture system in LB. (A) Schematic of tri-culture system in LB media, where PYO drives activity in phenazine sensitive strain, Population C. (B) NEB10β pZE-phzAG (Population A) and NEB10β pZE-phzMS (Population B) were co-cultured in LB media. After 5 hours of co-culture, 20 µL of an overnight culture of PH04 pSox-LasI, pSox- (Population C) grown in LB media were added to the co-cultures. After 1 additional hour of culture, the CM samples were taken. An AI-1 reporter assay was used to measure the extracellular AI-1 levels in the culture.[46] The controls "phzS" and "LacZα" indicate growth of monocultures NEB10β pZE-phzS or NEB10β pZE-lacZα, respectively, prior to addition of Population C. Error bars indicate s.d. of technical duplicates. C) and D) NEB10β pZE-phzAG (Population A) and NEB10β pZE-phzMS (Population B) were co-cultured in LB media. After 5 hours of co-culture, 20 µL of the co-culture was added to 180 µL of PH04 pCT10-pET-DsRedExpress2 or PH04 pCT10-pET-eGFP (Population C), C) and D) respectively, which were regrown from overnight culture to ~0.2 OD_{600} in LB media. After 4 hours of additional culture relative fluorescence was measured. The controls "phzS" and "LacZα" indicate monocultures of NEB10β pZE-phzS or NEB10β pZE-lacZα, respectively, in addition to Population C. Background fluorescence was subtracted out from samples by subtracting LacZα value from each sample. Error bars indicate s.d. of technical triplicates.

3.2.3 Growth control module for user-modulated growth rate in Populations A and B.

Having demonstrated that *E. coli* co-cultures expressing *phzAG* and *phzMS* produce PYO in both M9 media and LB media, and that the culture composition affects phenazine synthesis and signal propagation, we next sought to add our previously developed growth control module to either Population A or B in order to allow for user-modulated cell growth rate of either subpopulation [46]. The growth control module contains *ptsH* under the *lasI* promoter and *lasR* under a constitutive T5 promoter. As a population marker, DsRedExpress2 is also constitutively expressed. AI-1 addition upregulates HPr (encoded by *ptsH*), which increases cell growth rate in a *ptsH* mutant host strain.[46, 60] Here, we added these components to the pZE-phzAG and pZE-phzMS plasmids to create plasmids pZE-phzAG-ptsH and pZE-phzMS-ptsH. We transformed the *ptsH* mutant strain *E. coli* PH04 (see Figures 2 & 3) with these plasmids. We then grew *E. coli* PH04 pZE-phzAG-ptsH and *E. coli* PH04 pZE-phzMS-ptsH (separately) in M9 media and tested whether AI-1 addition changed the culture growth rate (**Figure 3.4**). As anticipated, increased AI-1 levels resulted in increasing cell growth rate. Notably, the calculated growth rates between 2 and 5 hours after AI-1 addition increased from 0.25 hr^{-1} and 0.28 hr^{-1} for *E. coli* PH04 pZE-phzAG-ptsH and PH04 pZE-phzMS-ptsH, respectively, with no AI-1 to 0.51 hr^{-1} and 0.57 hr^{-1} for *E. coli* PH04 pZE-phzAG-ptsH and PH04 pZE-phzMS-ptsH, respectively, with 1000 nM AI-1 effectively doubling the growth rate of these cells. The difference in growth rate between cultures grown with or without AI-1 is similar to what was previously observed[46] (in cells without the *phz* operon genes), and therefore the expression of the *phz* operon genes did not appear to interfere with the function of the growth control module in either strain tested. These results further demonstrate the portability of the previously described HPr growth controller to cells that have different functions, each synthesizing a part of a metabolic biosynthesis pathway towards pyocyanin.

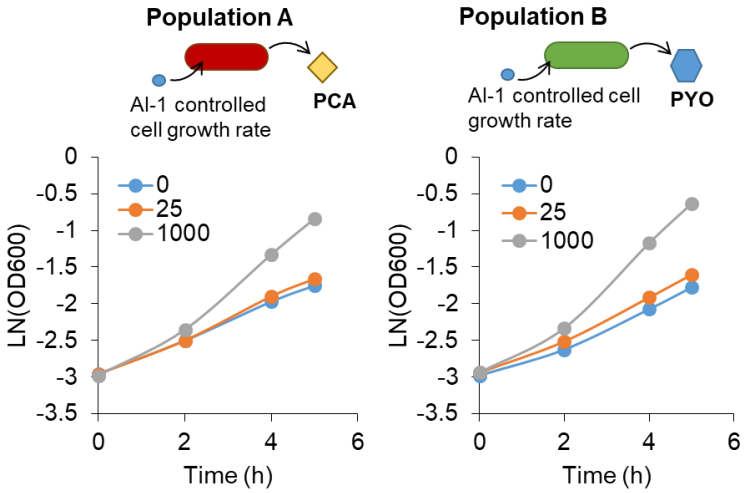


Figure 3.4 Growth control module characterization.

Growth control module enables regulation of cell growth rate in Populations A and B. PH04 pZE-phzAG-ptsH (left panel) and PH04 pZE-phzMS-ptsH (right panel) were grown in M9 media with glucose. Either 0, 25, or 1000 nM AI-1 (as indicated) were added at $t = 0$. Cell density was recorded over time, the plots above represent one replicate.

3.2.4 Tuning growth rate in Population A modulates outcomes of the phenazine signaling system.

We then tested co-cultures of Populations A and B, wherein we changed Population A so that it contained the HPr growth control module. We grew Population A (*E. coli* PH04 pZE-phzAG-ptsH) and Population B (*E. coli* PH04 pZE-phzMS) in M9 media using different initial ratios of A to B (20:1 and 100:1). As before, we ran these experiments with a total starting OD_{600} of ~0.1, and then added different concentrations of AI-1 one hour after inoculation. After overnight growth, we collected conditioned media to test for extracellular pyocyanin using the fluorescent reporter assay in LB media such that any residual PCA is not recorded. **Figure 3.5** shows the fluorescence, indicative of PYO, as measured by the reporter cells (*E. coli* SW101-DsRed). Increased AI-1 levels increased the level of PYO produced in the overnight co-cultures. The fold change in PYO produced was nearly 5-fold between the cultures without AI-1 and with 1 µM AI-1. Interestingly, this fold change was found irrespective of the initial culture ratio of A to B, suggesting that the

resultant PYO synthesis was influenced most heavily by the AI-1 addition (which modulates Population A growth rate). Here, AI-1 serves as additional knob to alter phenazine signaling via PYO production. Without the growth control module (**Figure 3.2D**) both populations are required to produce PYO, whereas here both populations plus AI-1 seem to be required for significant signal production. We suspect that the growth control module, which includes constitutive expression of DsRedExpress2 and LasR, may divert some resources away from PCA synthesis, and this effect of constitutive expression of DsRedExpress2 and LasR on PCA synthesis may be especially noticeable when *ptsH* is not induced (when there is no AI-1). This could explain why the co-cultures with the inducible growth control module in Population A produce little PYO when there is no AI-1, while the co-cultures without the growth control module (which do not have constitutive expression of DsRed) do produce PYO. The basal level of *ptsH* and/or the metabolic activity of Population A might be important. Consistent with this notion, by comparing Figure 5 with Figure 2d, when AI-1 is used to induce *ptsH* expression (i.e., increase Population A metabolic activity), we found an increase in fluorescence indicating increased PYO compared to co-culture without the growth control module.

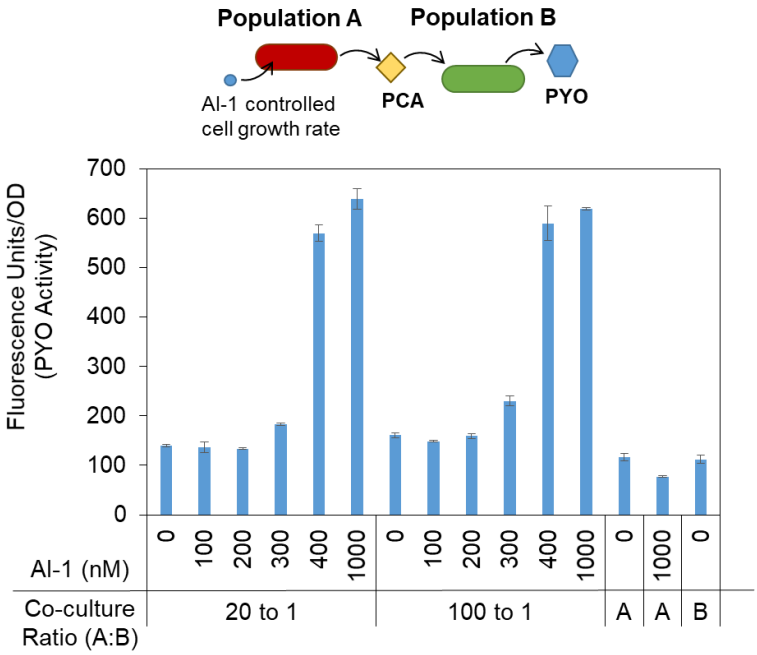


Figure 3.5 Signal output from growth control module co-cultures.

Signal output from growth control module toggles phenazine signaling co-cultures in LB media.

PH04 pZE-phzAG-ptsH and PH04 pZE-phzMS were co-cultured together in M9 media with the initial inoculation ratios, as indicated. Population A contains pZE-phzAG-ptsH and exhibits AI-1 modulated growth rate. One hour after inoculation, the indicated concentration of AI-1 was added. The cultures were grown overnight. Subsequently, CM samples were collected and PYO activity was tested using the fluorescent reporter assay in LB media (see Methods). Error bars represent s.d. of technical duplicates.

3.2.5 Growth control module enables user control of signal propagation in tri-cultures.

After demonstrating the growth control module's effect in our phenazine signaling co-culture, we tested capabilities of user control to modulate activity in a multi-population system. We grew Population A with the HPr growth controller and DsRed (*E. coli* PH04 pZE-phzAG-ptsH-DsRed) and Population B (*E. coli* PH04 pZE-phzMS) in either M9 or LB media at various initial composition ratios and AI-1 concentrations. Then, after overnight growth, we regrew the co-cultures in fresh media and supplemented with Population C, the same phenazine sensitive strain as used in Figure 3d (*E. coli* PH04 pCT10-pET-eGFP) that produces eGFP, to easily monitor activity via fluorescence. In all cases, we added Population C at roughly the same OD as the total OD of Populations A and B. We tested these tri-cultures in both M9 and LB mediums to assess the effects of environment on the growth control module's capabilities. In these experiments, we expected to see eGFP expression in Population C to increase with initial AI-1 concentration and initial Population A composition. In both cases, we hypothesized increased Population A would lead to increased PCA and PYO, leading to stronger signal propagation and higher eGFP expression. We also expected to see little or no fluorescence in Population C from cultures without Population A (i.e., Population A is required to produce intermediate, PCA, and downstream product, PYO).

After four hours of tri-culture growth, we measured DsRed and eGFP fluorescence of Population C to assess the co-culture's effect on the total system. **Figure 3.6A-B** shows average raw DsRed fluorescence, indicative of relative amounts of Population A in the multi-population culture. **Figure 3.6C-D** shows average raw eGFP fluorescence that is attributed to Population C activity. In both M9 and LB media, the initial ratio appeared to influence the composition and eGFP measured in cultures without AI-1, particularly when there were more Population A cells to

begin with. These results are consistent with **Figure 3.3,** as we found there that initial composition affects signal propagation in tri-culture. In **Figure 3.6A-D**, we also demonstrated that stimulation of Population C (eGFP expression) in a tri-strain culture is correlated to the relative quantity of Population A in culture as indicated by total DsRed fluorescence. Thus, we found here that one can alter the consortium by increasing AI-1 concentration to turn on the growth control module, which increases both the relative amounts of Population A and eGFP expression in Population C, regardless of the initial ratio of Population A to B.

It was interesting to note that in both M9 and LB media that as the ratio of A:B was increased, the relative fluorescence of population C increased, reaching an apparent maximum at the highest AI-1 levels when the fraction was between 19:1 and 99:1, suggesting that the influence of the altered growth rate was at a maximum in this region **(Figure 3.6C-D)**. In both environments, at higher initial ratios without AI-1 there appeared to be a saturation in Population C activity from the additional number of Population A cells.

We also noted that in M9 media, at initial ratios above 19:1, growth rate modulation seemed to have less of an impact in determining composition. Interestingly, while 400 nM AI-1 addition dramatically increased the relative amounts of Population A (DsRed fluorescence), further AI-1 (i.e., to 1000 nM) did not. That said, the additional AI-1 did appear to increase Population C's eGFP production. We plotted the ratio of eGFP to DsRed to qualitatively assess Population C activity per Population A for high initial ratios and with AI-1 addition (**Figure 3.6E)**; our data suggest that Population C activity was a bit higher at 1000 nM AI-1 than 400 nM. Perhaps in M9 media where both PCA and PYO are generated and also influence Population C, the added growth control module activity contributed to signal propagation as well as to the composition of Population A. The 1:0 culture, where additional AI-1 resulted in additional Population C activity

(**Figure 3.6E**) reinforces this in that this culture only had Populations A and C and no PYO from B. Thus, the added Population C activity was likely responding to additional PCA made per cell of Population A for producing a signal.

Comparatively, in LB media at ratios above 19:1 and with AI-1 addition, the effects of increased Population C activity actually diminished with increased Population A (**Figure 3.6F**). Here, incremental increases in PCA produced per cell of Population A would not be reflected in Population C activity, which is consistent with the ratios depicted in **Figure 3.6F**. Instead, at high initial ratios and with additional AI-1 where Population A increases without increase in Population C activity, suggests that Population B might be limiting in that there is less conversion of PCA to PYO. Further studies are currently underway to investigate the dynamic nature of the specific signal molecule levels, their uptake and subsequent temporal effects on gene expression.

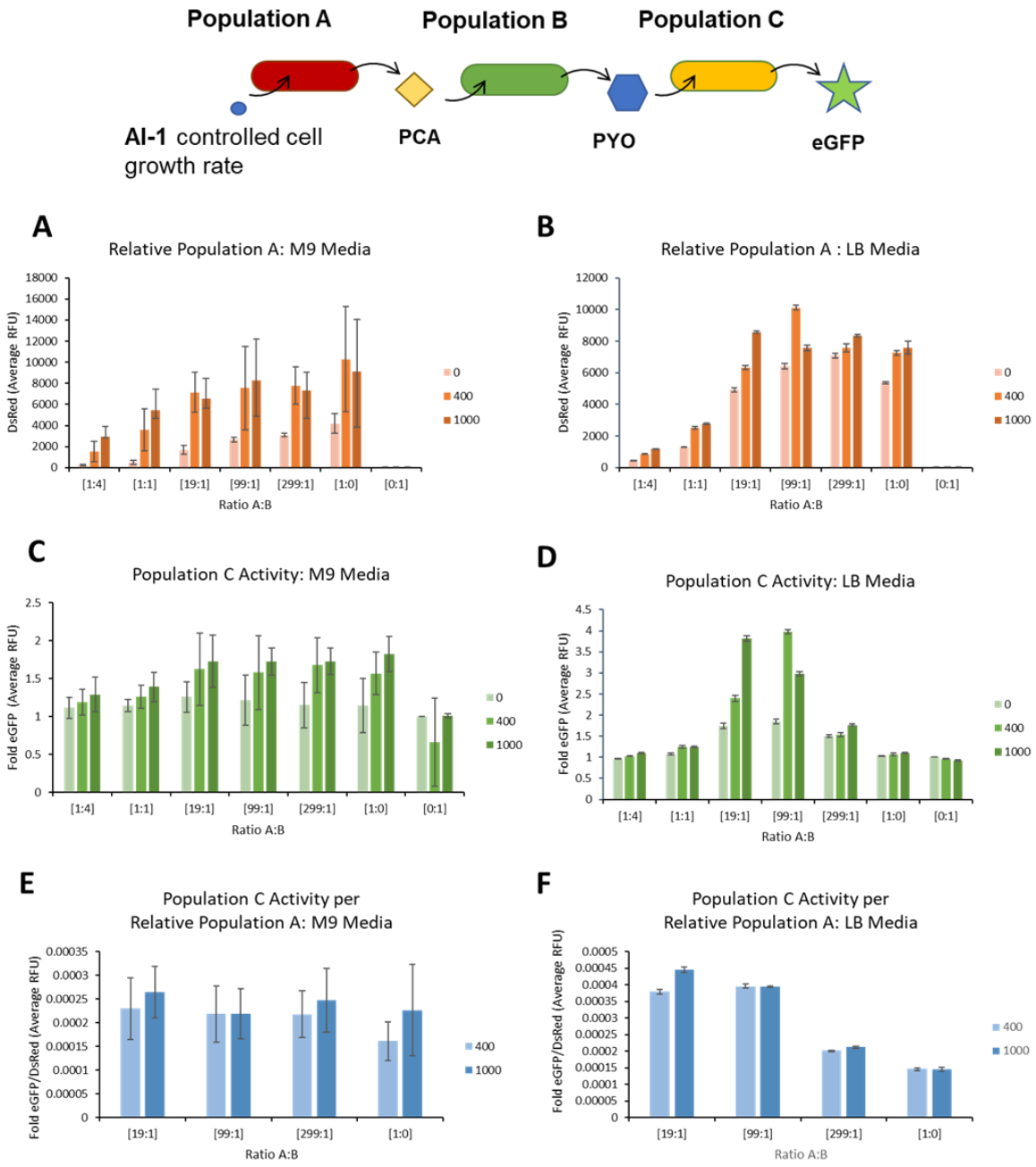


Figure 3.6 Growth modulation in tri-cultures.

Population A growth modulates Population C activity dependent on initial co-culture composition, AI-1 addition, and environment. Population A (*E. coli* PH04 pZE-phzAG-ptsH-DsRed) and Population B (*E. coli* PH04 pZE-phzMS) were co-cultured together at initial inoculation ratios as indicated on the x-axis and with nanomolar AI-1 concentrations as indicated in the legend. After overnight growth, 20 µL of the PYO producing co-culture (~1.5 OD_{600}) was added to a culture of ~0.2 OD_{600} Population C (PH04 pCT10-pET-eGFP) with 200 µL final volume in fresh media. The tri-strain culture was then incubated for 4 hours, and then measured for DsRed and eGFP fluorescence. **A and B:** Raw DsRed fluorescence of the tri-strain co-culture

at 4 hours, indicating Population A composition in culture. **C and D:** Average eGFP fluorescence of tri-strain co-culture, indicating Population C activity. **E and F:** Average eGFP divided by average DsRed, indicating relative Population C activity per Population A. Error bars indicate s.d. of triplicates.

3.3 Discussion

In this work, we have demonstrated the potential for combining quorum sensing and redox-based signaling in synthetic consortia. We employed a user-actuated genetic switch to control signaling mediated by pyocyanin synthesis in both two-member and three-member systems. Initially, using a two-member consortium, we employed culture composition as a parameter for controlling metabolic flux and signaling, as opposed to directly regulating transcription in either strain. In a three-member consortium, we used our previously reported growth control module to apply user-controlled quorum sensing communications for guiding cooperative signaling by phenazines PCA and PYO. Interestingly, we also found that the importance of the growth control module was most impactful at initial composition ratios below 99:1 and further found that outcomes were impacted less when initial ratios were high (~99:1 to ~299:1). That is, the outcome was already strongly determined by the predominance of Population A and further increases were less impactful. We suggest, however, that a strategy of altering the composition and metabolic activity may be broadly applicable in other systems as pyocyanin is derived from the shikimate pathway, a common starting point for many molecular products [64].

Strategies for modulating culture composition and activity may be of even more general utility as metabolic engineers and synthetic biologists consider ways to modulate microbiomes. Importantly for these contextual applications, we found that media composition influenced the *soxR*-driven gene expression, based on whether PCA or PYO was the predominant signaling molecule. For instance, the effect of the relative levels of Population A to B on the behavior of Population C varied based on media. Due to PCA and PYO's variable influence in *soxRS* gene

expression in various hosts (that was shown to be dependent on the growth media), we suggest that context-dependent signaling can be an advantage or a disadvantage but is certainly a factor that needs to be considered for future applications of microbial consortia. Previous studies have also shown consortia outcomes to be dependent on nutrient availability and environmental factors [65, 66]. That is, while it is attractive to consider unique orthogonality among signaling molecules, cultivation conditions and deployment context of consortia must be considered. Temperature, carbon sources, and the presence of other metabolites that may interfere with signal propagation and a consortiums functional success are important for robust design and optimization during bottom-up assembly.

The novel communication strategies demonstrated by this work also implies that user directed PYO producing co-cultures might have advantages for coordinating activity within cultures of higher order populations. We showed conversion between quorum sensing and redox communications by interplaying AI-1 and the PYO synthesizing coculture to influence one another's signaling. Thus, users might introduce and control a PYO producing co-culture to modulate orthogonal signaling molecules such as quorum sensing AHLs that, in turn, induce activity in downstream strains within a higher order multi-population system (i.e., > three strains). Additionally, the growth control module shown here might enable a finer level of control where initial consortia composition is limited, in that it may influence metabolic flux as well as population composition.

3.4 Conclusions

We showed that the pathway for synthesis of pyocyanin could be split between two populations and guided by initial composition of the consortia and subsequent growth modulation of the PCA producing strain. We further showed that PYO and even PCA could be used in *E. coli* as a molecule for cell-cell communication. By splitting the pathway for PYO synthesis between two populations, the behavior of the phenazine sensitive receiver is dependent on the relative composition of the PYO producing strains. We then demonstrated capabilities for user-directed control of this signal propagation in multi-population cultures by modulating the growth rate and therefore co-culture composition and metabolic activity.

The tools we have developed here support user-operated consortia control, however multi-population systems especially those built from the bottom up, are sensitive to the prevailing balance of population, molecular signaling, metabolic activity, etc., in which environment-specific context must be considered. Irrespective of the particular environment, we were able to use PCA and PYO as signaling molecules to drive downstream activity. We also effectively employed a user-directed quorum sensing induced growth control module to influence intercellular redox communication that propagated signal to guide a third strain's activity. Moving forward, these tools could be used to modulate a keystone organism of particular importance within a consortium or coordinate higher order activity. Such strategies that facilitate engineered cooperativity in synthetic biology systems may find utility in future applications including those where co-cultures and mini consortia are used for small molecule synthesis.

3.5 Methods

3.5.1 Strains and plasmids

The strains, plasmids, and primers used in this study are listed in **Appendix Tables 8.1** and **8.2**. Plasmids pZE-phzAG, pZE-phzMS, and pZE-lacZα are derived from the commercial vector pZE12MCS (ExpresSys), a high copy colE1 origin plasmid. Each plasmid contains the genes of interest under the LacO-1 promoter in the vector. The *phzA1-G1*, *phzM*, and *phzS* genes were amplified from *P. aeruginosa* PA01 using primers KpnI-phzA1-FWD and HindIII-phzG1-RS, HindIII-phzM-FWD and HindIII-phzM-RVS, and KpnI-HindIII-RBS-phzS-FWD and BamHI-phzS-RVS, respectively. The *phzA1-G1* fragment and the pZE12MCS vector were digested with KpnI and HindIII, and ligated to form pZE-phzAG. Plasmid pZE-phzMS was cloned in a multi-step process. When amplifying *phzS* a ribosomal binding site was placed upstream of *phzS* along with KpnI and HindIII restriction digestion sites. A BamHI restriction digestion site was added downstream of *phzS*. This fragment and the pZE12MCS vector were digested with KpnI and BamHI restriction enzymes, and ligated to form pZE-phzS. Then, the *phzM* fragment and pZE-phzS were digested with the HindIII restriction enzyme, and ligated to form plasmid pZE-phzMS.

To construct plasmid pZE-lacZα, *lacZα* was amplified from a wild type *E. coli* strain (W3110 derivative), using primers KpnI-lacZalpha-FWD and BamHI-lacZalpha-RVS. The fragment and pZE12MCS vector were digested with KpnI and BamHI restriction enzymes, and ligated.

Plasmids pZE-phzAG-ptsH and pZE-phzMS-ptsH were generated by adding the growth control module to the pZE-phzAG and pZE-phzMS plasmids. The growth control module, consisting of *ptsH* under the *lasI* promoter and dsRedExpress2 and *lasR* under a constitutive T5

promoter, was amplified from pAHL-HPr [46]. Primers PciI-t7-term-rvs and PciI-ptsH-rvs were used to amplify the control module and add a PciI restriction digestion site on either end of the fragment. Restriction digestion and ligation were used to insert the fragment into the pZE12MCS vector at the PciI restriction digestion site.

Plasmid pCT10 is a low copy plasmid derived from pFZY1 [67]. *soxR*, along with the divergent *soxR and soxS* promoters were amplified from *E. coli* using primers Fsoxp and Rsoxp. This fragment was inserted into the Invitrogen pCR-Blunt II-TOPO vector generating plasmid pTOPO-SoxRS. Restriction digestion of plasmids pFZY1 and pTOPO-SoxRS with restriction enzymes BamHI and HindIII, and ligation were used to insert *soxR* and *sox* promoters into pFZYI. T7 RNA polymerase was then inserted downstream of the *soxS* promoter at the HindIII restriction digestion site to generate plasmid pCT10.

Strain SW101 is derived from *E. coli* ZK126 (W3110 Δ*lacU169 tna-2*). The Datsenko and Warner method [68] was used to delete *soxR*, the divergent *soxRS* promoter region, and *soxS*. Primers soxHP1 and soxHP2 were used for the deletion.

3.5.2 Cell culture conditions

LB media was used for cloning and overnight growth of cultures inoculated from glycerol stocks. M9 minimal media (1x M9 salts, 2mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% casamino acids, 0.4% glucose) or LB media were used for experiments as indicated in the figure captions. Cultures were grown at 37ºC with 250 rpm shaking. Ampicillin (50 µg/mL) and/or kanamycin (50 µg/mL) was used to maintain plasmids.

3.5.3 PYO fluorescent reporter assay

A fluorescent reporter assay was used to quantify PYO in experimental samples. CM samples were collected by filtration through a 0.2 µM filter and stored at -20ºC until analysis. The fluorescent reporter cells, SW101 pCT10 pET-DsRed, consist of a dual plasmid system. A single copy plasmid contains the pyocyanin sensitive *soxS* promoter. T7 polymerase is under the *soxS* promoter and activates DsRed on the high copy pET plasmid.

For the assay, SW101 pCT10 pET-DsRed cells were reinoculated in LB media from overnight cultures and grown to approximately 0.2 OD_{600}. 180 µL of reporter cells and 20 µL of sample were added together in 96 well, black wall, clear bottom plates. Cultures were grown at 30ºC, 250 rpm shaking for approximately 4 hours. A SpectraMax M2e plate reader was used to read DsRed fluorescence and OD_{600}. For fluorescence, the excitation wavelength was 550 nm, the emission wavelength was 579 nm, and a cutoff of 570 nm was used. The reported value was divided by the cell density (OD_{600}).

3.5.4 AI-1 luminescent reporter assay

To measure extracellular AI-1, cell-free conditioned media samples were collected. *E. coli* AI-1 reporter cells with plasmid pAL105[61] were grown overnight in LB media, and then diluted 2500 fold in fresh LB media. 10 µL of samples were added to 90 µL of reporter cells. If necessary, samples were diluted prior to the assay to be in the linear range of the assay. Samples making up a standard curve of known AI-1 concentrations 0, 12, 24, 36, 48, and 60 nM AI-1 were also added to the reporter cells. After three hours, luminescence was recorded. A linear fit was used to determine the standard curve, which was then used to calculate the AI-1 in the experimental

samples. If the experimental samples were diluted prior to the assay, the results were multiplied by the dilution factor (usually 5 or 10) to back calculate to the AI-1 in the original sample.

3.5.5 Tri-strain co-culture assay

Population A (PH04 pZE-phzAG-ptsH) and Population B (PH04 pZE-phzMS) were co-cultured together overnight in M9 media with AI-1 concentrations ranging 0-1000 nM. Overnight cultures were incubated at 37ºC at 250 rpm shaking. After overnight growth, 20 µL of co-culture was added to 180 µL of ~0.2 OD_{600} Population C (PH04 pCT10 pET-eGFP) culture in 96 well, black wall, clear bottom plates. The tri-strain culture was then incubated for 4 hours at 30ºC with shaking in a TECAN Spark plate reader. DsRed (Ex/Em: 550/579 nm), eGFP (Ex/Em: 488/507 nm), and OD_{600} were measured every 30 minutes to quantify Population A relative composition and Population C activity.

Chapter 4 Development of a graph network model for the study of engineered microbial consortia

This chapter is adapted with permission from the following published work:

Chun, K., VanArsdale, E., May, E., Payne, G., Bentley, W. Assessing Electrogenetic Activation via a Network Model of Biological Signal Propagation. *Frontiers in Systems Biology* 4 (2024).

4.1 Introduction

Synthetic biology has enabled the production and sensing of biomolecules through the design, testing and implementation of genetic circuits. In addition to guiding complex biosynthesis processes for therapeutic and industrial applications [69-71], these engineered systems hold potential to communicate with and guide synthetic consortia and even native biomes [72]. Recently, synthetic consortia have been developed for leveraging the diversity of multi population systems in ways that expand biosynthetic potential and increase metabolic efficiency [7, 73-75]. The interactions within these engineered communities rely on robust cascades of molecular communication that convey information between cells [76, 77]. As such, system designs need to consider not only the genetic circuits within "designer" cells, but the communication networks that tie them together [9].

In this study we wanted to mathematically characterize molecular signaling that guided previously published experimental results [7] in which two cell-based systems synthesize a model product (green fluorescent protein, GFP) via chemical and electrical induction schemes exploiting different signaling pathways. These are depicted in **Figure 4.1A.** They are both based on induction by hydrogen peroxide. The base case (chemical induction) is actuated by the simple addition of hydrogen peroxide. Then, we had previously developed a means for electronically inducing cells; using simple electrodes, we altered the redox state of inducers [7, 8, 24, 78] and these activate genetic

circuits. We refer to the genetic expression induced by electronic input as electrogenetics[8] and have shown how one can electronically control gene expression, cell attributes [7, 78], and even cell consortia [7, 8, 20, 46, 79]. In our experimental work (**Figure 4.1A)**, we either added H_2O_2 (chemical induction) or we biased gold electrodes (2 mm diameter disk) immersed in the cultures with a –0.55V vs Ag/AgCl reductive potential [7]. This voltage is sufficient to electronically induce cells, it works by reducing oxygen dissolved in the growth media, creating hydrogen peroxide. Cells in the vicinity of the electrode genetically respond to the hydrogen peroxide through an engineered *oxyRS* regulon that activates a genetic circuit via the hydrogen peroxide sensitive transcriptional promoter, OxyR (**Figure 4.1B)**. OxyR endogenously regulates oxidative stress management genes by repressing transcription until its cysteine groups are oxidized into disulfide bonds. The resulting conformation change stabilizes the transcription complex, inducing downstream gene expression [17].

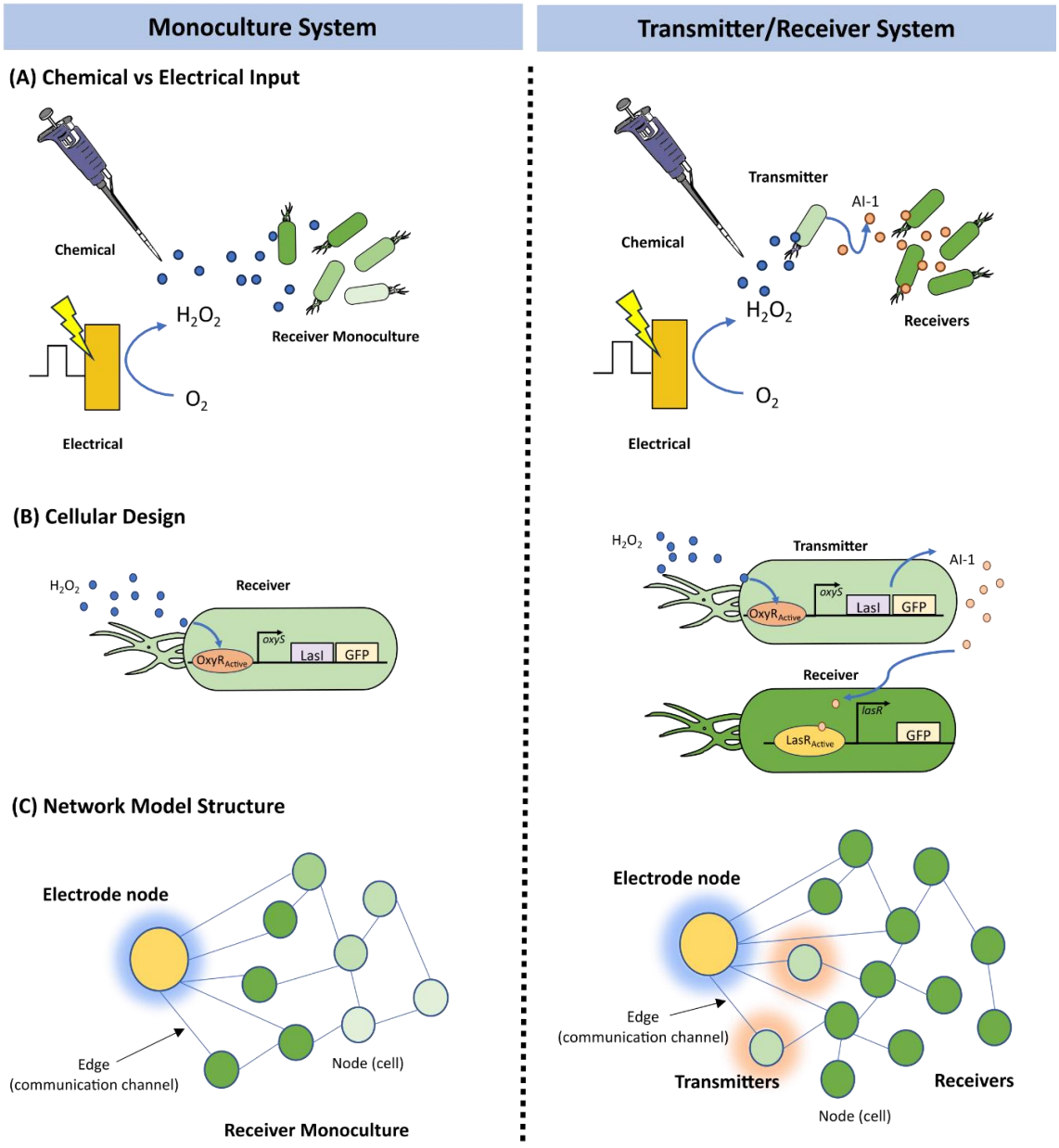


Figure 4.1 Electrogenetic systems overview.

Schematic of the **(A)** cellular design of a Monoculture System (left) in which "Receiver" cells express LasI and GFP in response to hydrogen peroxide induction via the *oxyRS* regulon, and of a Transmitter/Receiver System (right) in which the same "Receiver" cells of the Monoculture system are repurposed as "Transmitter" cells that convey molecular information (AI-1 by the expression of *lasI*) to a second population also denoted "Receiver", but that only express GFP in response to AI-1. **(B)** electrogenetic experimental setup where a biased gold electrode creates hydrogen peroxide as an initial input signal to the cellular systems, and **(C)** an example network model structure for the Monoculture and Transmitter/Receiver systems, saturation of green representing GFP levels and shading around nodes representing inducer production at those nodes.

In **Figure 4.1B**, we illustrate the design of the two systems: (i) a receiver Monoculture and (ii) a Transmitter/Receiver co-culture. In the former case, hydrogen peroxide stimulates LasI and GFP production (**Figure 4.1B**). GFP is the model product in both cases and is easily measured by its fluorescence. LasI synthesizes the quorum sensing molecule N-3-oxo-dodecanoyl-L-homoserine lactone, which we refer to as autoinducer-1 (i.e., AI-1). Quorum sensing signaling molecules enable a collection of cells to take on a population-wide phenotype. In the Transmitter/Receiver system, the same cells used in the Monoculture system are repurposed as "transmitters", where the hydrogen peroxide-induced quorum sensing signal is secreted and then encountered by the "receiver" cells and these respond by producing GFP [7]. Hence, in this two-strain culture one subpopulation turns the electronic signal into a biological signal for subsequent genetic activation and product synthesis in the second subpopulation. Autoinducer-1 is a very strong signaling molecule in that it activates gene expression at nanomolar amounts [46]. This amplifies the original signal to increase gene expression of the desired molecular product.

In this work, we employed a graphical modeling approach which enables a coarse grain interpretation of multicellular systems [80], thus, allowing us to capture agent-based intercellular interactions that fit population dynamics [81]. In **Figure 4.1C**, we depict our model in which each node represents a cell that possesses several weighted attributes: (i) local substrate concentration, (ii) the local inducer molecule concentration, and (iii) GFP expression level. The edges connecting nodes represent a communication channel where signaling molecules may transfer information between nodes. To characterize the movement of these signaling molecules, we implemented a previously developed overlay that approximates a formal diffusion model onto the network architecture [82]. This dramatically reduces computational demand while retaining dynamics of molecular communication and cellular connectivity.

With this model, we then characterized system performance in response to chemical and electrical induction by evaluating GFP production in both schemes. We further explored the effects of spatially fixed cultures (biofilms) in comparison to continuously stirred cultures by varying the edge dynamics in our model. Edges that are fixed reflect static cells, like would exist in a biofilm. Edges that are continually reconnecting between nodes reflect stirred cultures. Then, by utilizing modularity, a graph measure of a network's subcommunity structure [83], we related the network's spatial organization to its signal output. Overall, our model enables a kinetic understanding of signal propagation and GFP production among spatially varied bacterial populations that, in turn, exploit different signaling processes. This provides new hypotheses regarding modes of information transmission and their effectiveness, ultimately leading to new designs.

4.2 Materials and methods

4.2.1 Model formalism

Network initialization was performed by generating a random undirected G(n, m) graph [80] in which there are n total nodes and m total edges that are randomly distributed amongst the nodes. In this network, each node represents an individual cell and edges represent communication channels by which signaling molecules can be transferred between nodes. Each node N_i possesses the following dynamic node weights: $s_i(t)$, $H_2O_{2\,i}(t)$, $AI\text{-}1_i(t)$, and $GFP_i(t)$ corresponding to the cell's substrate, hydrogen peroxide, autoinducer-1, and green fluorescent protein concentrations at time *t*, respectively. In this graph, edges are unweighted and undirected, meaning they do not possess quantitative attributes, nor do they follow any directionality in their connections, i.e. signaling molecules can flow to in either direction between two connected nodes. In our model, time is discrete and represented by natural numbers, evolving forward with each iteration of the simulation as depicted in **Figure 4.2A**. At each timestep a transition is applied in which each attribute of the network is sequentially updated via the following modules: (i) Gene activation, (ii) Molecular production, (iii) Signal diffusion, (iv) Growth, and (v) Edge randomization. That is, a gene activation module is applied, and then activated nodes carry out their respective molecular production models, resulting in increased molecular concentrations at these nodes. Next a signal diffusion module is applied, and molecular concentrations are updated based on the calculated exchange of molecules. Lastly, a growth module is applied to nodes with available substrate; the concentrations of divided nodes are amended. After this, edge randomization may be applied to stirred culture simulations, and the time is forwarded to the next timestep. Thus, the state of the system can be described at any point by the nodes, each with their own set of state variables described by their weights and edges as depicted in **Figure 4.2A**.

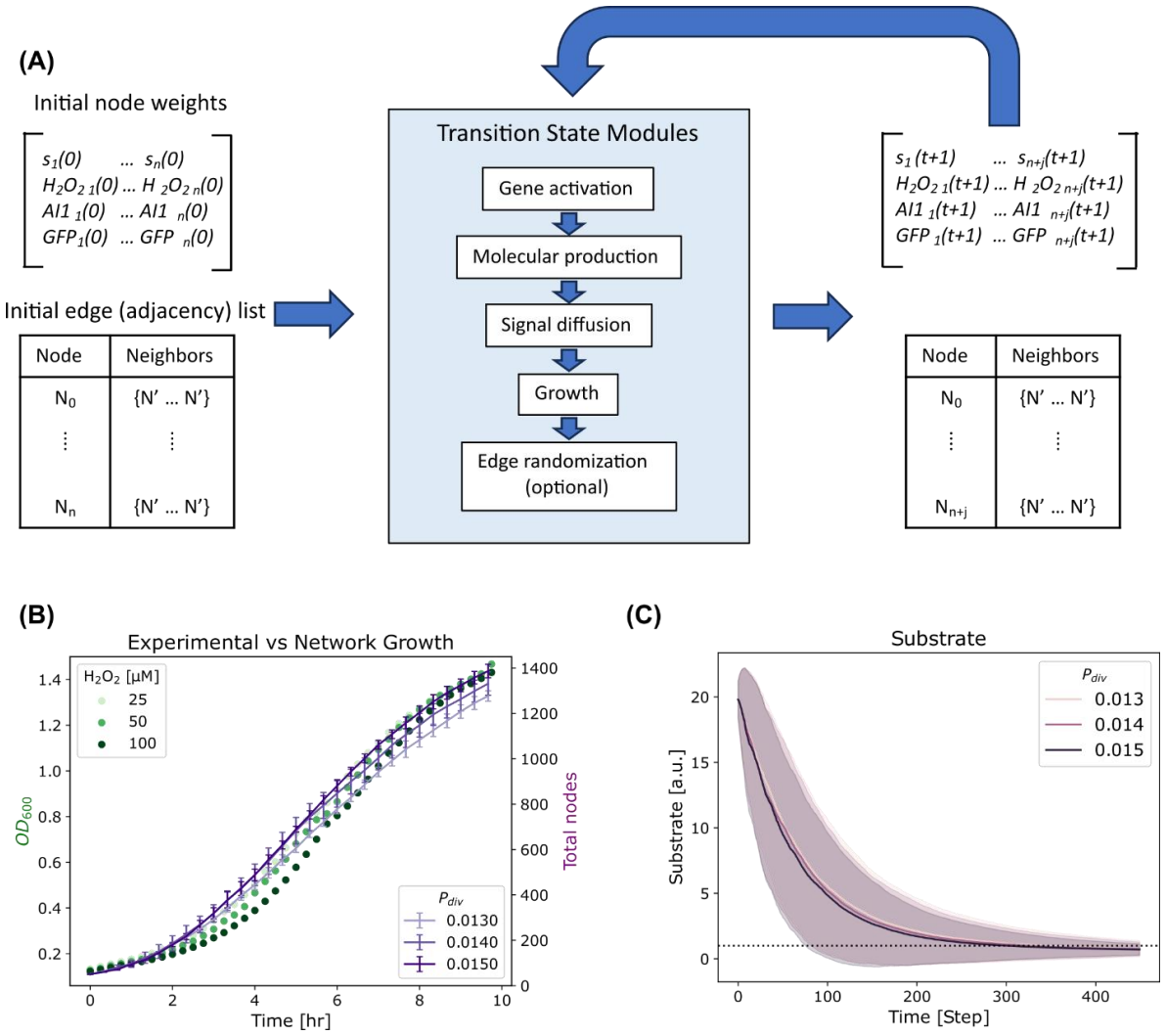


Figure 4.2 Simulation process and growth fit.

(A) Overview of the simulation iterations, where initial state variables and edge structure are updated via transition state modules at each timestep. The output length of the state variables matrix and edge list increase by *j* new nodes. **(B)** Growth measurements for *E. coli* strain OxyR-LasI-GFP (transmitters) with various hydrogen peroxide induction concentrations (chemical addition) are plotted in green, alongside average total nodes of 10 simulation repeats at various division probabilities (P_{div}) over time in purple. Bars represent standard deviation. **(C)** Average substrate per node for P_{div} in (B), the horizontal dashed line indicates a user-specified substrate threshold, *k=1*, below which a node will no longer divide. The shaded zones indicate standard deviation of the substrate concentration across the network for each probability.

4.2.1.1 Gene activation and molecular production

To capture genetic induction and subsequent molecular production we implemented a two step mechanism at each node at every timestep. First, the probability of gene activation is a function of inducer concentration (see **Table 4.1: Equations 1-6)**. H_2O_2 induced gene expression is described by a logistic curve (**Table 4.1: Equation 1-2, Appendix Figure 8.2A**) with a threshold of 12.5 μM. AI-1 dependent gene expression is implemented using a steeper and linear step function (**Table 4.1: Equation 3-4, Appendix Figure 8.2A**) reflecting the nanomolar requirements for induction [7, 84].

If a gene is activated at a node for a timestep (via probability function based on inducer concentration), it will produce the specified molecular product (GFP or AI-1) at a set expression rate based on the prevailing inducer concentration and substrate availability (**Table 4.1: Equation 5-6, Appendix Figure 8.2B-C**). For production based on AI-1, the rate is linear while for hydrogen peroxide it is a saturation function so that at low concentrations there is a steep peroxide dependence and at high concentrations the rate is saturated [7, 9, 46]. These transitions occur at each timestep prior to the diffusion and growth modules, such that molecular production occurs with the concentrations from the previous timestep. The discrete equations are described in **Table 4.2**.

Description	Equation
Hydrogen peroxide induced gene activation probability.	(1) $Prob_{H_2O_2}(H_2O_2 > 0) = \frac{1}{1+e^{-\frac{[H_2O_2]-8}{2}}}$ (2) $Prob_{H_2O_2}(0) = 0$
AI-1 induced gene activation probability.	(3) $Prob_{AI1}(AI1 > 0) = \frac{1}{1+e^{-50([AI1]-0.25)}}$ (4) $Prob_{AI1}(0) = 0$
Hydrogen peroxide induced molecular production rate.	(5) $Rate_{H2O2} = \frac{2[S]}{1+e^{-[H_2O_2]}}$
AI-1 induced molecular production rate.	(6) $Rate_{AI1} = 0.2[AI1][S]$

Table 4.1 Equations for gene activation and subsequent protein production for the inducers: hydrogen peroxide and AI-1.

Variable	Equation
H_2O_2	$H_2O_{2\,i}(t+1) = H_2O_{2\,i}(t) + \alpha\left[\sum_{j\in N_i} H_2O_{2\,j}(t) - H_2O_{2\,i}(t)\deg(i)\right]$
AI-1	$AI\text{-}1_i(t+1) = AI1_{\,i}(t) + Prob_{H_2O_2} * Rate_{H_2O_2} + \alpha\left[\sum_{j\in N_i} AI1_j(t) - AI1_i(t)\deg(i)\right]$
s	$s_i(t+1) = \frac{s_i(t)}{2}$ (if node *i* divides)
GFP	$GFP_i(t+1) = Prob_{AI1} * Rate_{AI1} + GFP_i(t)$ (AI-1 induced) or $GFP_i(t+1) = Prob_{H_2O_2} * Rate_{H_2O_2} + GFP_i(t)$ (H_2O_2 induced)

Table 4.2 State variable dynamics equations.

4.2.1.2 Signal diffusion

Signaling between nodes occurs across edges, such that only nodes connected by an edge may transfer H_2O_2 and AI-1. Signal molecule movement across edges are defined by a discrete approximation of diffusion derived by the following equations as previously described by Sayama[82]:

(1) $\frac{dc_i}{dt} = \alpha \sum_{j \in N_i}(c_j - c_i)$

(2) $c_i(t + \Delta t) - c_i(t) = \left[\alpha \sum_{j \in N_i}(c_j - c_i)\right]\Delta t$

(3) $c_i(t + \Delta t) = c_i(t) + \alpha\left[\sum_{j \in N_i} c_j(t) - c_i(t)\deg(i)\right]\Delta t$

where c_i is the concentration of signaling molecule at a given node i, c_j is the concentration at that node's neighbor j, $deg(i)$ is the number of edges at node i, and α is a diffusion coefficient (See **Appendix Table 8.4** for all coefficient values). In **Equation 1**, diffusion is generalized to the change in concentration at a node with respect to the difference between its own concentration and its neighbors. This can be discretized **(Equation 2)** and solved to find that the change in concentration at a node is determined by the difference between the sum of its neighbors' concentrations and the product of its own concentration and number of edges (**Equation 3)**. At every timestep, the concentration is calculated from **Equation 3** for each node and updated prior to growth module implementation. This process applies to the following state variables and occurs prior to the calculation of network growth: $H_2O_2(t)$ and $AI\text{-}1(t)$. The equations for these variables prior to network growth can be found in **Table 4.2**.

4.2.1.3 Network growth

The network grows with time, depending on substrate availability and growth probability, P_{div}. Initially, each node is assigned the same initial substrate weight, s_0. At each time step, if a node has a substrate level above a minimum threshold, k, the node has the probability P_{div}, that it may divide into two. Following a division event, the substrate (**Table 4.2)**, H_2O_2 and AI-1 node

weights are divided equally between daughter nodes at each timestep. As noted above this occurs after the diffusion module, such that the newly calculated H_2O_2 and AI-1 concentrations may be divided in two upon a division event. As depicted in **Figure 4.2A,** with each iteration the network will increase by *j* nodes, determined by substrate availability at each node and P_{div}. We note that daughter nodes maintain fluorescence (GFP) of their parent's. This assumption is in agreement with previous experiments [77]. We additionally neglect protein degradation, again in agreement with experimental results [77].

After a division event, the resulting daughter nodes share an edge and maintain their parent's edges, limited to a maximum of 10 neighbors. Note, as commonly defined within the field of network science, we refer to neighbors as nodes which share an edge [85]. These 10 neighbors are randomly sampled from the parent's neighbors including those that have divided at that timestep. In a case where a dividing node has 10 neighbors that also all divide at that time step, out of the 20 surrounding nodes only 10 will be randomly selected to share an edge with each daughter.

In **Figure 4.2B,** we depict growth curves for the *E. coli* strain OxyR-LasI-GFP grown with various hydrogen peroxide concentrations [7]. These cells are the receivers in the Monoculture case and transmitters in the Transmitter/Receiver case [9] (**Figure 4.1)**. Alongside we show the total number of nodes over time for a simulated network of with 50 initial nodes, an s_0 of 20, and a *k* of 1 for various P_{div}. With these s_0 and *k* values, each node can divide five times during the growth phase, allowing us to fit the initial node count to 50 and total possible number of nodes to 1600 which approximates 1 node to 0.001 OD_{600}. The P_{div} values assigned helped to ensure that the growth phase of the network translated well to experimental results, such that 45 timesteps represented ~1 hour of cell culture. Our simulation thus mimicked the log phase growth of the cell cultures. We note that flexibility for fitting experiments is enabled by altering the division

probability, P_{div}. Additionally, we note that as the network grows the average substrate per node decreases over time (**Figure 4.2C**), until reaching below the threshold value of k=1. As described above, below this threshold, nodes may no longer divide. The shadowed area in **Figure 4.2C** represents the full ranges of substrate levels across the network and for each division probability. While the substrate defined in our model represents general nutrient availability, the trend shown in **Figure 4.2C** emulates the decrease in glucose over time in *E. coli* cultures demonstrated experimentally [86]. That is, while the network model formalism does not include a typical deterministic Monod model for growth with a maximum specific growth rate and saturation constant, the configuration here well represents the overall culture dynamics.

4.2.1.4 Edge randomization

To describe the spatiotemporal effects of various modes of cell culture such as stirred, immobilized biofilms (static), and combinations thereof, we implemented edge randomization. In the absence of stirring, edges which are assigned during network initialization and at each node division, remain fixed for the duration of a simulation. To simulate a stirred batch culture, we randomized the edges amongst all nodes at every timestep. We simulated two base cases with either static or randomized edges and with or without network growth to demonstrate the effects signaling dynamics: one case where inducers may come from a highly concentrated source node and another case where an electrode may generate inducers at its surface over a specific time period (**Appendix Figure 8.4)**. As anticipated, cases that include network growth and edge randomization resulted in faster homogeneity of signaling molecule concentration across the network than non-growing networks or those growing with static edges. From these tests, we found a set of parameters that when used, enabled reasonable agreement between our previously published data (s_0 =20, k=1, P_{div}=0.015, α=1 and an initial average of 4 edges per node).

4.2.1.5 Electrical hydrogen peroxide generation

To mathematically characterize the production of hydrogen peroxide at the surface of a biased electrode as a mode of information transmission into bacterial cells, we model the input as a signal generated from an individual source node, then link this source to the various nodes. In our network architecture, the electrode is represented by a single node which produces hydrogen peroxide at each time step that it is turned "on". To simulate the actual experimental conditions in which electrical stimulus resulted in negligible growth during the time of induction[8], we set the growth probability parameter, P_{div} , to zero when the electrode is "on" until that time when the growth was observed to increase. We fit the hydrogen peroxide production for an initial network size of 100 to produce 46 μM hydrogen peroxide per timestep to approximate experimental results (**Appendix Figure 8.5A**).

Previously reported experimental results demonstrated that electrical induction yielded lower GFP output compared to a chemical addition, suggesting that the spatiotemporal heterogeneity resulting from the localized inducer production at the electrode's surface effects output. To recapitulate these findings in our model we limited the number of nodes connected to the electrode to 5 percent of the total network at every timepoint. In **Appendix Figure 8.5B**, we plotted the Monoculture response for chemically and electrically induced simulations to demonstrate that the limitation of electrode connectivity to the network reproduces experimental trends, via reduced GFP production compared to chemical induction.

4.2.2 Code and data availability

Graph simulations were performed in Python using NetworkX[87], and modifying and implementing the Simulation class from A First Course in Network Science[88]. Graph generation and initialization and graph transition states were defined and are contained in supplemental notebooks.

Visualizations were performed using Python's matplotlib and seaborn libraries[89, 90]. Experimental data used for parameter fitting are from VanArsdale et al., 2022[7].

Python notebooks and simulation data are available online at github.com/kaychun29/bio-network-simulations.

4.3 Results

4.3.1 Chemical and electrical induction of Monoculture and Transmitter/Receiver Systems

We first simulate the two cellular systems in response to the chemical addition of hydrogen peroxide. We aimed to capture the experimental results depicted in **Figure 4.3A-B** (reproduced with permission), where identical levels of hydrogen peroxide were added to the Monoculture system and to the Transmitter/Receiver System. We later measured GFP expression in all cells via flow cytometry after 3 hours [7]. Flow cytometry provides for the distribution of GFP among all cells in a population. Especially at high concentrations, a chemical addition of hydrogen peroxide should result in a homogeneous input [7] wherein there is little "noise" accompanying induction. In the Monoculture system, increases in GFP became obvious at initial concentrations of 12.5 uM H_2O_2. Further increases in H_2O_2 had relatively little effect on GFP. Interestingly, for the Transmitter/Receiver system, lower initial concentrations of H_2O_2 resulted in significant GFP expression owing to the AI-1 signal propagation. In the end, the yield of GFP for this Transmitter/Receiver system was nearly 10-fold higher than the case with just H_2O_2 added to the monoculture, even at the highest concentrations [7].

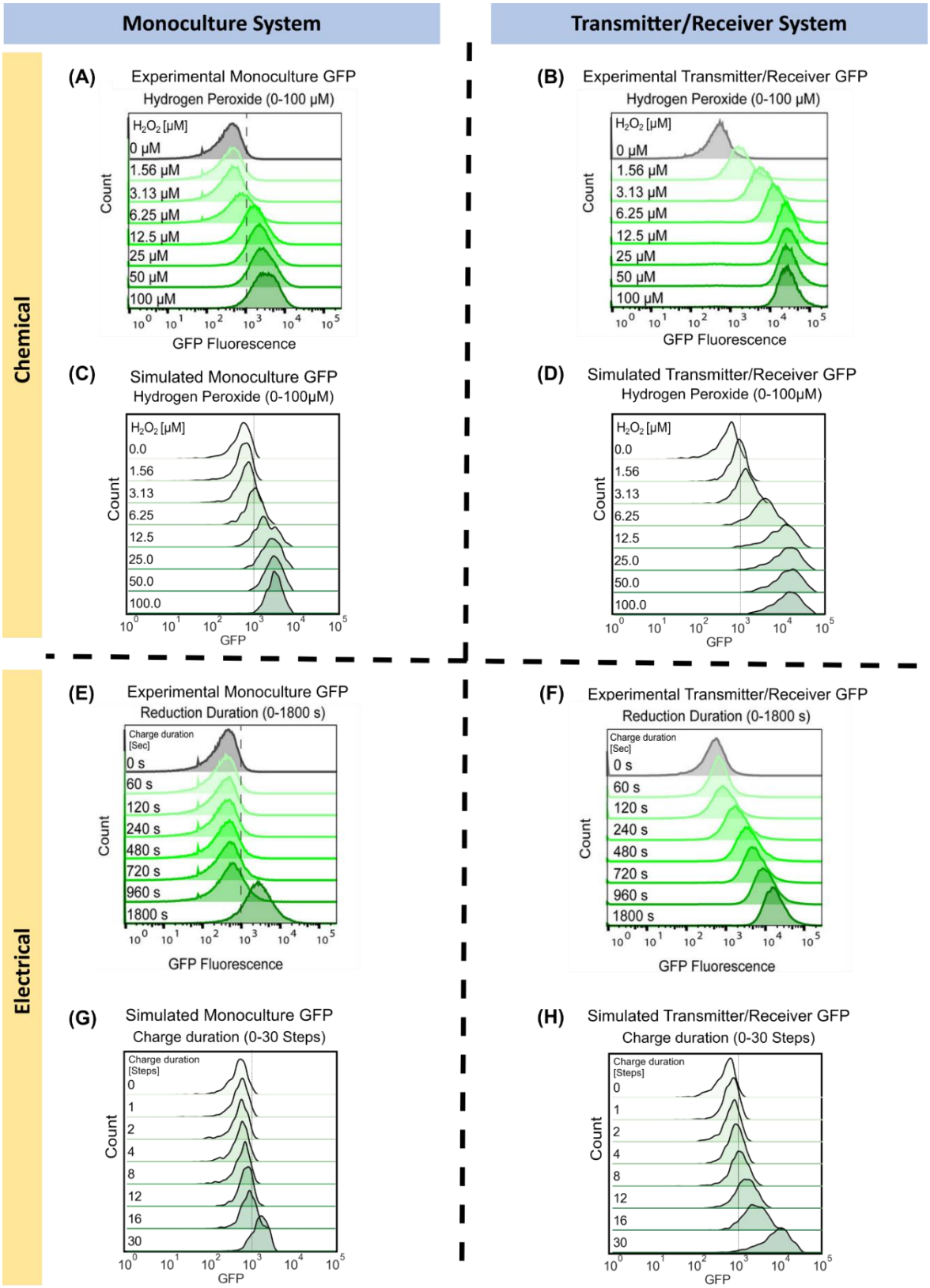


Figure 4.3 Monoculture and Transmitter/Receiver GFP distributions.

Figure 4.3 caption: Chemically induced **(A)** Monoculture and **(B)** Transmitter/Receiver system at 3 hours hydrogen peroxide addition, reproduced with permission from VanArsdale et al [8]. **(C)** The simulated monoculture system GFP distribution at 180 timesteps is shown for an aggregate of 10 replicates, with initial hydrogen peroxide concentration ranging from 0-100 μM. **(D)** The simulated Transmitter/Receiver system's GFP distribution across all nodes at 180 timesteps is shown for an aggregate of 10 replicates, with initial hydrogen peroxide concentration ranging from 0-100 μM. Experimental flow cytometry data of the **(E)** Monoculture and **(F)** Transmitter/Receiver system at 3 hours post charge application, reproduced with permission from VanArsdale et al.[7] GFP distributions of simulated electrical induction for the **(G)**Monoculture system and **(H)** Transmitter/Receiver's receiver GFP distributions across all nodes at 180 timesteps post charge application. Distributions shown are an aggregate of 10 simulated replicates, with charge durations ranging from 0-30 timesteps.

To simulate these results, we assigned each node the same initial hydrogen peroxide weight based on the initial experimental concentration. We set initial GFP weights randomly using a Gaussian distribution with a mean of 500 and standard deviation of 250. This allows for all nodes to have fluorescence background, which fit our previously published experimental distribution for uninduced cells, **Figure 4.3A-B** [7]. For the following simulations the initial network size was 100 nodes, with an average of 4 edges per node. These initial conditions enabled reproducible network propagation, while conserving computational time. We implemented network growth and edge randomization at each timestep to recapitulate the well-mixed growing culture, according to the methods previously described. For the Monoculture system, chemical induction was simulated using the gene activation probability ($Prob_{H_2O_2}$, **Table 4.1:Equations 1-2**) and the molecular production rate ($Rate_{H2O2}$, **Table 4.1: Equation 5**). To model the Transmitter/Receiver system in which a two-strain co-culture is used to amplify the initial hydrogen peroxide signal, we partitioned the initial network into 10 percent transmitter nodes, which function the same as the Monoculture's receivers, and 90 percent receiver nodes which activate GFP production by AI-1 induction. In both systems, AI-1 freely diffuses between nodes at each timestep [46], while in neither case does the GFP diffuse out of the cell [47]. In this Transmitter/Receiver system, GFP production is probabilistically

activated ($Prob_{AI1}$, **Table 4.1: Equations 3-4**) and produced at a rate ($Rate_{AI1}$, **Table 4.1: Equation 6**) dependent on AI-1 and substrate concentration. In **Figure 4.3C-D** we plotted the simulated GFP distributions across the entire network for both the Monoculture and Transmitter/Receiver networks at mid-log growth (180 timesteps) as a function of initial H_2O_2 level. Consistent with the experimental results, the range of expression in the Transmitter/Receiver network reached 10^5, while the Monoculture network's maximum values were ten-fold lower.

We next simulated the electrogenetic approach wherein an applied reducing potential on the electrode generates hydrogen peroxide and this, in turn, stimulates the cells. Naturally, a major difference between this mode of induction is that the hydrogen peroxide is generated at the electrode and while the system is mixed, the peroxide level increases with the extent of its generation rate. The experimental results from earlier work are shown in **Figure 4.3E-F** (reproduced with permission)[7]. In the Monoculture system, small increases in GFP were observed until the cells were exposed to -0.55 V for 1800 seconds. In the previous work, a solution exposed to this reduction duration produced approximately 15 μM of H_2O_2 [7]. Thus, the experimental results for the electrogenetic case were roughly equivalent to the chemical addition of H_2O_2. It was interesting to see that in the case of the Transmitter/Receiver system, a continuous increase in GFP was observed with increased charge. This was previously described as a result of cells near the electrode experiencing sufficient peroxide to induce AI-1, which, in turn, is stable and can be mixed throughout[7].

To simulate electrical induction, we utilized the same model structure as described prior for chemical induction with the exception of initial hydrogen peroxide concentrations. For electrical induction, initial hydrogen peroxide weights were set to zero across the whole network and hydrogen peroxide was produced over a designated charge duration as described in **Methods**. In

Figure 4.3G, we found the simulated GFP distribution of the electrically induced Monoculture system did not increase significantly until greater than 30 timesteps of applied charge (equivalent of 30 minutes), aligning with experimental results in **Figure 4.3E**. For the Transmitter/Receiver system (**Figure 4.3H)**, activation increased nearly immediately, and full activation was attained with 30 steps of electrode charge. Our network model, in all cases, corresponded well with the actual data in **Figure 4.3E-F**, wherein the Monoculture distribution was essentially unchanged until over 960s and the Transmitter/Receiver distribution increased across the span of 960s to reach full activation.

An advantage of the network approach is that one can examine state variables that are otherwise difficult to obtain experimentally. Also, one can more easily align results with underlying mechanisms. In **Figure 4.4A-B**, we plotted the estimated AI-1 distributions for the Transmitter/Receiver networks. While not measured experimentally [7], these simulated values are consistent with expectations. The AI-1 distributions suggest significant heterogeneity within the network. We found this heterogeneity was a result of the variance in activation and spatial distribution of the transmitter nodes and we note this heterogeneity has been reported in chemically induced bacterial cell cultures [77]. We also note that such heterogeneity is not characterized with commonly implemented population scale ODE models, but it can be manipulated experimentally via quorum sensing and genetic circuit design [91]. Our initial network model suggests that there is a level of heterogeneity that is innate to the system and is introduced when amplifying an initial homogenous input through a subpopulation of cells.

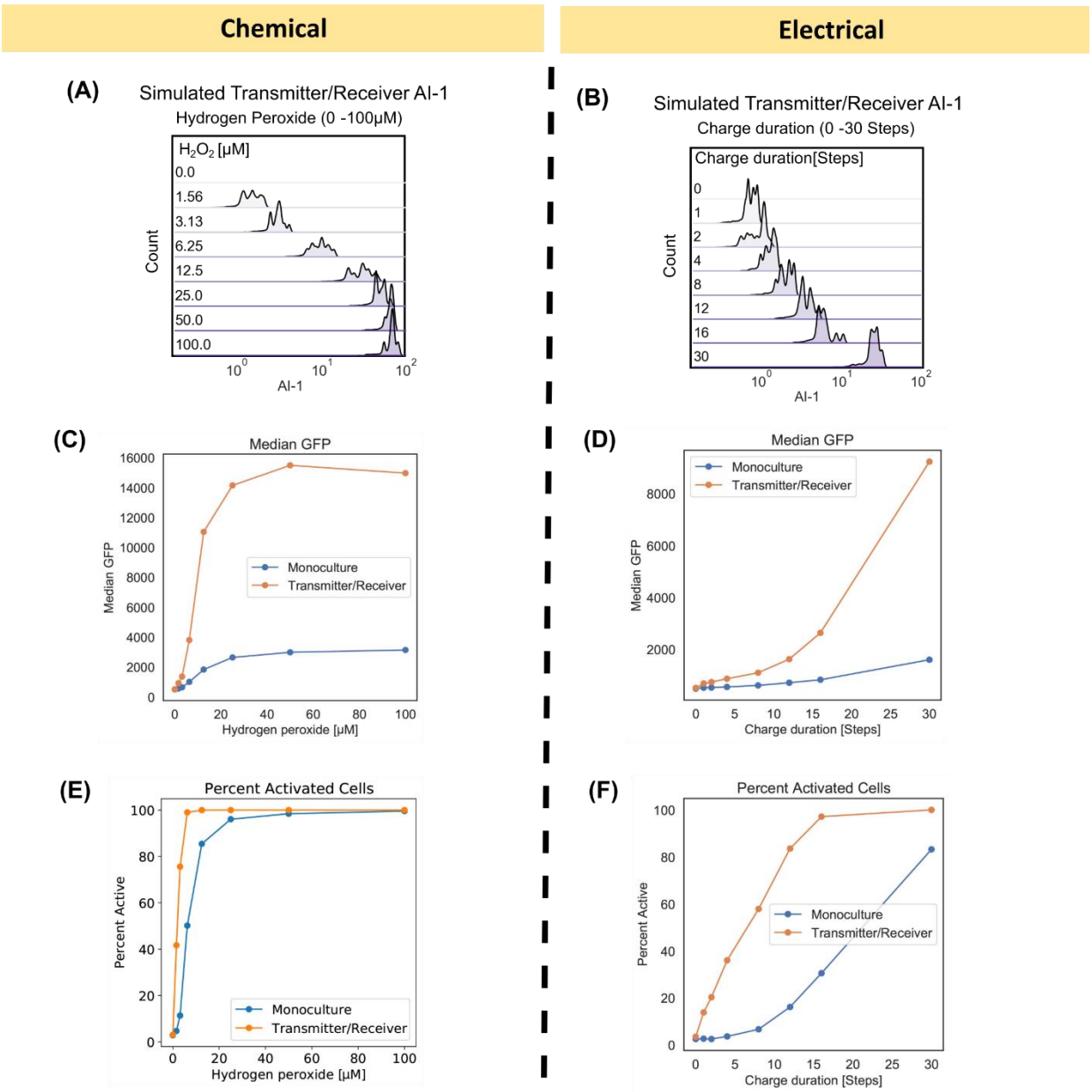


Figure 4.4 Transmitter/Receiver AI-1 distributions and signal metrics.

(A) The AI-1 distribution amongst all nodes in the Transmitter/Receiver network at 180 timesteps is shown for an aggregate of 10 replicates. **(B)** The AI-1 distribution amongst all of nodes in the Transmitter/Receiver network at 180 timesteps post charge application is shown for an aggregate of 10 replicates. (**C-D)** Calculated median GFP from the distributions data shown in **Figure 4.3** plotted over their initial inducer concentration **(C)** and charge duration **(D)**. **(E-F)** Calculated percent active nodes from the distributions data shown in **Figure 4.3** plotted over their initial inducer concentration **(E)** and charge duration **(F)**, threshold for activation was defined at 1000 GFP.

Interestingly, we found that the range of GFP for both Transmitter/Receiver systems was reflected in the AI-1 distributions in **Figure 4.4A-B.** In the chemically induced system, the AI-1 concentrations were between 10^1-10^2 for initial H_2O_2 concentrations above 6.25 µM. Comparatively, for the electrically induced system the AI-1 distribution across the entire network increased incrementally with only the highest charge duration of 30 timesteps producing above 10^1 of AI-1. We further evaluated signal transmission by assessing the median GFP and fraction of activated cells for chemical and electrical induced systems. These serve as metrics for final signal output. The median GFP shows that with electronic induction, expression was generally lower than with chemical induction (**Figure 4.4C vs Figure 4.4D)**, suggesting the signal was attenuated when the inducer was produced at a point source (the electrode node) and needed to diffuse outward among the cells to provide induction.

When comparing the Monoculture to Transmitter/Receiver systems, we observed the amplified response enabled by the Transmitter/Receiver system was readily apparent; the median GFP was above 1.4×10^4 versus 2.5×10^3 for the Monoculture (**Figure 4.4C**), an approximate 5-fold increase, when chemically induced with 100 µM. With electrical induction the median GFP of the Transmitter/Receiver system reached about 8.5×10^3 at the longest charge duration (30 steps), whereas the Monoculture system did not increase above 2.0×10^3, an approximate 4-fold difference. In addition to median GFP we also calculated the percent activated nodes in the network for each initial inducer concentration (by measuring the number of nodes with GFP above a 10^3 threshold). In **Figure 4.4E**, we plotted chemically induced systems and observed that although both systems ultimately reached 100% activity, the Transmitter/ Receiver system reached this peak at lower H_2O_2. For the electrically induced systems, the portion of active nodes increased incrementally and monotonically with charge (**Figure 4.4F**). We note that the Monoculture system

had a consistently lower percentage of active nodes than the Transmitter/Receiver system, as expected, and never reached 100% by with 30 timesteps of induction. Overall, our model simulations corresponded well with the previous data (**Figure 4.3A-B, E-F**). Our simulations also suggest that despite the heterogeneity or "noise" that is introduced by amplifying the initial signal through a subset of cells (electrode induction), the molecular amplification that was enabled by transforming the H_2O_2 into a stronger secondary signaling molecule, in particular one that evokes a quorum sensing response, overcame that disruption, and produced high levels of signal and activation.

In **Figure 4.5**, we explored further the dynamics of H_2O_2, AI-1, and GFP for the chemically and electrically actuated cases by plotting their average (lines) and standard deviation (shaded) across the network over time. We chose representative cases with similar average H_2O_2 concentrations. In **Figure 4.5A**, we depict the simulated H_2O_2 dynamics for the chemical addition of 6.25 mM H_2O_2 and for the electrical induction at 12 timesteps of applied charge (~ 6 μM of hydrogen peroxide generated). The widely distributed H_2O_2 level in the case of electrical induction was expected, but the average concentration simulated was quite similar. We note, **Figure 4.5A** depicts Transmitter/Receiver H_2O_2 dynamics, however Monoculture dynamics were nearly identical suggesting the type of cellular system does not affect hydrogen peroxide diffusion and generation. Despite the comparable average H_2O_2 levels in the systems over time, the AI-1 concentration of the Transmitter/Receiver system was nearly 2-fold higher that of the chemical induction (**Figure 4.5B)**. In general, the GFP levels produced by both the Monoculture and Transmitter/Receiver systems (**Figure 4.5C-D**) were higher for the chemical addition relative to the electronically induced systems. This was understandable because the electrode produced H_2O_2 levels were found to be widely dispersed, indicating that many cells likely encountered minimal

levels of inducer (**Figure 4.5A**)**.** When comparing the Monoculture versus Transmitter/Receiver GFP dynamics (**Figure 4.5C vs 4.5D)**, GFP expression in the Monoculture increased consistently over time whereas the Transmitter/Receiver network expression was slightly delayed initially during which time AI-1 was produced (~50 steps corresponding to peak AI-1) and subsequently accumulated. For both modes of induction, the Transmitter/Receiver GFP yields were higher irrespective of a delay in production.

Overall, we note that the large standard deviations depicted in **Figure 4.5** reflect substantial heterogeneity within the network. We suggest this heterogeneity is rooted in the wide signaling molecule distribution that can occur when cell numbers are low (early on) and when electrodes are used to generate hydrogen peroxide. In the latter case, this signal molecule interacts with cells in a random and distributed manner. In the experimental system, an uninduced cell needs to be transported near an electrode to receive H_2O_2. At further distances the peroxide could be depleted so that cells far away never experience high levels. Interestingly, our network model seems to well characterize the extent of signal propagation and the effects of its design structure in determining system outcome. The tradeoffs between the delay in responses and expression levels provide insight on system design. They also suggest spatial heterogeneity, we explore this as a potential design feature as follows.

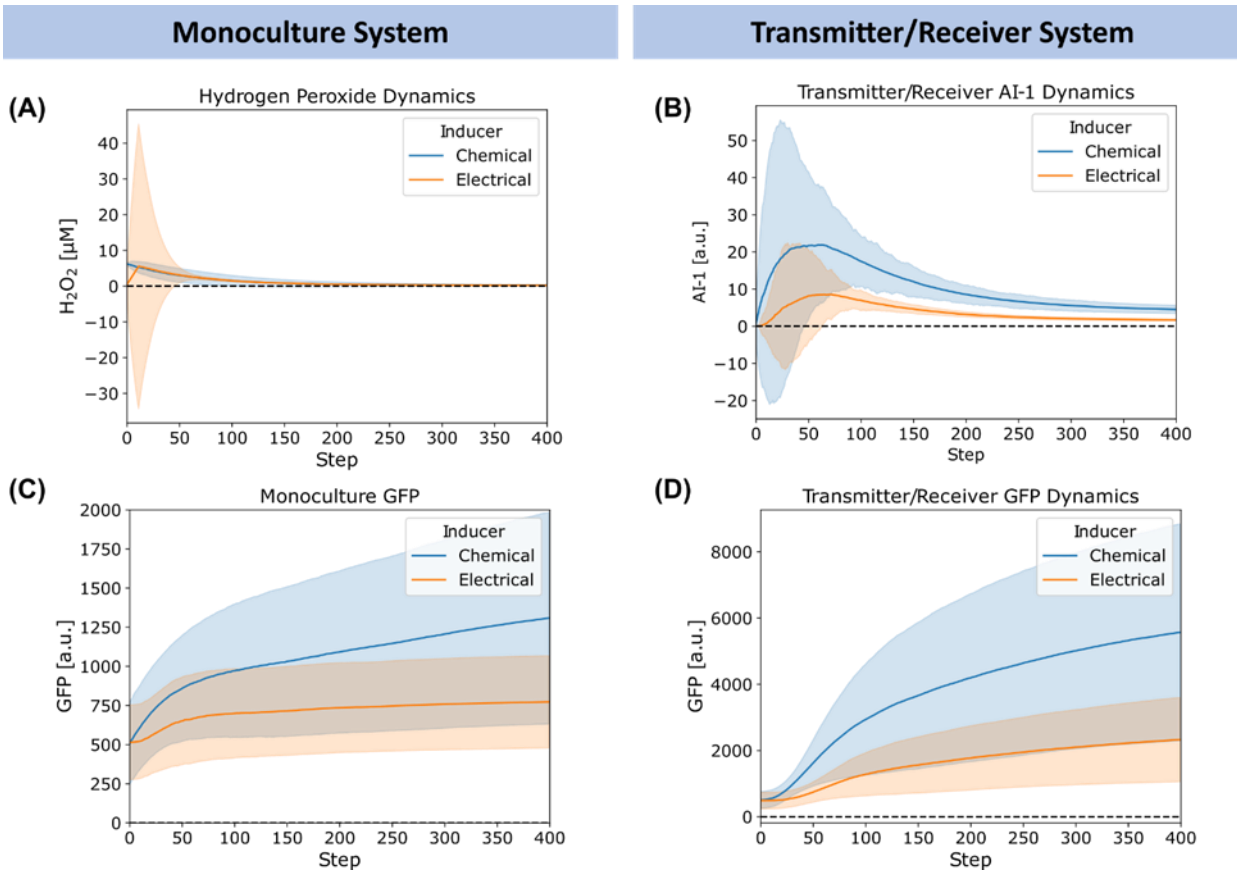


Figure 4.5 Chemical versus electrochemical signaling dynamics for edge randomized networks.

(A) Average hydrogen peroxide, **(B)** Average Monoculture GFP, **(C)** Average Transmitter/Receiver GFP, and **(D)** Average Transmitter/Receiver AI-1 concentrations over time for a 6.25 μM hydrogen peroxide induced chemical addition (blue) and 12 step charge duration (orange) across the entire network. Error bars appear as shaded regions, representing standard deviation of aggregated network data from 10 simulation replicates.

4.3.2 Spatial design via network topology: Graph modularity and edge dynamics' effect on signaling

Based on our successes in characterizing experimental data from both the chemical addition of hydrogen peroxide and its electrode-based generation for both the Monoculture and the Transmitter/Receiver systems, we decided to interrogate the design space for altered induction methodologies. Specifically, we next explored how the relative spatial distribution of cells (nodes) could affect the signaling. We decided to test a case where we retain transmitter cells directly onto the electrode. Thus, in **Figure 4.6A**, in addition to the (i) a chemical addition and (ii) electrical induction cases previously described, we added (iii) electrical induction of transmitter cells that are fixed to its surface. This last network structure captures experimental designs in which cells are either engineered to bind to gold electrodes [9] or that are retained in an assembled hydrogel film [26]. Cells localized in this manner could receive electronic signals (hydrogen peroxide) and then transmit their "message" to cells outside of the film through signal synthesis, secretion, and transport to cells occupying the liquid proximal to the electrode and beyond [26]. For affixed cells, instead of randomizing edges at time steps, we fixed edges and maintained them throughout. This mimics a static system, representative of a biofilm [92, 93] or a set of cells localized on an electrode [9].

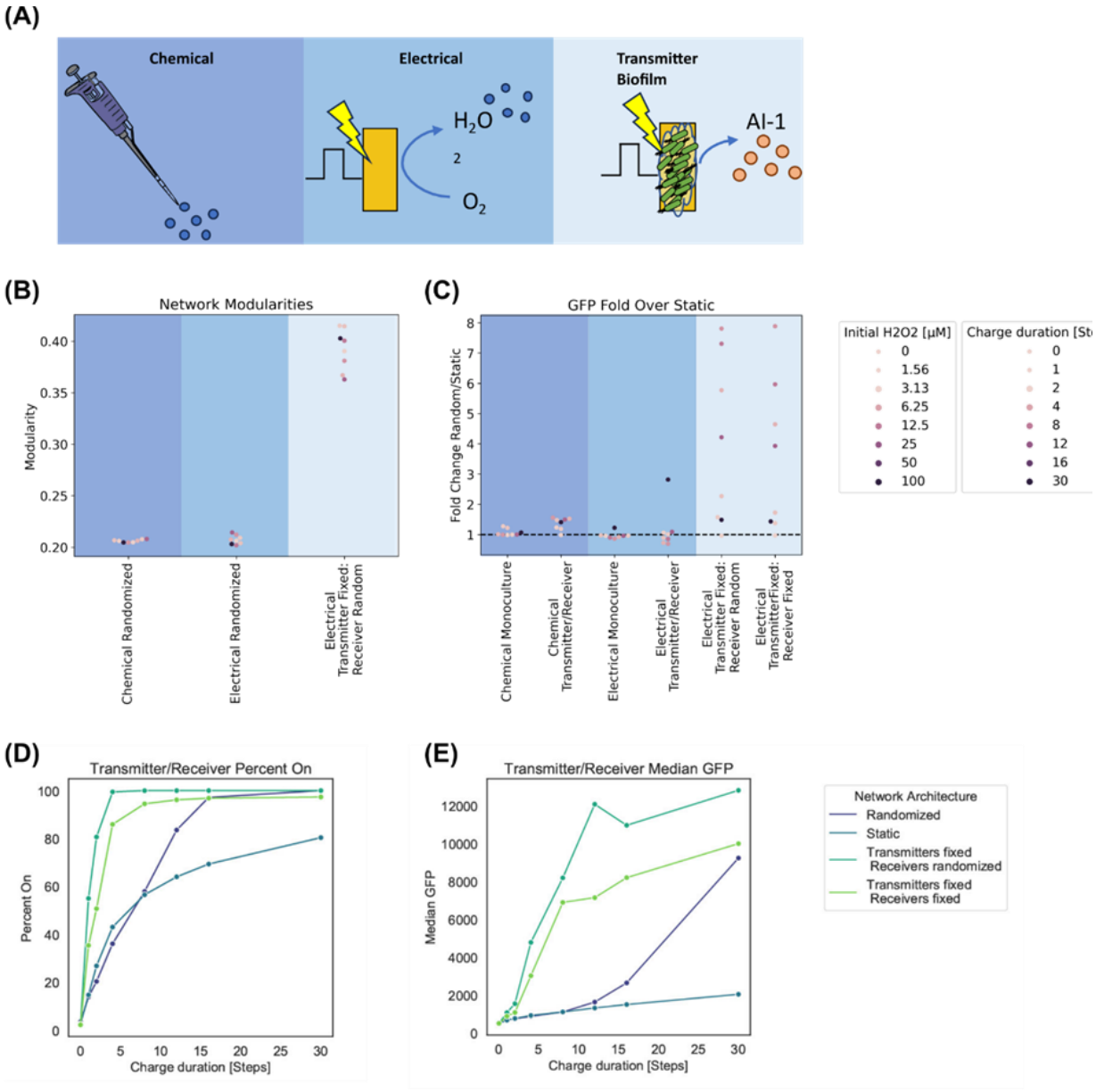


Figure 4.6 Modularity and fold change dependent on network structure.

(A) Graph schematic of the three spatial configurations tested. **(B)** Network modularity of differing node arrangements and edge dynamics at timestep 180. **(C)** Fold change in GFP of randomized edge networks over static edge networks for the Monoculture and Transmitter/Receiver systems with either chemical or electrical induction at 180 timesteps post induction. **(D-E)** Signal transmission metrics for Transmitter/Receiver network architectures at 180 steps post 30 steps of electrical induction, calculated from an aggregate distribution of 10 simulation replicates. (**D)** Percent active nodes for varying charge duration times. (**E)** Median GFP across network for varying charge duration times.

To quantify structural variation that emerges due to growth and edge dynamics, we used modularity [83, 85, 94] as a measure of network structure (**Appendix Figure 8.6A)**. In general, modularity describes how well a network is partitioned into various sub-communities [85]. A single community wherein the connections are near random is represented by a modularity value of zero, while a network where all edges fall within the same community would have a modularity of 1 due to its strong community structure[83]. For our calculations, we use the Louvain method to calculate the modularity as it is computationally efficient in finding high modularity partitions of large networks [94].

In **Figure 4.6B**, we depict the calculated Louvain modularity at 180 timesteps for the cases above (chemical and electrode induction for mixed cultures), as well as the new case where transmitter cells are fixed to the electrode (initial 10 nodes) and the receiver cells are not fixed. For the Transmitters fixed onto the electrode, dividing nodes inherit the edges from their parent nodes without further edge randomization. As expected, our results show that there was increased modularity calculated in the case where some cells are fixed (Transmitters) and some are free to move (Receivers). In general, we found that the modularity of randomized networks was lower than static networks (see **Appendix Figure 8.6B** for simulations of completely fixed systems, not shown here). This is understandable because randomized distribution of edges among the nodes yields an unorganized network structure. In comparison, as static networks grow, they maintain structure.

We further ran simulations with fixed edges for different charge durations and hydrogen peroxide concentrations as in the earlier simulations, to examine static biofilm cultures relative to well stirred systems. We found differences in charge duration and initial hydrogen peroxide concentrations did not affect the modularity as molecular concentrations that are represented by

node weights do not affect the spatial structure of the network. We then analyzed the output (i.e., GFP level) for these simulations. To compare the output of these static cultures we calculated the ratio of average GFP at 180 timesteps from randomized networks to the static networks. We use this as a way to measure the benefit of cells in the traditional well-mixed system to those in a fixed or partially fixed system (i.e., cells fixed to an electrode propagating signals to those in fluid nearby). In **Figure 4.6C**, we plotted these ratios for each inducer and system type. For the new case of transmitters fixed to an electrode, we also tested a case in which the receivers are also fixed to emulate multilayer deposition of cells onto an electrode as a potential design. This is more representative to a complete biofilm. The fold change calculated from these transmitter fixed cases were done relative to static networks of electrically induced Transmitter/Receiver simulations.

Here we see that for chemical addition, there was little difference between the network structures. This results from the fact that all nodes experience the same initial inducer concentrations. For electrically induced systems, there was minimal effect on the Monoculture at all charge durations. In the Transmitter/Receiver system, we found that for 30 steps of charge there was an approximately 3-fold increase in signal when randomizing the network. In fixed transmitter simulations, we found a substantially larger range for the overall system output. These fold increases are indicative of how edge randomization generally increases output while strategic spatial arrangement of the co-culture with respect to inducer sources can largely amplify signal throughput.

Finally, we assessed how topological effects leading to increased modularity affect signaling within the network. We calculated the percentage of cells that are active (GFP above a 10^3 threshold) and the median GFP for these Transmitter/Receiver simulations with various edge dynamics (**Figure 4.6D-E)**. We observed that the static networks had both the lowest median GFP

and the fraction of active cells **(Figure 4.6D-E)**. Interestingly, our simulations suggest that introducing transmitters that are fixed to the electrode increases the overall activation and median GFP over completely randomized networks, and this is irrespective of receiver conformation (fixed or not). We suggest this is due to the faster and increased AI-1 production that is enabled by transmitter proximity to the electrode **(Appendix Figure 8.6C)**. Randomizing the receivers further increased estimated output. This is a consequence of allowing the whole receiver population's increased contact with the transmitter population, as the AI-1 source. This was evident as the network's GFP distributions where increasing static network components correlate to a wider range in GFP values (**Appendix Figure 8.6D)**. Overall, these results reveal that while high modularity yields increased signal heterogeneity, it also lowered signal output compared to low modularity networks. That said, the strategic or intentional organization of subpopulations can drastically increase output, despite increased modularity.

4.4 Discussion

In this work, we developed a graphical network approach for modeling multi-population bacterial cultures. By coarse graining the cell-to-cell signaling interactions that are known to occur in complex bacterial systems [4] and leveraging intrinsic network properties that attempt to simulate spatial distributions, we have elucidated signal dynamics that would be very difficult to ascertain using traditional deterministic population scale multicellular modeling. The implementation of a graph-based model allowed us to vary network structures that we had previously implemented experimentally. We were able to determine network parameters (probabilities of growth, molecule production, gene activation) that when employed in the model, accurately recapitulated the experimental observations. Then node weights (other state variables such as inducer levels, substrate levels, etc.) were examined to better understand the experimental results. Perhaps more importantly, with this agreement we then tested hypotheses regarding the spatial composition of microbial systems. Further, by implementing various edge architectures, we attempted to mimic various engineered and endogenous culture structures. We mimicked stirred batch conditions common to biomanufacturing settings via edge randomization. Static edge conformations imitate biofilms found in nature and other immobilized or hydrogel-assembled cell systems. Additionally, we could easily accommodate varied edge profiles in our model so that we could test how relative spatial structures affect communication between different populations.

Owing to the natural tendency to think in terms of subpopulations and quorum sensing [77], we introduced the notion that network modularity would be a valuable tool in analyzing bacterial networks when organized in the various experimental configurations. In testing fixed spatial conformations we found that for increased modularity, meaning more subcommunities in the network, maximum signal throughput is reduced and delayed for simulations with an electrode as

an input source. We suggest this is attributed to the need for the input signal to diffuse into each subcommunity for the secondary signal to then be produced and diffused back out for further signaling. We suggest this introduces an increase in noise at each step of signal transmission due to structural constraints. That said, these decreases in signal can be overcome by spatially orienting transmitter nodes close to the electrode as the input signal source. We further tested fixing all transmitter nodes to the input signal source (the electrode) and found that regardless of whether the receivers were fixed or randomized this restored signal in fixed networks and resulted in higher expression than in randomized simulations. Correspondingly, in Terrell et al. [11], they demonstrated that by fixing microbes to a gold electrode they could produce AI-1 with electrochemical stimulation, and this was shown to be quite successful in signal propagation (more so than in VanArsdale et al.[7], where the transmitter and receiver populations were fully mixed in a stirred vessel). Unfortunately, in neither case was it experimentally feasible to monitor the AI-1 diffusion and activation across the system boundaries [9]. Here, our work may provide theoretical insight into the signaling occurring in these types of experimental configurations and those found in natural biofilm systems, where measurements in real time and at small distances is difficult.

Additionally, we suggest that models such as this can be further extended to simulate other spatial conformations of cell populations to provide insight into how much input and signal transmission is necessary for successful outcomes [84]. These include cases where synthetic assembled consortia of higher complexity may be cultured together in batch or spatially fixed within gels [95], between membranes or 3D printed niches [96], or within varying ecological niches [15, 16, 92, 93, 97]. For example, the field of biomaterials has implemented the spatial confinement of cells within hydrogel structures and microcapsules for the use in generating functional living materials and to recreate micro communities found in nature [98-101].

Chapter 5 Mediated electrochemical probing for rapid L-cysteine quantification and antibody quality assessment using machine learning

This chapter is adapted from an unpublished manuscript.

5.1 Introduction

The monitoring of media composition and product quality throughout bioproduction is of great interest for process optimization and development. In this chapter, we developed electrochemical methods to quantify L- cysteine and reduced IgG antibody to contribute a novel rapid approach to monitoring biomanufacturing processes. Although considered endogenously non-essential through its production in the liver, L-cysteine is an important media component for biomanufacturing of proteins [102]. Cysteine residues within protein structures are critical for forming vital tertiary structures through the formation of disulfide bonds [103]. The presence of cysteine is also necessary for retaining a redox balance in cell cultures as it is an essential for the production of antioxidants such as glutathione, hypotaurine, and taurine [104]. However, cysteine is relatively unstable as it can easily autoxidize its thiol groups to form thiyl radicals and its dimer L-cystine, both of which can contribute to an increased oxidative state of the media [104-106]. Studies examining L-cysteine concentrations in basal media during antibody producing fed-batch CHO cell culture showed that increased initial L-cysteine concentrations increased oxidative stress due to excessive consumption and subsequent depletion of intracellular antioxidants [107]. Whereas studies with decreased cysteine concentration in their feed composition can lead to decrease in protein titer and increased oxidative stress due to the depletion of antioxidant production [108, 109]. Thus, the development of methods that can rapidly monitor L-cysteine with the intent of maintaining concentrations that balance the redox state for optimized productivity is desirable. The need for monitoring thiols is specifically relevant

in antibody production, as monoclonal IgG1 antibodies contain 16 thiol residues; 4 of which are critical to antibody structure in that they hold together the heavy and light chains [110]. The reduction of these key disulfide bonds results in fragmentation of the biologic, and ultimately a decrease in activity and stability of the desired therapeutic [103, 111-113].

Traditional analytical methods to measure the fragmentation of antibodies, such as capillary electrophoresis and other chromatography methods, can have long lead times due to labor intensive processing and logistics [114, 115]. We suggest that the development and implementation of rapid electrochemical approaches for monitoring L-cysteine and reduced antibody concentration could provide quick on-demand quantification to complement gold standard analytics, allowing manufacturers to catch mishaps and intervene in real time.

Here, we expand on the previous methods for electronically sensing molecular structure by developing a pipeline for rapid sampling and quantification of L-cysteine and reduced monoclonal antibodies. In Motabar et al., researchers described the use of mediated electrochemical probing (MEP) enabled the detection of L-cysteine and the free thiols on reduced IgG in solution [116]. This methodology consisted of adding the redox mediator Ferrocene to IgG samples and electrochemically measuring them using both chronocoulometry and cyclic voltammetry. The measured electrochemical signals were then used to calculate metrics against a background reading of mediator alone in solution. This analysis revealed how interactions between the mediator and IgG thiol groups effect amplification and attenuation the signals. In this method, electrochemical measurements took 4 minutes and the metrics calculated were used for binary classification; to determine if samples of the same total protein concentration were reduced or intact. In all cases, the methodology required parallel testing of the test sample and the control. In this chapter, we optimized an MEP method for 10 second sampling time (24-fold faster), from which data was fed

into a trained machine learning regressor to determine its L-cysteine or reduced IgG concentration; the overview of this workflow is depicted in **Figure 5.1**. This obviated the need for a parallel mediator-only control.

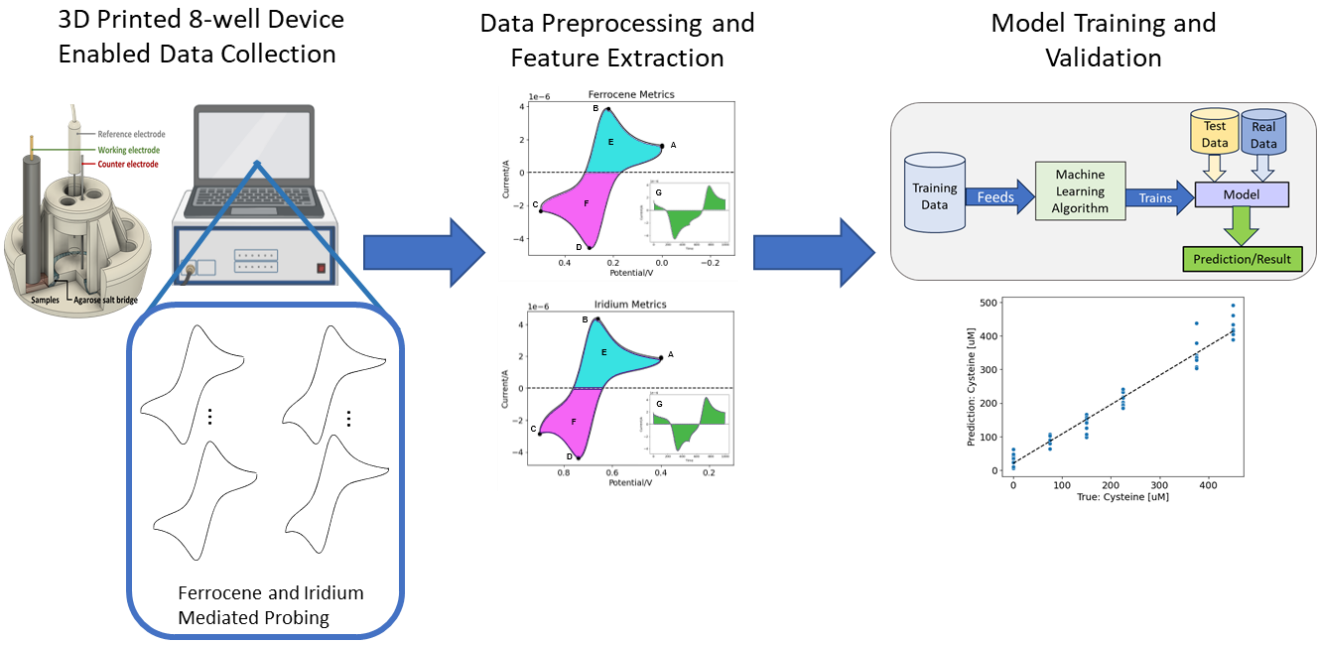


Figure 5.1 L-cysteine and IgG workflow overview.

To develop this workflow, we began by screening and optimizing electrochemical parameters on L-cysteine dissolved in Phosphate Buffered Saline (PBS) and Dulbecco's Modified Eagle Medium (DMEM). These represent low and high background solutions, respectively. Screening of various parameters were streamlined by utilization of a 3D printed 8-well sampling device and 8-channel potentiostat (see **Methods**). The datasets generated by this initial screen were then used to test various preprocessing approaches and machine learning models. Top L-cysteine quantification pipelines demonstrated high correlation of predicted concentrations with true concentrations (R^2>0.8) on test sets. The electrochemical parameters with the highest performing L-cysteine quantification pipelines were then selected to measure antibody samples wherein the antibodies were intentionally reduced to expose cysteine thiols to the mediators. These antibody datasets were then used to test various pipelines. The resulting top performing models were found to quantify reduced IgG concentration in both low and high background solutions with a positive correlation to its true concentrations (R^2>0.74). In this report, we describe the development of this methodology as a proof-of-concept workflow, which we believe may be transferred to other biologically relevant systems for thiol detection or for detection of other redox accessible residues enabled by alternative redox mediators.

5.2 Results and discussion

In mediated electrochemical probing [117], redox active molecules that serve as "mediators" (i.e., they exchange electrons with electrodes and other molecules) are intentionally added to a biological sample and then measured voltammetrically to quantify the mediator's redox state using a simple electrode setup. The interactions between the mediators and biologics cause the ratio of the reduced to oxidated form of the mediator to become altered. This is captured by voltametric measurements at the electrode surface, where the recorded current's redox peaks are altered compared to a mediator only control. Here, we utilized cyclic voltammetry [25] as the electrochemical method to measure our samples, as it provides both information on the reductive and oxidative nature of the sample by sweeping the applied potential in both directions.

5.2.1 Electrochemical parameter optimization

We explored the following parameters of our electrochemical measurement to compare their effects on L-cysteine dependent signal: scan range, mediator selection and mediator concentrations. We first optimized these parameters to a 100 mV/s scan rate such that the measurement time would be 10 seconds per sample, a 24-fold improvement upon previous methods that used 2 mV/s for a total of 4 minutes per sample [116]. We explored two mediators which have shown prior interactions with L-cysteine; ferrocene dimethanol (Ferrocene/Fc) and potassium hexachloroiridate(III) hydrate (Iridium/Ir). Ferrocene has a redox potential of E_0= 0.25, and as a mild oxidant it has shown selectivity towards redox cycling with L-cysteine [26]. Iridium has a redox potential of E_0=0.7, and is a stronger oxidant than ferrocene that has recorded redox interactions with multiple amino acids such as L-cysteine, L-tyrosine, and L-lysine [26].

Previous methods that implemented Ferrocene for MEP used a scan range of 0 V to 0.5 V, capturing a range of +/- 0.25 V of the redox potential, E_0=0.25 V [26, 116, 117]. We sought to test

whether increasing the scanning voltage range could increase the richness of the data collected. We found that in a low background solution, in this case PBS, there was minimal qualitative differences between the cyclic voltammograms captured at various cysteine concentrations with 50 µM of Ferrocene at 0 V to 0.5 V compared to -0.2 V to 0.9 V (**Figure 5.2A**). However, when measuring these same conditions in a high background solution, DMEM cell culture media, we found that expanding the scanning voltage range resulted in decreased signal range for the ferrocene peaks and the appearance of an oxidative peak at higher voltage ranges (0.6-0.9 V) indicative of oxidation of media components (**Figure 5.2A**). The oxidation of these media components at higher voltages reduced the range in signal of the ferrocene peak. We suspect this is due to the oxidation of media components at higher voltages, rendering them unable to further cycle with ferrocene in the lower voltage ranges. We demonstrated this by measuring one cyclic voltammogram cycle of cysteine samples with 50µM of ferrocene at the higher voltage range, and then subsequently measuring them at the lower voltage range in **Supplemental Figure 5.1**. The sample measured at a high and then low scan range resulted in reduced current signal and a weak correlation to cysteine concentration compared to measurements taken only at the lower voltage range. This phenomenon was repeated several times and was consistent (not shown here).

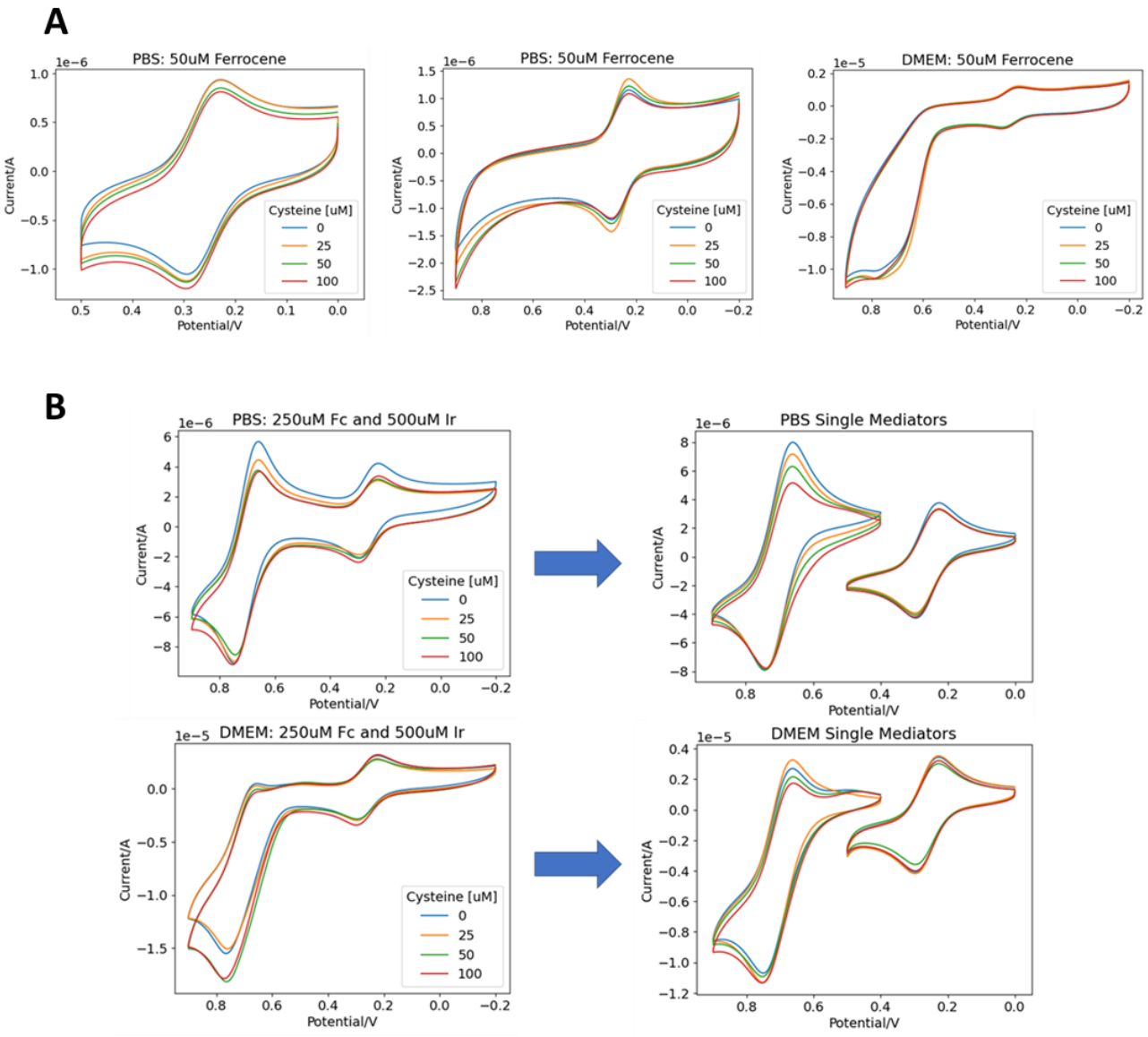


Figure 5.2 Scan range screening and mediator multiplexing.

(A) Scan range screening. Cyclic voltammograms were captured using 50µM Ferrocene probing for samples containing 0, 25, 50, and 100 µM cysteine. Scan ranges and background solutions were as followed from left to right: PBS 0V to 0.5V, PBS -0.2V to 0.9V, and DMEM -0.2V to 0.9V. **(B) Multiplexing.** Cyclic voltammograms captured with 250µM Ferrocene and 500µM Iridium probing in PBS (top) and DMEM (bottom). Left plots were probed by multiplexing both mediators in solution at the same time. Right plots were measured with each mediator seperately. All plots captured voltammograms samples of 0, 25, 50, and 100 µM cysteine.

We also investigated whether multiplexing mediators would increase the data's richness by adding both ferrocene and iridium to samples of various cysteine concentrations. However, in both PBS and DMEM solution backgrounds, we observed that having both mediators over a wide voltage range caused poor separation between cysteine concentrations compared to the single mediators within their own voltage ranges at the same concentrations (**Figure 5.2B**). Similarly to expanding the voltage range, we suggest this signal reduction when multiplexing may be attributed to both mediators interacting with overlapping components in solution. In this case, once one mediator oxidizes or reduces a component in solution the other mediator may be left with no substrate to interact with. That is, there are likely interactions between entities in solution that are not well defined and that degrade the quality of the data owing to the attenuation of characteristic peaks specific to the added mediators.

After establishing that increasing scan range and multiplexing mediators do not enhance our voltametric measurements, we collected datasets by probing a range of cysteine concentrations (0-500 µM) with various concentrations of each mediator (0-1 mM) individually. That is, we prepared cysteine samples in solutions of PBS and DMEM, then added stock solutions of mediator to each sample immediately prior to its measurement to obtain the desired mediator concentrations . This constituted 80 conditions per solution. We collected these measurements using a custom 3D printed 8-well device which fitted a 2mm standard glassy carbon disk electrode in each well and allowed for measurement volumes of 100 µL per well, see **Methods**. Using this device, we were able to collect 80 cyclic voltammograms per condition by utilizing each well simultaneously and collecting 10 sequential cycles, yielding a total of 640 samples across the screen. In **Figure 5.3A**, we provide results of various concentrations of Ferrocene and Iridium in PBS and DMEM. Across mediator concentrations in both solutions, we observed their current amplitudes increase with mediator concentration, as expected, given increased availability for electron transfer with higher mediator concentrations. We also observed a significant oxidative background current in the DMEM measurements without any mediator in voltage ranges above 0.6 V, owing to the many DMEM components that can be oxidized. Interestingly, measurements of the mediators in DMEM revealed how cell culture components modify the mediator's current signature, notably causing a background oxidative peak amplification and reductive peak attenuation across iridium concentrations.

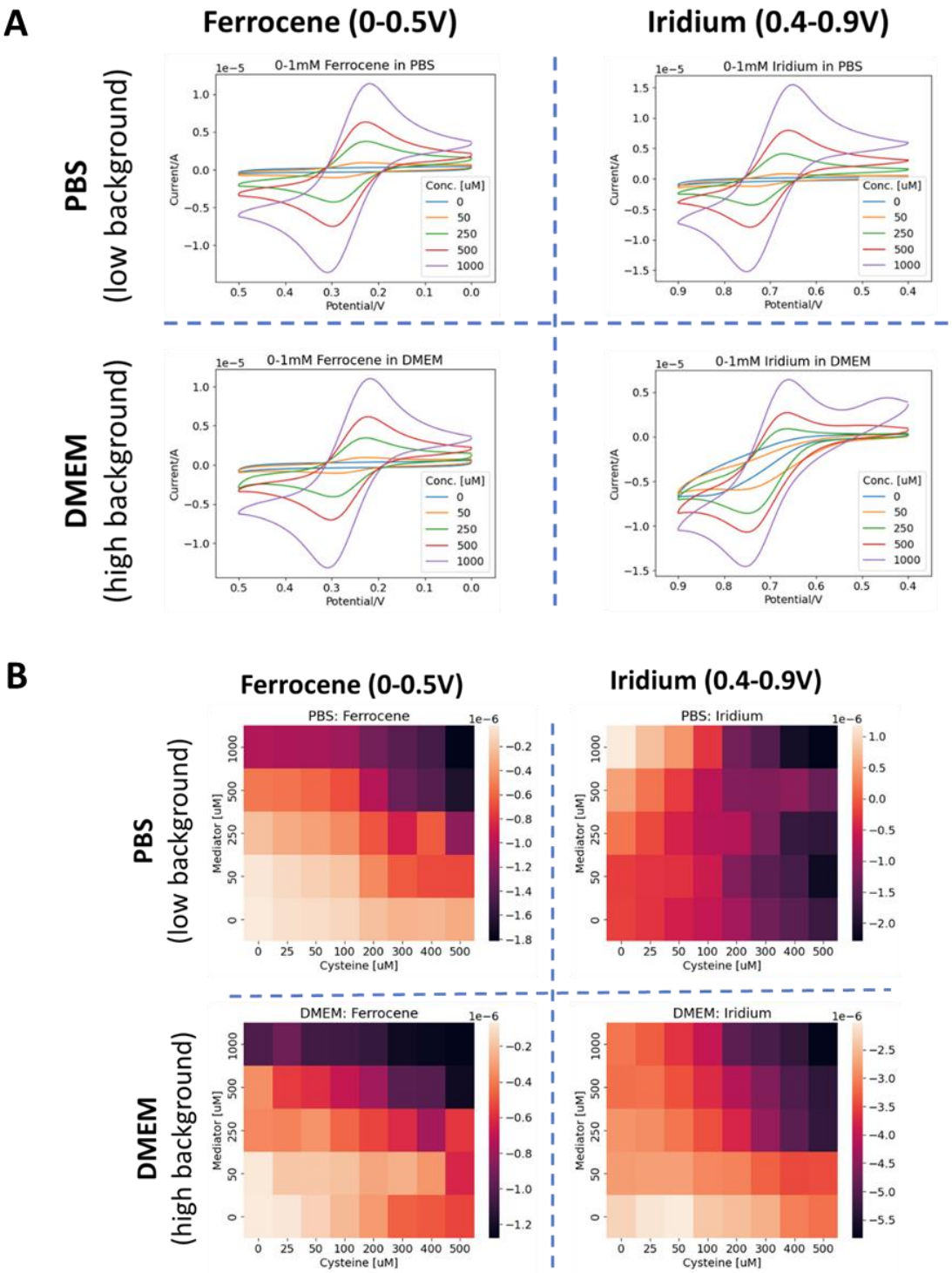


Figure 5.3 Mediator characterization.

(A) Cyclic voltammograms of mediator in background solutions. Top plots were measured in PBS, bottom plots were measured in DMEM. Left plots were probed with 0-1000 µM Ferrocene at 0V to 0.5V and right plots were probed with 0-1000 µM Iridium at 0.4V to 0.9V. **(B) Average Area-under-the-curve heatmaps.** Average area under the curve per condition was calculated (n=10) from single mediator probed cyclic voltammograms of cysteine in PBS and DMEM.

In **Fig 5.3B**, we plotted the average Area-under-the-curve (AUC) for each mediator in both PBS and DMEM to gain initial insight as to which mediator concentrations provide the best resolution to detect changes in current signal across cysteine concentrations. We observed that for Ferrocene, midrange mediator concentrations (50-500 µM) had the highest range in AUC values, indicating their potential for parsing cysteine concentrations. Whereas Iridium in PBS demonstrated similar AUC values across a wide range of mediator concentrations; this was also true in the case where there was no mediator added. We attribute this to the higher potential scanning range of the iridium, in which even without any mediator the cysteine can still get oxidized. We showed this in **Appendix Figure 8.7**, where we quantified free thiol concentration using Ellman's assay for samples with and without Iridium before and after measuring one cyclic voltammogram cycle. After taking a cyclic voltammogram in a cysteine sample without mediator, the free thiol concentration decreased as indicated by the stark reduction in absorbance. This was also the case for samples with 250 µM iridium, where adding iridium reduced absorbance even without taking a cyclic voltammogram measurement but after taking a measurement, the absorbance further decreased. Apparently, the iridium is a very effective cysteine oxidant. Interestingly, we noted further that there was a strong correlation between AUC and cysteine concentration even without any mediators over both voltage ranges and background solutions, **Appendix Figure 8.8**. This clearly demonstrates that even without mediators, cysteine residues can get oxidated by the electrode. It is well known, for example, that thiols can react with gold and foul gold surfaces, even if they are not used as electrodes [118]. We note, however, that the range of signals between the various cysteine solutions was higher when using mediators.

5.2.2 Data pipeline development

From our initial AUC analysis, we anticipated that mid-range ferrocene concentrations and all of the iridium concentrations would enable a model to predict cysteine concentrations with high correlation and accuracy. To find high performing models that could predict cysteine concentrations given a cyclic voltammogram measured using MEP as an initial input, we generated a pipeline (**Figure 5.4A**) to test various preprocessing methods and machine learning (ML) models across the datasets of various mediator concentrations collected. We tested three approaches to preprocessing the data; first electrochemical feature extraction in which we extracted features that are traditionally calculated when analyzing cyclic voltammograms (**Figure 5.4B**) [6, 25, 116, 119]. Second, we applied principal component analysis to the current data and extracted the top ten components as a method of dimensionality reduction. Lastly, we used both the electrochemical metrics and the top ten principal components as model inputs. After preprocessing with each of these approaches, we fed the extracted features into a grid search, where we tested multiple existing regression models across various hyperparameters and used 8-fold cross validation to calculate performance metrics for each dataset [120]. The list of models and hyperparameters tested in the grid search can be found in **Appendix Table 8.5**. Brief technical descriptions of the tested models can be found in **Methods**. The 8 folds consisted of 8 subsets of data, meaning the data retrieved from each well of the device was separated such that multiple cycles from the same well would not be in both the training and test sets. This was implemented to avoid data leakage; i.e., the model would not encounter data in the test set from the training set and yield unjustly high performing models. R-squared (R^2) values and negative mean absolute errors (MAE) were calculated as performance metrics for all cases.

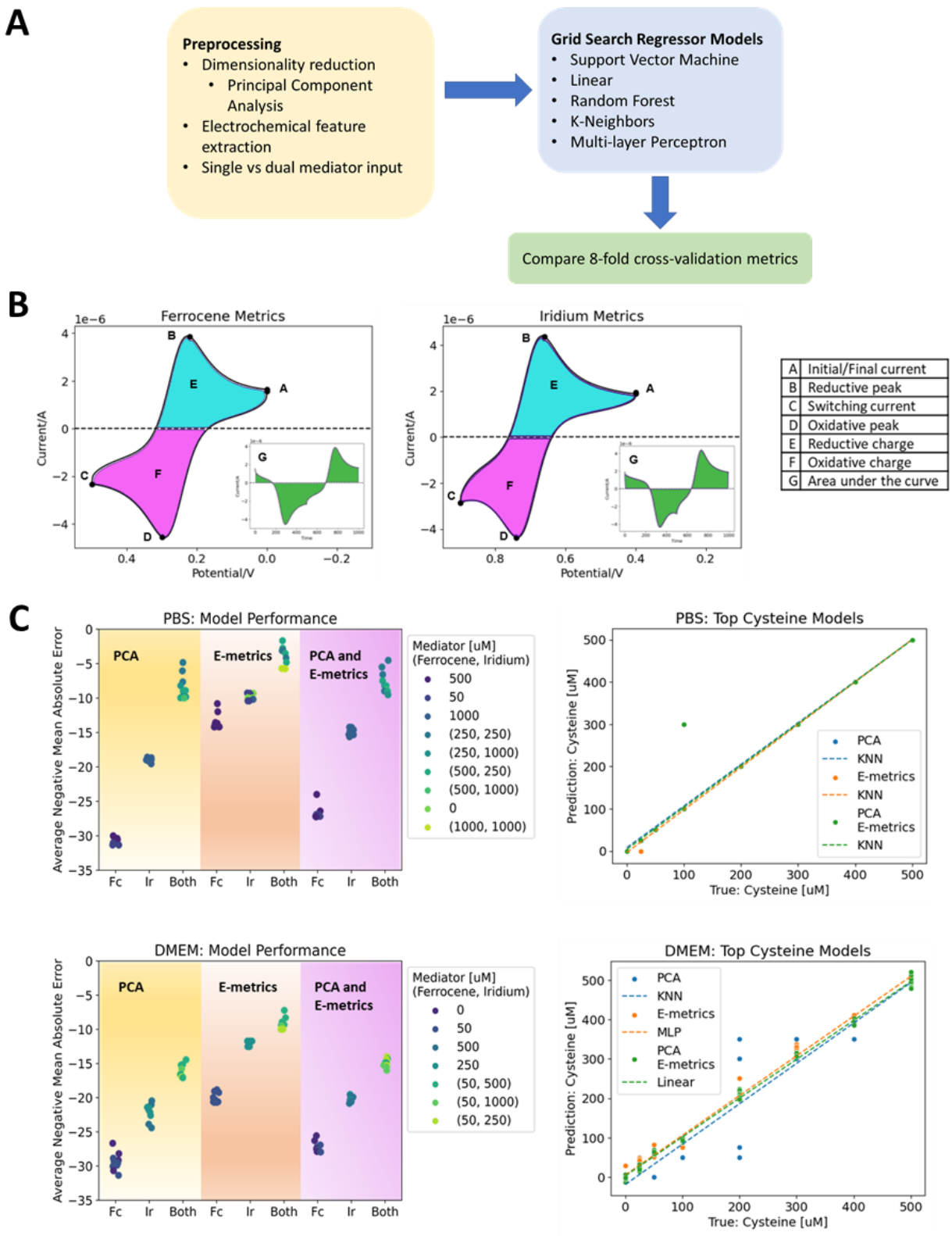


Figure 5.4 Data pipeline and L-cysteine grid search results.

(A) Data pipeline development scheme. (B) Electrochemical metrics. Diagram of electrochemical metrics from ferrocene (left) and iridium (right) which are used as training features. **(C) Cysteine model grid search results.** Average MAE values from the top ten models from each preprocessing and mediator selections were plotted (left). Cysteine predictions on dual mediator datasets from models with the smallest MAE per preprocessing method were plotted against their true values (right). Mediator concentrations for PBS and DMEM were: 250 µM Iridium and 500 µM Ferrocene, and 250 µM Iridium and 50 µM Ferrocene, respectively. Performance metrics from these prediction plots are found in **Table 5.1**.

In **Figure 5.4C**, we plotted the top ten models with the lowest cross validation average MAE values for each preprocessing method and mediator approach. We found that across all preprocessing approaches and both background solutions, datasets with features from both mediators had higher performing models as denoted by lower average cross validation MAE values. In terms of models trained on single mediator datasets, pipelines trained on Ferrocene probed datasets generally had larger MAE values than those probed with Iridium. With respect to the various preprocessing approaches, the top models for each ranked as follows from lowest to highest MAE: electrochemical metrics, PCA components and electrochemical metrics, and PCA components. This emphasizes how the use of features relevant to the data collection method may yield better model performance.

Overall, model performances for predicting cysteine concentrations in DMEM were lower than in PBS as indicated by their larger MAEs. As shown in **Figure 5.4C** and **Table 5.1,** PBS models could achieve MAE values below 5 µM whereas the top performing DMEM model had a MAE of 8.63 µM. This is not surprising given the previously demonstrated background in DMEM measurements due to its media components. Both grid searches yielded many models with R^2 values above 0.9, indicating that both mediators in either solution could predict cysteine concentrations with a strong correlation to their true values. The full grid search results for are available in **Appendix Table 8.6**. In both solutions, the highest performing models per preprocessing methods were trained dual mediator datasets. For PBS, the top ten models using

dual mediators for all preprocessing approaches were K-Neighbors regressors. We suspect that due to the lack of background in PBS, the input features naturally group together based on their cysteine concentrations. This type of input data is ideal for implementation of a K-neighbors approach, where predictions are based on the unknown sample's proximity to the known samples in the training set within the vector space. For DMEM, a pipeline using dual mediator electrochemical metrics and PCA components as inputs to a simple linear regressor performed the best with a MAE of 8.63 and R^2 of 0.9967. Whereas DMEM pipelines with just electrochemical metrics had top performances with MLP regressors, and pipeline with only PCA component inputs worked well with K-neighbor regressors. This may speak to the complexity of electrochemical metrics, as some of these features take in to account current data as voltage dependent variables. Comparatively, PCA preprocessing is applied to raw current values, in which each individual current value is considered an independent variable.

From these results, we moved forward with 250 µM Iridium and 500 µM Ferrocene for probing samples in PBS and 250 µM Iridium and 50 µM Ferrocene for probing samples in DMEM. These concentrations were high performing in dual mediator datasets but were not the highest concentrations for the sake of reducing the need for mediator reagent. In **Figure 5.4C,** we plotted the prediction results using these mediator concentrations and their respective top model per preprocessing methods for one fold in our cross validation, such that we trained the models on data from wells 2-8 and presented the model predictions for the data from well 1. The performance metrics for these plots can be found in **Table 5.1**. These plots illustrate the performance variability between the various pipelines and background solutions, notably DMEM has an overall larger MAE across all pipelines.

Solution	Preprocessing	Model	Hyperparameters	R^2	Slope	MAE
PBS	PCA	KNN	n_neighbors=5	0.9682	0.9841	5.0
PBS	E-metrics	KNN	n_neighbors=5	0.9989	1.0088	1.56
PBS	PCA+E-metrics	KNN	n_neighbors=5	0.9838	0.9920	2.5
DMEM	PCA	KNN	n_neighbors=2	0.9313	1.0257	28.59
DMEM	E-metrics	MLP	activation=logistic, alpha=0.01, hidden_layer_sizes=(100,) max_iter=10000, solver=sgd	0.9933	1.008	12.46
DMEM	PCA+E-metrics	Linear	N/A	0.9967	0.9862	8.63

Table 5.1 Cysteine grid search top pipeline parameters and model performance metrics.

We then tested dual mediated probing pipelines with their top models and preprocessing methods on a test set which contained measurements of cysteine concentrations not present in the training set in both PBS and DMEM (**Figure 5.5A**). To do so, we trained the highest-ranking pipelines (same as **Figure 5.4C**) on the full datasets from our grid search with their solution specific mediator concentrations. We then used the trained pipelines to predict the cysteine concentrations in our test set. The prediction results from these pipelines are shown in **Figure 5.5B**. The predictions of cysteine in PBS had high correlation to the true cysteine concentrations with R-squared values above 0.88 for all preprocessing approaches (**Table 5.2**). The best pipeline consisted of electrochemical metrics as inputs to a K-Neighbors regressor, which was also the top performing pipeline from our grid search. The DMEM test set predictions were not as accurate as the PBS predictions, where using the same top pipelines from the DMEM grid search yielded MAE values of around 100 µM. Therefore, we explored other models from our grid search and found

that using PCA components and electrochemical metrics together as inputs to a linear regression model yielded higher R^2 values than other top models previously tested, however the MAE was still 122.56 µM. Whereas when testing other models from our PBS grid search, we found that using a MLP regressor on electrochemical metrics yielded the lowest MAE (27.76 µM) of the pipelines tested. This demonstrates that the highest-ranking from our grid search results may not necessarily yield the best results on our test set. This is common as models can hone in on features and trends in the training dataset that are not always significant or present in new data [121].

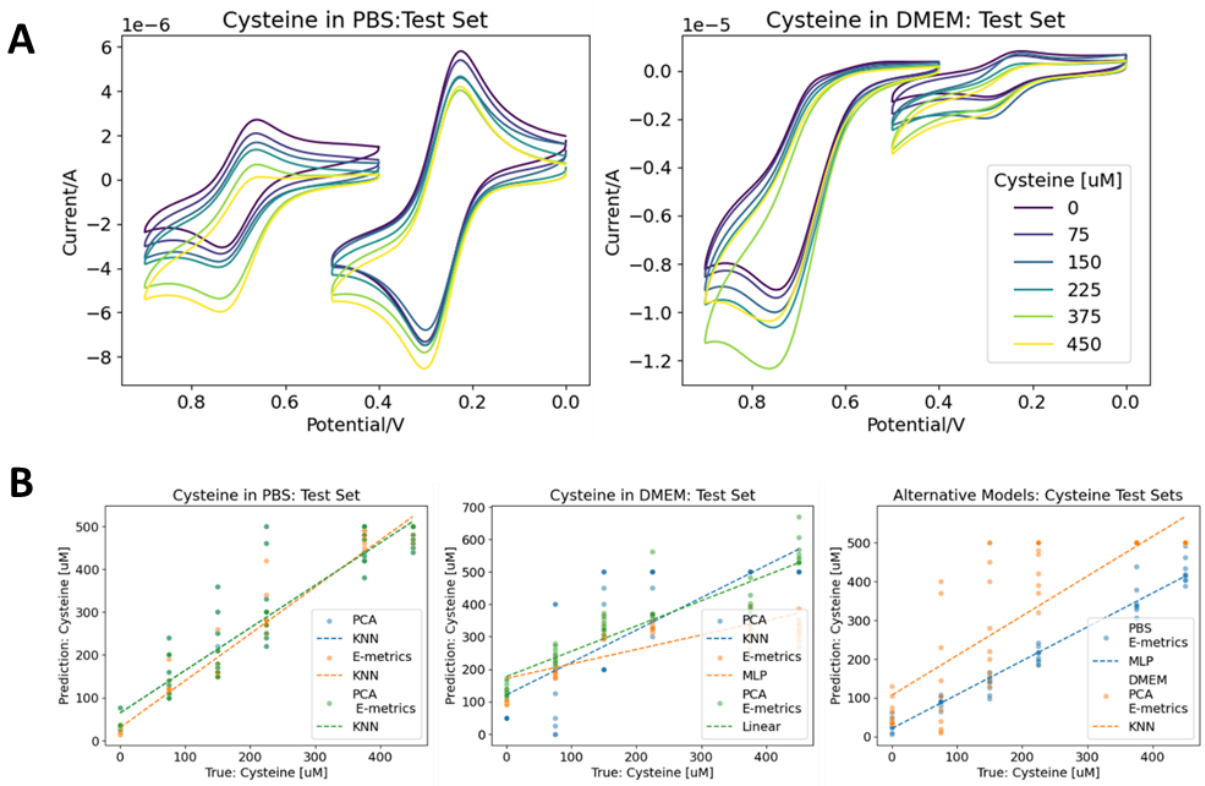


Figure 5.5 L-cysteine test set data.

(A) Cysteine test set data. Cyclic voltammograms of the test set, containing samples of 0, 75, 150, 225, 375, and 450 µM cysteine in PBS (left) and DMEM (right). PBS measurements were probed with 500µM Ferrocene and 250µM Iridium, DMEM measurements were probed with 50µM Ferrocene and 250µM Iridium. (**B) Cysteine model test set performance.** Prediction plots from top cysteine quantification models on test set data (left and center). Prediction plots from alternative high performing models (right). Performance metrics from these predictions can be found in **Table 5.2**.

Solution	Preprocessing	Model	Hyperparameters	R^2	Slope	MAE
PBS	PCA	KNN	n_neighbors=10	0.8839	0.9924	63.45
PBS	E-metrics	KNN	n_neighbors=10	0.9602	1.0939	50.21
PBS	PCA+E-metrics	KNN	n_neighbors=10	0.8840	0.9928	63.29
DMEM	PCA	KNN	n_neighbors=2	0.6822	0.9965	127.92
DMEM	E-metrics	MLP	activation=logistic, alpha=0.01, hidden_layer_sizes=(100,) max_iter=10000, solver=sgd	0.6141	0.4460	109.65
DMEM	PCA+E-metrics	Linear	N/A	0.8223	0.7785	132.56
PBS	E-metrics	MLP	activation=tanh, alpha=0.01, hidden_layer_sizes=(100,) max_iter=10000, solver=sgd	0.9677	0.8733	27.76
DMEM	PCA+E-metrics	KNN	n_neighbors=10	0.7065	1.024	119.88

Table 5.2 L-cysteine test set pipelines and model performance metrics.

5.2.3 Method transfer to IgG from human serum

After demonstrating the feasibility of developing a pipeline to quantify L-cysteine, we next tested whether we could transfer this approach to quantifying reduced monoclonal antibody. All antibodies possess interchain disulfide bonds that are crucial to their structure. Specifically, these interchain bonds dictate whether antibodies are in their correct formatting as depicted in **Figure 5.6A**. When these bonds are reduced fragmentation and subsequent reoxidation can occur into incorrect formats [103, 111, 122]. Previous work has demonstrated that MEP can be used to probe the thiol groups on reduced human IgG despite their increased complexity due to the cysteine residues' embedded nature compared to free L-cysteine in solution [116]. Thus, we applied the top performing electrochemical parameters from our L-cysteine pipelines (500µM Ferrocene and 250µM Iridium for PBS; 50µM Ferrocene and 250µM Iridium for DMEM) to collect datasets of reduced and intact IgG from human serum. To prepare these samples, we intentionally chemically reduced stock IgG to generate free thiol groups to probe with our method. These datasets were used to develop data pipelines that can quantify the concentration of reduced IgG from cyclic voltammetry data. The sample space of each dataset consisted of samples with only reduced or intact antibody from 0-3 g/L and samples at a total of 3 g/L at various ratios of reduced to intact antibody (**Figure 5.6A**). We then applied the same methodology where we tested three various preprocessing approaches (PCA, electrochemical metrics, and both) and used a grid search to test various regression models on these datasets (**Figure 5.4A**). The results from these grid searches are also in **Appendix Table 8.6.**

From the top ten grid search results we saw that in PBS, the electrochemical metrics preprocessing approach had the smallest MAE values. We anticipated this result owing to the relatively clean background. Additionally, models trained on just Ferrocene probed data had largest error across all preprocessing approaches compared to models trained on Iridium and dual

probed datasets. This too was anticipated based on cysteine results above in that Ferrocene is weakly oxidizing and can participate in reduction and oxidation reactions. Whereas Iridium is a strong oxidant and likely participates in a more uniform manner via primarily oxidation reactions. Comparatively in DMEM, models trained on Iridium datasets preprocessed with PCA had higher MAE values than Ferrocene. Interestingly, Iridium datasets preprocessed with electrochemical metric extraction had lower MAE values than Ferrocene and dual mediator datasets. This suggests that when probing IgG in DMEM with Iridium, the media background and increased sample complexity may reduce the effectiveness of PCA extracted components. We also noted that unlike previous L-cysteine grid search results, DMEM models trained on just Ferrocene probed data had lower MAE values than those trained on just Iridium probed data for PCA and PCA plus electrochemical metrics preprocessing methods. This also speaks to the added complexity of samples in DMEM, which may cause altered signal of iridium through the oxidation of other residues in solution. Overall, the top performing models in from our grid search were K-Neighbors and Random Forest for all of the preprocessing methods tested (**Figure 5.6B**). Both models demonstrated strong correlations between true and predicted values with R-squared values above 0.9, as well as low MAE metrics (≤0.1 g/L).

We then prepared human IgG samples at different ratios of intact to reduced at a total concentration of 2 g/L per sample and measured them electrochemically (**Figure 5.6C**). We trained models on the full IgG dataset from the grid search and then used them to predict the reduced concentration of IgG per sample. In PBS, we found that the top performing K-Neighbors models from the grid search yielded poor prediction results with R-squared values below 0.6 and high MAE values. However, we found that the Random Forest model on dual mediator datasets were able to increase the correlation to R-squared values around 0.7, although the MAE values were

quite high around 0.5-0.7 g/L. We noted that the performance of these antibody models were generally lower than the L-cysteine models, as the R^2 values were lower and MAE values were larger relative to the predictions. This may be expected as the samples are inherently more complex as the additional residues on the IgG protein can contribute to the increased noise in the sample and due to the additional structure of the protein the accessibility of cysteine residues are not as readily available as pure L-cysteine in solution [123, 124]. While this is the case, the reduced performance on the antibody test set could also be a sign that the sample space is too sparse and increased performance may just be a matter of increasing the diversity of the training set's sample space or increasing the amount of model training data [121].

Solution	Preprocessing	Mediator	Model	Hyperparameters	R2	Slope	MAE
PBS	PCA	Both	KNN	n_neighbors=1	0.9097	0.9235	0.1107
PBS	E-metrics	Iridium	KNN	n_neighbors=1	0.9967	0.9863	0.0161
PBS	PCA+E-metrics	Iridium	KNN	n_neighbors=1	0.9843	0.9660	0.0304
DMEM	PCA	Both	KNN	n_neighbors=10	0.9658	0.9879	0.0554
DMEM	E-metrics	Iridium	RF	max_features=25, n_estimators=1000	0.9945	0.9952	0.0534
DMEM	PCA+E-metrics	Ferrocene	RF	max_features=10, n_estimators=500	0.9780	0.9950	0.0761

Table 5.3 Antibody grid search top pipeline parameters and model performance metrics.

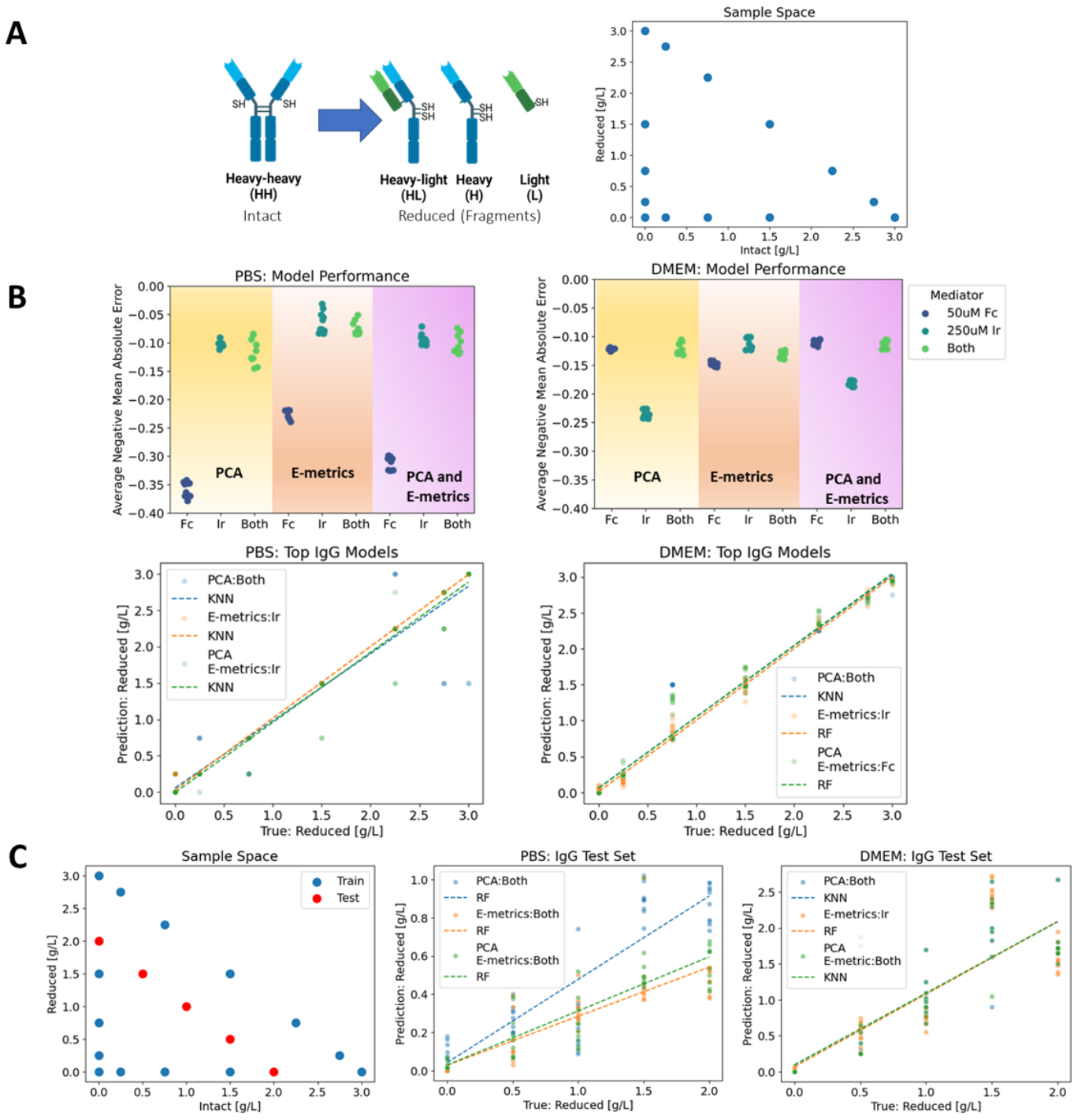


Figure 5.6 Antibody grid search and test set results.

(A) Diagram of antibody and IgG sample space. Diagram created with BioRender. **(B) IgG model grid search results.** Average MAE values from the top ten models from each preprocessing and mediator selections were plotted (top). Reduced IgG concentration prediction plots of models with the smallest MAE per preprocessing method were plotted against their true values (bottom). Performance metrics from these prediction plots are found in **Table 5.3**. **(C) IgG model test set sample space and performance.** Test set

samples in the sample space (left). Prediction plots from top IgG quantification models on test set data (center and right). Performance metrics from these predictions can be found in **Table 5.4**.

Solution	Preprocessing	Mediator	Model	Hyperparameters	R2	Slope	MAE
PBS	PCA	Both	RF	max_features=5, n_estimators=1000	0.7605	0.4194	0.5679
PBS	E-metrics	Both	RF	max_features=5, n_estimators=1000	0.7141	0.2566	0.7181
PBS	PCA+E-metrics	Both	RF	max_features=5, n_estimators=1000	0.7209	0.2831	0.7003
DMEM	PCA	Both	KNN	n_neighbors=10	0.7156	0.9955	0.317
DMEM	E-metrics	Iridium	RF	max_features=25, n_estimators=1000	0.6744	1.010	0.3610
DMEM	PCA+E-metrics	Both	KNN	n_neighbors=10	0.7131	0.997	0.3175

Table 5.4 Antibody test set pipeline parameters and model performance metrics.

5.2.4 Prospective method transfer to cell culture supernatant

We additionally measured samples from CHO cell shake flask culture over the course of their 11 day growth. These cultures were seeded at approximately 500,000 cells per mL in CHO cell specific media (AMBIC 1.1 basal media) and were fed every odd day with its suggested paired supplement (AMBIC Feed A and B). We measured the cell free supernatant from these samples using the same mediator concentrations optimized for DMEM cell culture media and found the signal at Iridium scan ranges (0.4-0.9V) were much 2-6 folder higher than measurements in DMEM (**Figure 5.7A**). We suspect this is indicative of the media's increased complexity, such that additional components resulted in more oxidizable residues. We also observe that at 250 µM of Iridium, the measurements appear to have lost their distinguishable redox peaks. In DMEM, we also observed the attenuation of the reductive Iridium peak at these but were still able to see its oxidative peak at approximately 0.75 V. We note, that DMEM does not contain L-cysteine (Corning 17-205-CV). Ferrocene mediated measurements appeared to be within the same range as measurements in DMEM, approximately +/- 1 µA.

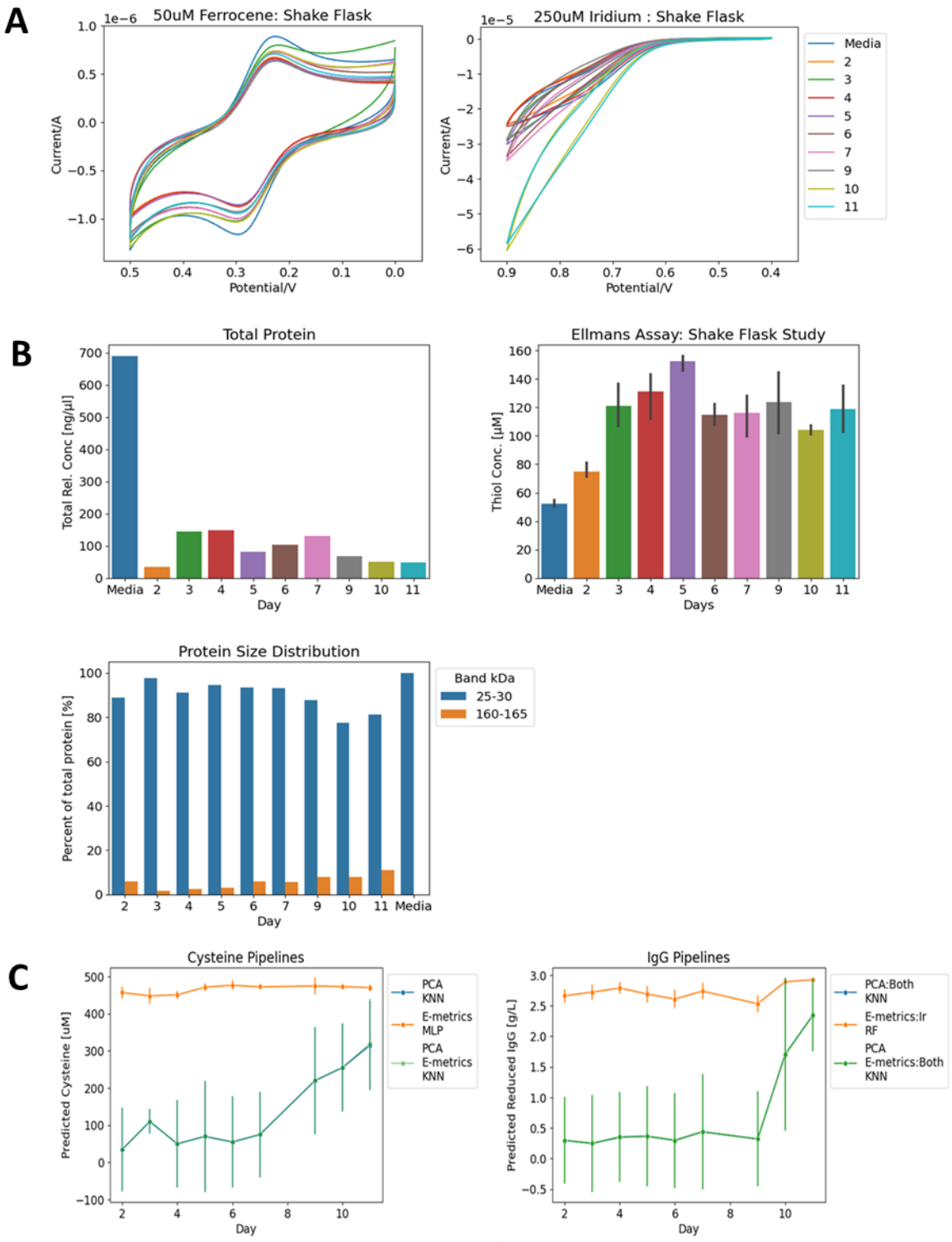


Figure 5.7 Shake flask culture MEP.

(A) Shake flask cell culture cyclic voltammograms. Plots of cell free supernatant from shake flask cultures, probed with 50 µM Ferrocene (left) and 250 µM Iridium (right). **(B) Shake flask CE-SDS and Ellman's assay results.** Total relative protein concentrations of shake flask culture supernatant from CE-

SDS were plotted across time in days (upper left). The 25-30 kDa (bottom left) and 160-165 kDa (bottom right) proteins were plotted as percent of the total protein over time. The total thiol concentration quantified by Ellman's Assay was plotted over time (upper right) (**C) DMEM model predictions on shake flask samples.** Cysteine (left) and IgG (right) model predictions on shake flask probed data using DMEM optimized mediator concentrations; 50 µM Ferrocene and 250 µM Iridium. Error bars represent standard deviation of voltammogram cycles, n=10.

We then characterized the amount of free thiol and antibody fragments in the cell culture supernatant to establish the true concentrations. To measure free thiol, we used Ellman's assay and found that the cysteine concentration increased across the first 4 days, then decreased after day 5 (**Figure 5.7B**). This trend correlates to previous reports in which antibody production greatly increases after viable cell density reaches its peak [125, 126]. This suggests that cysteine consumption may increase once cells enter their protein production phase. To quantify protein in the supernatant, we ran microchip capillary electrophoresis and found that the total protein fluctuated over time, with it decreasing at the culture's endpoint. It is notable that the basal media had approximately 7-fold more total protein than the cell culture supernatants, indicating some aggregates or proteins in the media components that are consumed by the cells during culture. This is also indicated by a large band at 20-30kDA from the basal media sample. Before day 5, most protein can be attributed to these media components as this 20-30kDA band was above 90% of total protein. After day 5, we observe an increased percent of protein at 160-165kdA which correlates to intact IgG. Although we note, this only contributed to about 10% of the total protein at its highest.

We then applied the models which performed the best on test sets of L-cysteine and reduced antibody in DMEM to the cell free supernatant measurements. Prediction results showed pipelines using PCA components predicted increasing cysteine and reduced antibody concentrations after day 8 (**Figure 5.7C**). This suggests that the clear increase in amplitude for day 10 and 11 on the raw iridium CVs, likely contribute to these predictions (**Figure 5.7A**). The pipelines which used

solely electrochemical metrics did not predict variable amounts of cysteine and reduced antibody over the course of the cell culture, instead predicted concentrations close to the maximum concentrations the model was trained on. These results suggest the Iridium's increased amplitude in this cell culture media causes the models previously trained on DMEM to just approximate the signals to the highest concentrations it has encounter. The inaccuracy of these predictions indicated that the pipelines trained on the datasets collected on DMEM samples are not transferrable to samples of different basal component composition. We suggest for further application of these pipelines, additional collection of datasets in cell line specific media would be necessary or a library of cell culture samples at various conditions should be used. Given the high production volume and analytics performed at industrial bioprocessing sites, gathering these datasets would be quite feasible; enabling rapid pipeline development that could be transferred to earlier process development workflows or used for production scale monitoring.

5.3 Conclusions

In this chapter, we have demonstrated a workflow for generating rapid detection methods for cysteine and reduced antibody concentrations. The development of these methods aimed to provide simple, inexpensive, and rapid approaches to monitor cell culture and product quality during bioprocessing. To establish our workflow, we first screened and optimized data pipelines for quantifying L-cysteine in low and high background solutions. This data set is made available in our supplemental and would be appropriate for others as a means to do an initial screen. Datasets generated for this screening were captured efficiently through the use of an 3D printed device that increased measurement throughput. Additionally, the rapid nature of electrochemical measurements enabled quick turnaround times for data collection. This yielded electrochemical methods and data pipelines that could predict L-cysteine concentrations in low and high backgrounds with high correlations to their true concentrations.

We then transferred the highest performing pipelines from our L-cysteine screen to human IgG to quantify the concentrations of reduced antibody. We collected datasets with L-cysteine optimized probing parameters and used them to test various data pipelines. Here, trained models on IgG in PBS and DMEM were able to predict reduced antibody concentrations with a positive correlation between the predicted and true concentrations (R^2>0.7). However, we suggest that increasing the diversity and quantity of data in these training sets could increase model performance. This may take the form of collecting additional samples measurements with more ratios of intact to reduced IgG to enhance the distribution of conditions that the model is trained on. Measuring more samples of conditions that the dataset currently obtains may also increase performance by helping expose the model to more noise introduced by experimental variance.[121]

Lastly, we tested L-cysteine and reduced antibody prediction pipelines on antibody producing CHO cell culture supernatant. We found that DMEM probing parameters may not be ideal for this CHO specific cell culture media due to its increased complexity yielding additional oxidizable components. Subsequently, the pipelines trained on DMEM datasets did not perform well for quantifying mAb levels on the cell culture samples. However, we suggest that future dataset acquisition and pipeline development demonstrated in this chapter can be easily transferrable to other cell culture medias or real cell culture samples. The rapid analysis and ease of measurements enabled by electrochemical measurements and efficient low sample volume sampling setups make technology transfer simple and feasible. This is further supported by the existing throughput and analytics already being implemented by industrial manufacturers which could be used to develop these pipelines.

5.4 Methods

5.4.1 Materials

1, 1'-Ferrocene dimethanol (Santa Cruz sc-222778) was dissolved in DMSO at 100 mM and stored at 20C until use, potassium hexachloroiridate (III) hydrate (Thermo Scientific 011887.03) was dissolved in water at 10 mM and stored at -20°C until use. Phosphate buffered saline prepared from tablets (Sigma Aldrich P4417) and Dulbecco's Modified Eagles Medium with 4.5 g/L glucose and sodium pyruvate and without phenol red and L-glutamine (Corning 17-205-CV) were used as sample diluents.

5.4.2 Electrochemical measurements

Cyclic voltammograms were captured using a CHI1040c Multichannel Potentiostat with a scan rate of 0.1 V/s, sensitivity of 1e-5 A/V, and a positive scan polarity. Electrochemical measurements taken with a glassy carbon electrode working electrodes were acquired using a 3D printed 8-well device. Samples were measured by placing 100 uL per well and placing the working electrode into each well. Measurements were taken as one cell as all samples were connected to the same reference and counter electrode.

5.4.3 Device fabrication

The 3D printed 8-well device for small volume samples was designed based on a standard 3-electrode system including a device with 8 satellite sample wells for standard working electrodes, and a central salt bridge well for an Ag/AgCl reference electrode and a Pt wire serving as the counter electrode (**Appendix Figure 8.10**). Specifically, to make the 3D printed device housing, designs were created using Autodesk® Fusion 360 and Autodesk® Inventor (Autodesk, Inc, US). The CAD files were then converted to writing-path using the Chitubox 1.9.4 (Guangdong, China). Mars 3 Pro was used to fabricate the housing. When the printing process was completed, the substrate was placed in a bath of methanol with gentle shake for 1 minute and then dried with

nitrogen. This step was repeated three times to thoroughly remove any uncured residues. The prints were then further cured under a 20-Watt UV lamp for 10 seconds followed by a 30-second rest at room temperature. This step was repeated three times to achieve full strength.

To prepare the device for use, a 1% agarose prepared in 1M KCl solution is first heated, and 200 µL of a liquid agarose solution is injected into the designed grove located in central well to cast the agarose salt bridge noted above. After solidification, the central well was filled with 4 mL of salt bridge solution (1M KCl). An Ag/AgCl reference electrode and a Pt counter electrode were then inserted into the central well, completing the standard 3-electrode system. In all cases, 100 µL of test samples are then loaded into the satellite sample wells followed by insertion of the working electrodes for further interrogation.

5.4.4 Reduction of human IgG

Lyophilized IgG from human serum (Sigma Aldrich I4506) was dissolved in PBS and was reduced in 10 mM TCEP at 4°C overnight. Reduced IgG was then buffer exchanged in either PBS or DMEM to remove TCEP using desalting spin columns (Thermofisher Zeba Spin 7K Cutoff Desalting Columns).

5.4.5 Free thiol quantification

Ellman's reagent, 5, 5'-dithio-bis-(2-nitrobenzoic acid) or DTNB, was used to quantify free thiol concentration of samples against a L-cysteine standard curve.[127] Briefly, 2.5 µL of 4 mg/mL of DTNB dissolved in reaction buffer (0.1M sodium phosphate pH 8.0 containing 1mM EDTA) was added to each sample well in a 96-well plate. Next, 125 µL of reaction buffer was added to each well along with 12.5 µL of either L-cysteine hydrochloride monohydrate standards or unknowns. All wells were mixed and incubated at room temperature for 15 minutes and then their absorbances were measured at 412 nm using a TECAN Spark Plate Reader.

5.4.6 Capillary electrophoresis

Microchip capillary electrophoresis was performed using a 2100 Bioanalyzer (Agilent). Samples were processed according to Agilent's Protein 230 kit protocols under nonreducing conditions.

5.4.7 Dataset acquisition

Datasets were prepared by measuring 11 cycles of cyclic voltammograms per condition with all 8 channels on an CHI1040c 8 channel potentiostat. The first cycle is removed from the dataset, as it served as an initial conditioning cycle, yielding 80 cyclic voltammograms per condition.

5.4.8 Data preprocessing and model development

The following electrochemical metrics were calculated per each cyclic voltammogram: Total Area Under the Curve (AUC), Reductive Charge (Q_red), Oxidative Charge (Q_ox), Redox charge ratio (Q_red/Q_ox), Max peak current (Ip_max), Minimum peak current (Ip_min, Initial Current (i_initial), Switching Current (i_switch), and Final Current (i_final). In addition to these metrics, the top ten components from principal component analysis performed on the training set were computed and utilized as input features. Data was scaled to its mean using standard scaling prior to principal component analysis. Lastly, the cycle numbers of the cyclic voltammogram were included as an input feature for model training.

Model development was performed using python's Scikit-Learn library [120]. Data cleaning, analysis, and visualization done using the following python libraries: NumPy [128], Pandas [129], SciPy [130], and Seaborn [90].

5.4.8.1 Model descriptions

Below are brief technical descriptions of the models tested in this chapter.

5.4.8.1.1 Linear Regression

Linear regression uses inputs as independent variables and computes regression coefficients to best linearly fit input variable to their corresponding dependent variables [131].

5.4.8.1.2 Support Vector Machine

Support vector machines (SVMs) are models that utilizes data points as vectors and identifies hyperplanes that are used to separate the samples [132, 133]. The kernel hyperparameter refers to different kernel functions that provide a basis for hyperplane The hyperparameter C is a regularization parameter that modifies the margins of the decision boundaries, in other words a C can toggle the model's sensitivity to outliers.

5.4.8.1.3 Decision Tree

Decision Tree (DT) models contain tree like structures where each node is an input feature and each leaf is a class [134, 135]. The hyperparameter max_features hyperparameter determines the max number of splits each node can undergo at each level of the tree.

5.4.8.1.4 Random Forest

Random Forest (RF) models consist of ensembles of decision trees, in which subsets of data will randomly be used to generate decision trees to form a random forest [134-136]. Like decision trees, the max_features hyperparameter is used to determine the max number of splits per node. The n_estimators hyperparameter is used to determine the number of trees in the forest.

5.4.8.1.5 K-Neighbors

K-nearest-neighbors (KNN) models predict class values by sampling the input's k-nearest neighboring data points [137]. The hyperparameter n_neighbors determines the number of neighboring data points the model uses to predict the class of the input sample.

5.4.8.1.6 Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) models consist of multiple layer of connected nodes with the following architecture: 1) Input layer: A layer that consists of one neuron per input feature, 2) Hidden layer: One or more layers whose weights are updated during training to optimize error, 3) Output layer: Final predictions are generated using the output layer [138]. The solver hyperparameter determines the solver used for weight optimization, activation determines the activation function

for the hidden layers, max_iter is the max number of iterations the solver undergoes if it does not converge, hidden_layer_sizes determines the number of hidden layers and the number of nodes in each layer, alpha is a regularization term.

5.4.9 Cell culture

A standardized IgG_1-producing recombinant CHO-GS (-/-) (glutamine synthetase) cell line referred to as NISTCHO [139] was cultured in AMBIC 1.1 basal media (MilliporeSigma, Burlington, MA) in 250 mL shake flasks with 70 mL media. Pre-cultures were passaged every three days and three times out from thaw prior to inoculating the experiments. Cultures were seeded at 0.5 x 10^6 cells/mL, incubated at 37°C with 5% CO_2, and shaken at 130 rpm. AMBIC 1.1 Feed A and Feed B (MilliporeSigma, Burlington, MA) were fed at 10% and 1% of the working volume every other day beginning on Day 3, respectively.

Chapter 6 Electrochemical identification of bacterial vaginosis associated microbes

This chapter is adapted from an unpublished manuscript.

6.1 Introduction

The vaginal microbiome is of great importance to female reproductive health and fertility. Previous studies have found that the vaginal microbiome is typically dominated by one *Lactobacillus* species, or is otherwise presenting as a polymicrobial state of dysbiosis [140-142]. These are referred to as community state types (CST), in which CSTI, CSTII, CSTIII, and CSTV are dominated by *L. crispatus, L. gasseri, L. iners, and L. jensenii*, respectively . Whereas CSTIV consists of heterogenous biomes that typically include higher levels of strictly anaerobic bacteria, indicative of dysbiosis. CSTIV commonly presents itself as the disease bacterial vaginosis (BV), which is defined by a shift in vaginal microbiota from a beneficial dominance of *Lactobacilli* to a diverse set of anaerobic bacteria, often forming polymicrobial biofilms, and an increase in vaginal pH to above the normal range of pH 3.5-4.5.[143, 144] BV is characterized by itching, burning, and malodor and is linked to increased risk of gynecological complications such as preterm labor [145], urinary tract infection and acquisition and transmission of sexually transmitted infections [146, 147]. In addition to CSTIV, CSTIII which describes a dominance of *L. iners* has been observed in clinical temporal analyses to often be a precursor to shifting into CSTIV [35, 148]. *L. iners* has also been documented to be present in BV positive patients, although whether the species is causal to the disease is undetermined due to lack of evidence [35, 149, 150]. Comparatively, other *Lactobacillus* dominated CSTs are generally considered healthy [148, 149]. These findings have led to increased interest in determination of CST in the clinical setting beyond traditional methods of BV diagnosis.

The gold standard of diagnosis of BV is the Nugent scoring method in which patient swabs are Gram stained and scored by a trained technician according to a rubric that quantifies the presence of various microbial species morphologies [36, 37]. However, this methodology does not provide the resolution to a determine patient's CST. Nugent scoring also requires outsourcing samples to a clinical laboratory, such that results are not available at the time of a patient visit. Additionally, methodologies utilizing molecular diagnostics have been developed in recent decades to increase the diagnostic accuracy and speed. However, these assays are often expensive and labor intensive resulting in limited clinical translation. This often leads clinicians to preemptively prescribing antibiotics at the time of care, based on patient symptoms. The over use of antibiotics has been documented to be a key accelerator of antibiotic resistance, especially in low resource settings where inadequate diagnostic facilities lead to over prescription of antibiotics [151]. In addition to symptomatic patients, monitoring the vaginal microbiome would also be beneficial to general female reproductive health and fertility. Such that, symptoms may not always align with microbiome composition although certain CSTs may indicate a patient is at increased risk for disease or fertility and labor complications [30, 148, 152]. Generating a quick point-of-care assay for determining the CST of the vaginal microbiome would allow for preventative care as well as help provide better informed treatments.

In this work, we implemented mediated electrochemical probing [6, 116, 117, 119] techniques and machine learning analysis [e.g., from Chapter 4] to develop a simple and rapid methodology to determine the majority species in vaginal associated microbe samples as shown in **Figure 6.1**. Mediated electrochemical probing (MEP) is an indirect technique that uses redox-active chemical species as electron shuttles that can exchange electrochemical information from a biologic sample in the local environment to an electrode surface.[24, 117] Here we transferred MEP approaches to a

disposable and inexpensive platform by modifying commercially available screen-printed electrodes (SPEs) with mediators. We utilized two mediators traditionally used in colorimetric detection: 1) 2,6-dichlorophenolindophenol (DCPIP), which is used for the colorimetric quantification of ascorbic acid, and 2) Ninhydrin which is used for the detection of amines. Previously studies have shown significant differences in metabolomic profiles between CSTs [153], such that we were able to use these two mediators to probe for these differences in vaginal microbe cultures as they pertain to pH and amine concentration. We grew monocultures of vaginal bacteria strains in the same growth media and collected their supernatants to characterize with these modified SPEs. We then generated datasets from the supernatant samples to train a machine learning model that would discriminate between species in test cultures. We then transferred this methodology (e.g., training sets based on individual monoculture growth profiles) to test patient mucus samples, in order to determine if a model trained on *in vitro* bacterial samples could accurately classify clinical samples.

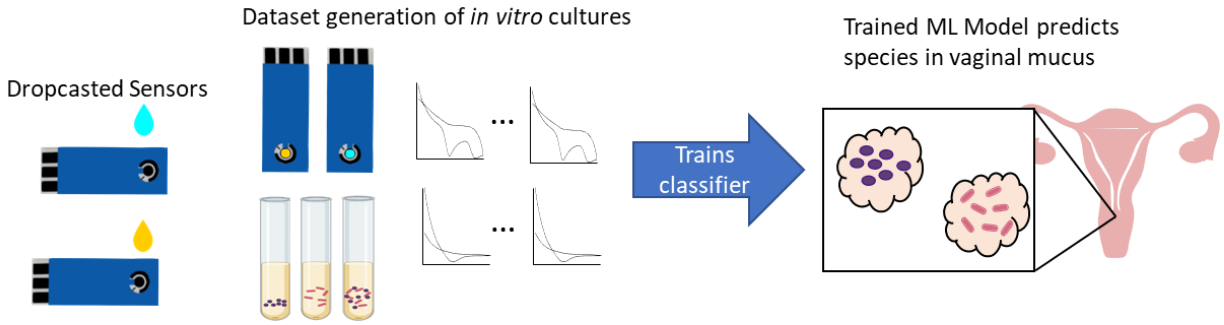


Figure 6.1 Vaginal microbe detection overview. Created with BioRender.

6.2 Results and Discussion

6.2.1 Sensor characterization

We first performed colorimetric assays for DCPIP and Ninhydrin using vaginal microbiome relevant parameters. In **Figure 6.2A,** the mechanism for each colorimetric assay is shown. For both mediators, reduction reactions occur in the presence of their respective analytes and yield color changes. Specifically, DCPIP gets initially reduced at is center amine group which yields a pink color and then again on its terminal alcohol group to convert to colorless indicating the presence of ascorbic acid (Vitamin C) as a reducing agent [154]. Ninhydrin, initially a colorless solution, is reduced by alpha amine groups and subsequently complexes with a non-reduced ninhydrin and free amine to form Ruehlmann's Purple [155-157]. It may also form a yellow color when complexed with proline [158]. We show in **Figure 6.2B,** DCPIP incubated with buffer at a vaginal microbiome relevant pH range of pH 3-5, using citrate buffer to accommodate for a suitable buffering capacity for this range. Visually, it is observed that lower pH samples yield more pinkish color compared to higher pHs. To quantitatively assess the samples, we measured absorbance at 600 nm for 30 minutes at room temperature. Over time the absorbance values decrease more for samples at lower pH, indicating the continued reduction of DCPIPH to $DCPIPH_2$ over time. For Ninhydrin, we tested its colorimetric performance using 0-100 mM of cadaverine, a diamine that was found to be significantly higher in BV positive patients [153]. For the initial 5 minutes at room temperature, we observed the absorbance at 570 nm positively correlated with cadaverine concentration. However, from 10 minutes onwards the absorbance of 10 mM samples exceeded that of the 100 mM samples.

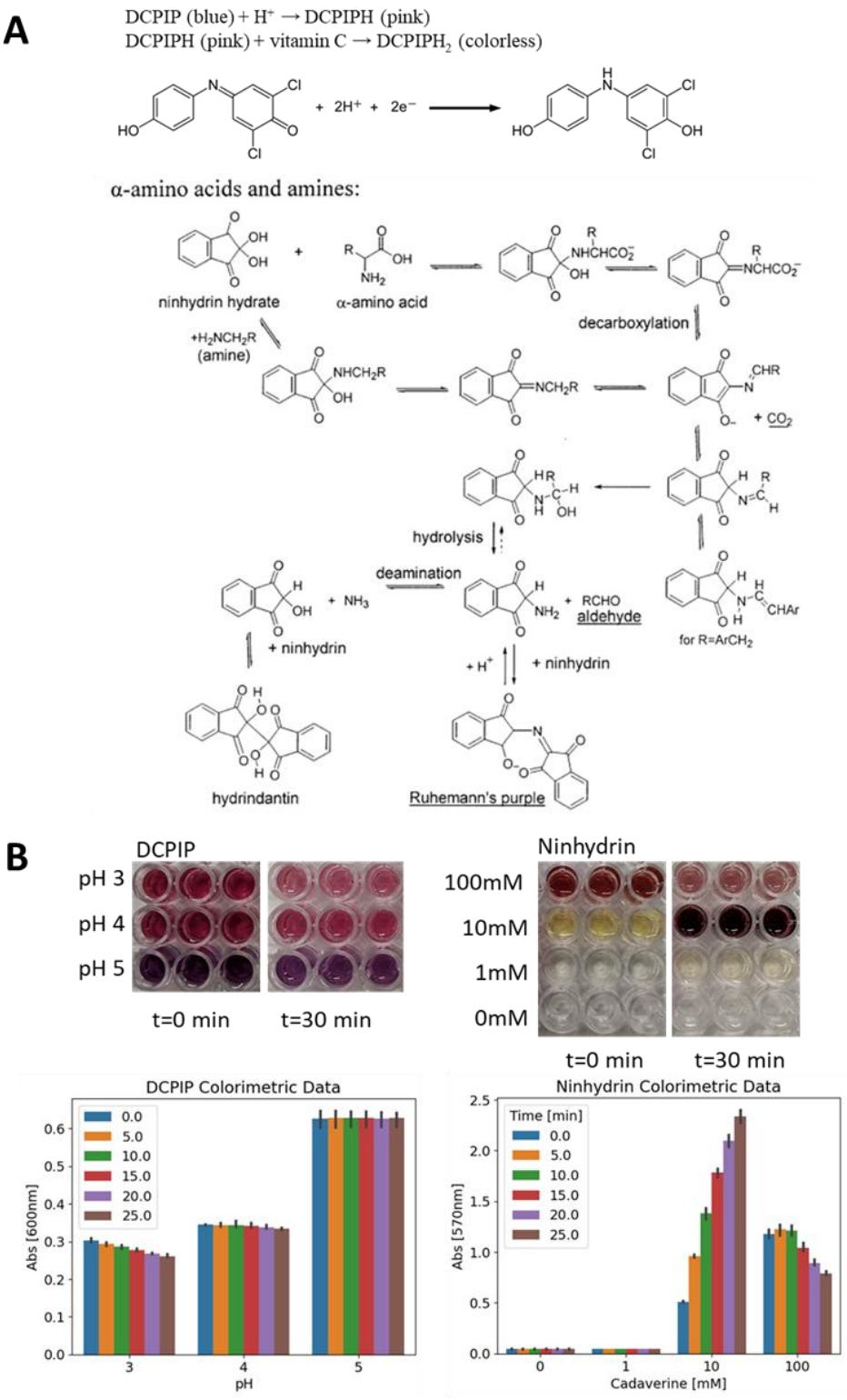


Figure 6.2 Mediator structures and colorimetric properties.

(A) Structure of mediators and reactions. (B) Colorimetric mediator assays. Error bars represent s.d. of technical triplicates.

To characterize the DCPIP and Ninhydrin modified screen printed electrodes, which we will refer to as DCPIP-SPE and Nin-SPE, we first compared their electrochemical measurements to unmodified screen-printed electrodes. To fabricate mediator modified electrodes, we used commercially available carbon ink SPEs (Zensor) as a base. These consisted of a carbon working electrode, a carbon counter electrode, and a pseudo-Ag/AgCl reference electrode printed onto cardboard material. We performed mediator modifications by simply drop casting mediator solutions onto the surface of the working and drying them at 60°C (see Methods). This elevated "curing" temperature helped in stabilizing the mediator during subsequent use. In **Figure 6.3A**, we measured cyclic voltammograms of the unmodified and mediator modified electrodes in buffer. In the cyclic voltammograms of DCPIP-SPEs in buffer compared to unmodified SPEs, we observe two oxidative peaks, reflecting the two steps of reduction of the DCPIP to DCPIPH and DCPIPH2. When comparing unmodified electrode measurements to Nin-SPEs, we observe a single oxidative peak around -0.25V which aligns with a previous report that used ninhydrin as a probe for a DNA hybridization sensor [159].

We then characterized the modified SPEs using the same vaginal microbiome relevant conditions previously tested colorimetrically. We observed that with lower pH, both oxidative peak potentials increased as seen in **Figure 6.3B**. This finding aligns with previous reports in which the redox potential of DCPIP increases with lower pH [160]. We then measured 0-100 mM of cadaverine in PBS using Nin-SPEs and observed that with increasing cadaverine concentration the amplitude of the oxidative peak increased and the oxidative peak potential shifted negatively. To capture these behaviors, we plotted the oxidative peak voltages from DCPIP-SPE voltammograms and the oxidative charges from Nin-SPE voltammograms in **Figure 6.3C**. The oxidative charges

(Q_ox) of these voltammograms were calculated by taking the total area under the curve for current negative current values. Here, we see the oxidative charge increasing linearly across 10-fold increases in cadaverine concentrations. This may be due to the larger amount of reducible resides available when ninhydrin complexes with amine groups. These various intermediates are depicted in **Figure 6.2A.**

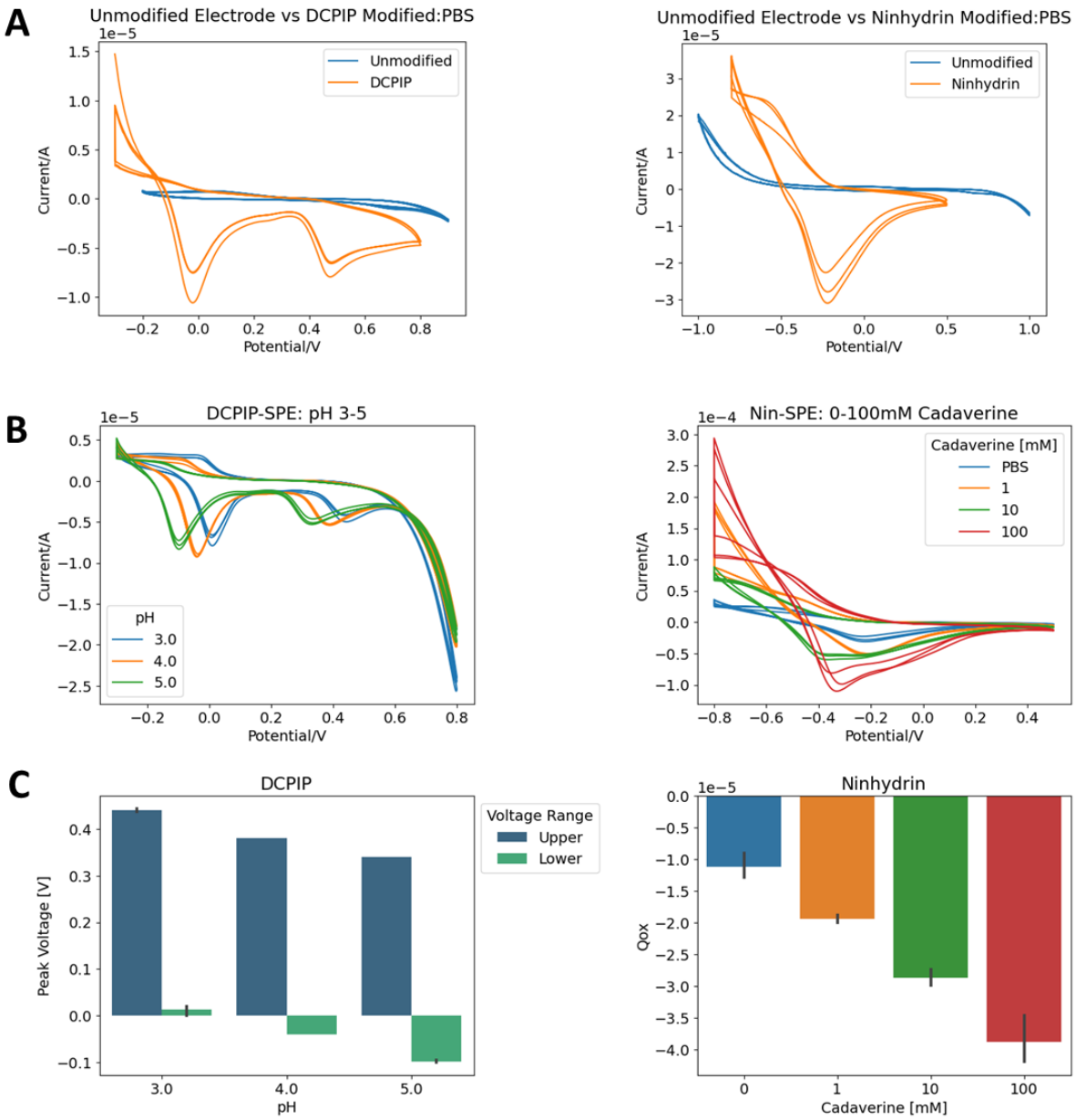


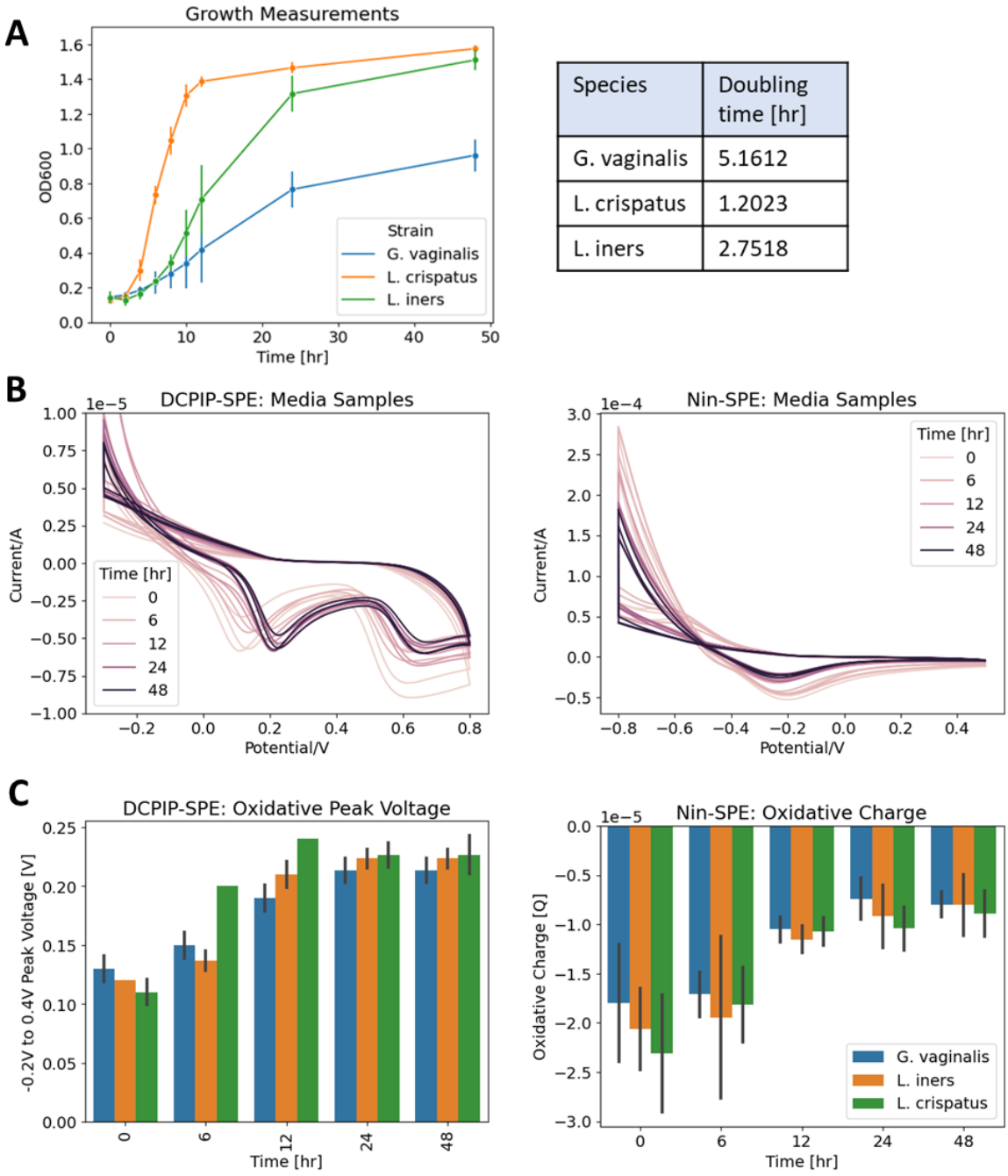
Figure 6.3 Modified SPE characterization.

(A) Cyclic voltammograms of modified vs unmodified SPEs. (B) Cyclic voltammograms of modified SPE controls (C) Metrics from control voltammograms. Error bars represent s.d. of technical triplicates.

6.2.2 Bacteria monoculture testing

We next tested the sensor on *in vitro* cultures of three vaginal associated microbes. We used *G. vaginalis*, *L. iners*, and *L. crispatus* as representative cultures of the following clinical states unhealthy (CSTIV), healthy with probability of disease (CSTIII), and healthy (CSTI), respectively. We cultured all of these species as monocultures in the same media (NYCIII) , see **Methods** for strain origins and storage protocols, and sampled their cell free supernatant over time to assess whether the sensors could discriminate between species at various growth stages over the course of 48 hours (**Figure 6.4A**). We note that these species have varying growth rates in the order of fastest to slowest growth: *L. crispatus*, *L. iners*, and *G. vaginalis*. The doubling times of these strains cultured in NYCIII media are located in **Figure 6.4A**.

We electrochemically measured the samples from each monoculture at 0, 6, 12, 24, and 48 hours with each sensor. Regardless of species, we observed that DCPIP-SPE oxidative peak potentials shift positively over time (**Figure 6.4B**), correlating to the expected acidification of the media as nutrient components were consumed and *Lactobacillus* species produce lactic acid. We also observed with Nin-SPE measures, that regardless of species the oxidative charge decreased with time, indicative of the consumption of amino acids in the media (**Figure 6.4B**). With respect to species-to-species differences, we typically saw *L. crispatus* with the highest DCPIP-SPE oxidative peak voltage (**Figure 6.4C**) which can be expected as it has been recorded to produce the most lactic acid between the two Lactobacillus species investigated [35]. Additionally, we found that *L. iners* had the second highest oxidative peak voltage as it also produces lactic acid, whereas *G. vaginalis* solely acidifies the media by nutrient consumption rather than acid production. These trends also appear with respect to the Nin-SPE measurements, where *L. crispatus* cultures have the largest peak amplification as noted by their oxidative charge values (**Figure 6.4C**). This may suggest that *L. crispatus* have consumed the least metabolites containing alpha amines.



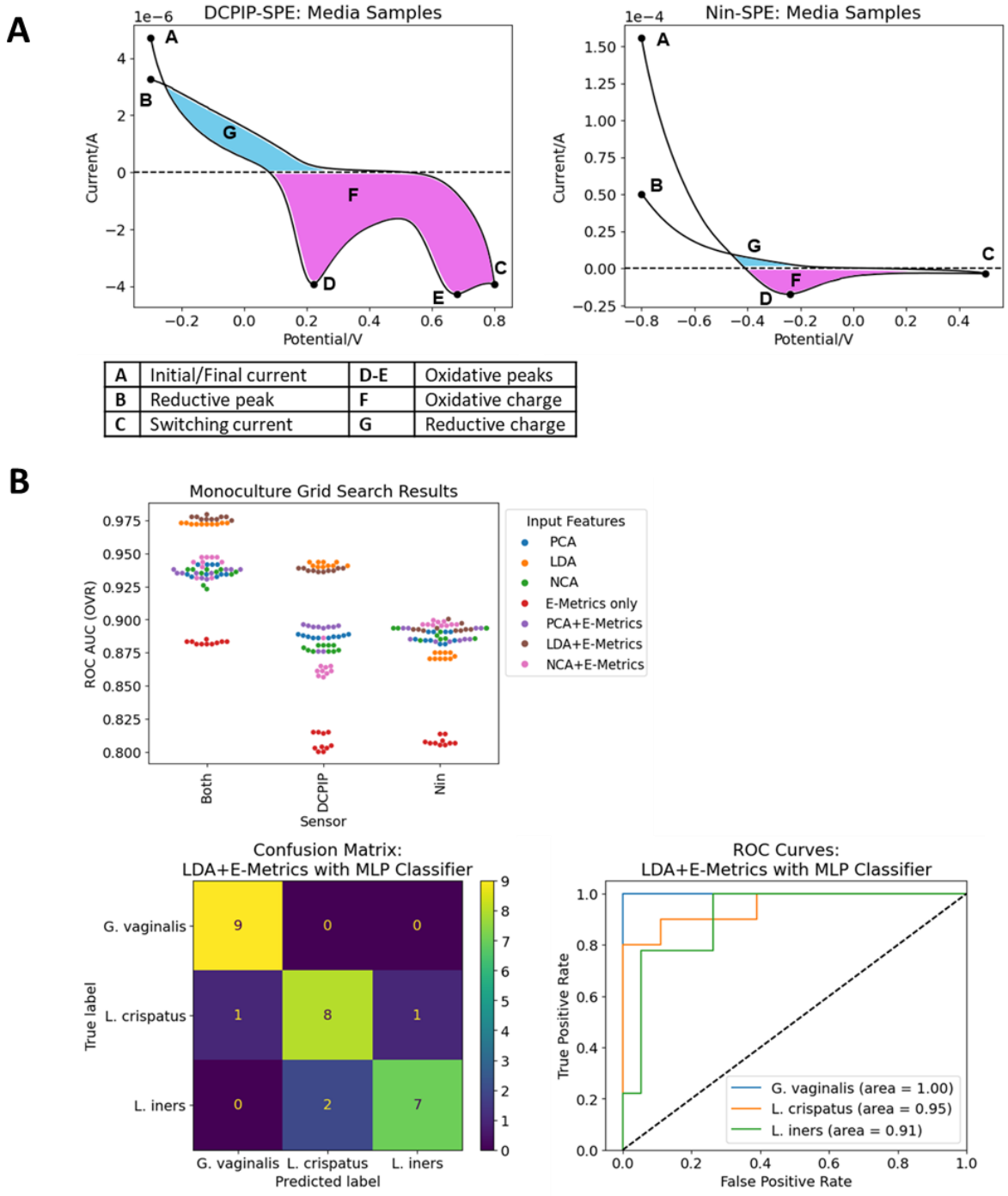
Species	Doubling time [hr]
G. vaginalis	5.1612
L. crispatus	1.2023
L. iners	2.7518

Figure 6.4 Microbial monoculture MEP.

(A) Monoculture growth curves. Error bars represent s.d. of biological duplicates. **(B) Cyclic voltammograms of monoculture culture media measured with modified SPEs. (C) Electrochemical metrics of monoculture culture media.** Error bars represent s.d. of triplicates of biological duplicates (n=6).

We then utilized the datasets generated from these monocultures to test and train a model to classify which species the electrical measurement is derived from. The dataset consisted of biological duplicates of the monocultures timepoints at 0, 6, 12, 24, and 48 hours that were measured in technical triplicates, resulting in a dataset with 90 samples, 30 samples per species class. Each sample consists of two cyclic voltammograms; one from each DCPIP-SPE and Nin-SPE measurement. Multiple approaches to preprocessing were tested to assess the optimal pipeline for model performance, all of which can be found in **Appendix Table 8.7**. We tested using the data from individual sensors as well as both sensors, and applied electrochemical feature metric extraction (**Figure 6.5A**) and various methods of dimension reduction to the voltammogram. We then applied these various preprocessing pipelines to a cross-validated grid search, in which we tested the various classifiers with different hyperparameters also listed in **Appendix Table 8.8**. The Receiver-Operator-Curve Area-Under-the-Curve (ROC-AUC) was used as a performance scoring metric for the grid search [6]. The ROC-AUC of the top 10 models from each preprocessing pipeline were plotted in **Figure 6.5B**, which showed that across all dataset groupings (i.e. one sensor's data or both sensors') using only electrochemical metrics worked the least effectively for predicting a sample's species. We also found that using the data from both sensors as inputs for the pipelines increased the model's overall prediction accuracy compared to models trained on a single sensor's data. Although these plots only display the top ten performing models per each preprocessing method, all grid search results can be found in **Appendix Table 8.9.** In **Figure 6.5B**, we also plotted the confusion matrix and ROC curves for the top pipeline, in which applied Linear Discriminant Analysis to the cyclic voltammograms from each modified SPE. Then, the resulting components in addition to electrochemical metrics were used inputs for a Multi-Layer Perceptron

classifier. The plotted predictions were generated using 30% of the dataset while the other 70% of the samples were used to train the classifier prior.



A	Initial/Final current	**D-E**	Oxidative peaks
B	Reductive peak	**F**	Oxidative charge
C	Switching current	**G**	Reductive charge

Figure 6.5 Electrochemical metrics and monoculture grid search results.

(A) Electrochemical metrics. (B) Monoculture grid search results. ROC Curves and Confusion Matrices.

6.2.3 Methodology transfer to vaginal mucus

We next tested our methodology on vaginal mucus samples obtained from human volunteers. Mucus samples were diluted 20-fold in water and vortexed to mix before measuring on the modified SPEs. We measured a total of 21 of samples and plotted their cyclic voltammograms in **Figure 6.6A**. The measurements collected with DCPIP-SPEs revealed differences in pH of samples through their variance in peak oxidative voltages (**Figure 6.6B)**. The Nin-SPE oxidative charge was fairly low, within range of the 48-hour monoculture media samples. This was indicative of the lower concentration of amines from the clinical samples, relative to *in vitro* bacteria cultures. This was unsurprising as the vaginal environment was expected to be less nutrient dense as culture media. Studies using analytical chemistry methods on vaginal secretion fluid have reported amine concentrations within hundreds to thousands of micromolar dependent on CST [161]. Comparatively, NYCIII media is a fairly rich media containing peptone, yeast extract and horse serum all of which are likely sources of amines. We then used the best performing pipeline trained on monoculture data, LDA and electrochemical metrics features into a MLP classifier, to predict the majority species of these mucus samples. In parallel, we performed Nugent Scoring and qPCR using primers for *L. crispatus, L. gasseri, L. iners, L. jensenii*, and *G. vaginalis* to determine the BV diagnosis and approximate CST type, respectively. qPCR results are shown in **Appendix Figure 8.11.**

In **Table 6.1**, we summarized the model's prediction results alongside the qPCR and Nugent scoring data. When interpreting Nugent Scores, the following rubrics is used for BV diagnosis: 0-3 is normal flora, 4-6 is intermediate or mixed flora, and 7-10 is considered BV positive. We note that 3 out of 21 samples yielded qPCR results of *L. gasseri* dominance, such that our model had never encountered samples from this species before. For all of the *L. gasseri* samples, the model predicted them to be *L. iners* and interestingly two of those samples had

intermediate Nugent scores. Of the remaining 18 samples, the model predicted 9 of them with the correct highest abundance species from the qPCR results. Comparatively, a dummy classifier which randomly predicts one of the three species with uniform distribution, was only able to predict 1 out of the 21 samples. It should be noted, that for most samples (13 out of 21) the qPCR resulted in amplification of multiple Lactobacillus strains. This is expected, as with single Lactobacillus species dominance it is common to observe the co-occurrence of other Lactobacillus strains at lower population densities, which has been documented by previous metagenomic analyses of patient samples [141, 142]. Interestingly, for most cases which the model incorrectly predicted *L. crispatus,* the qPCR results showed that in addition to *L. iners* other Lactobacillus species which characterize healthy CSTs were present. The one exception was the incorrect prediction of *L. crispatus* on a *G. vaginalis* sample, however it is notable that this sample was positively identified as BV with *Mobiluncus* species via Nugent scoring. Findings from studies of *Mobiluncus mulieris* within the context of BV has shown that it does not elevate biogenic amines but potentially induces proinflammatory immune responses due to increase in lipids and metabolites linked to oxidative stress [162]. This may affect our sensor's ability to classify this sample accurately, as the training data has not previously encountered samples with *Mobiluncus spp.*

Samples in which the model incorrectly predicted *L. iners* were all cases of *L. crispatus*. Interestingly, in all of these cases the qPCR results determined *L. crispatus* to be the only Lactobacillus species present out of the five tested. This may speak to the differences in metabolite composition between *L. crispatus* grown *in vitro* cell cultures compared to vaginal mucus microenvironment. Overall, these results show promising potential for the developed dual sensing pipeline. The model does not predict false positives in that, cases where *L. crispatus* were predicted on *L. iners* dominated poly-*Lactobacillus* samples the Nugent scores indicated negative

BV diagnosis. False positives, in which *L. iners* was assigned to *L. crispatus* individually dominated samples would be caught in a clinical which applies our methodology as an initial point of care screen. Such that, follow up clinical laboratory testing with Nugent scoring could be arranged after initial flagging by our method. To improve the developed methodology for applications in preventative care, in which intermediate Nugent scores could be more accurately screened, we recommend additional data collection to increase the diversity of the *in vitro* training set or a large number of available patient samples to serve as training data, the pipeline could increase prediction accuracy. This would open the possibility of decreasing false positives for intermediate and BV positive samples to the point at which following up with further clinical laboratory methods for result confirmation would be unnecessary.

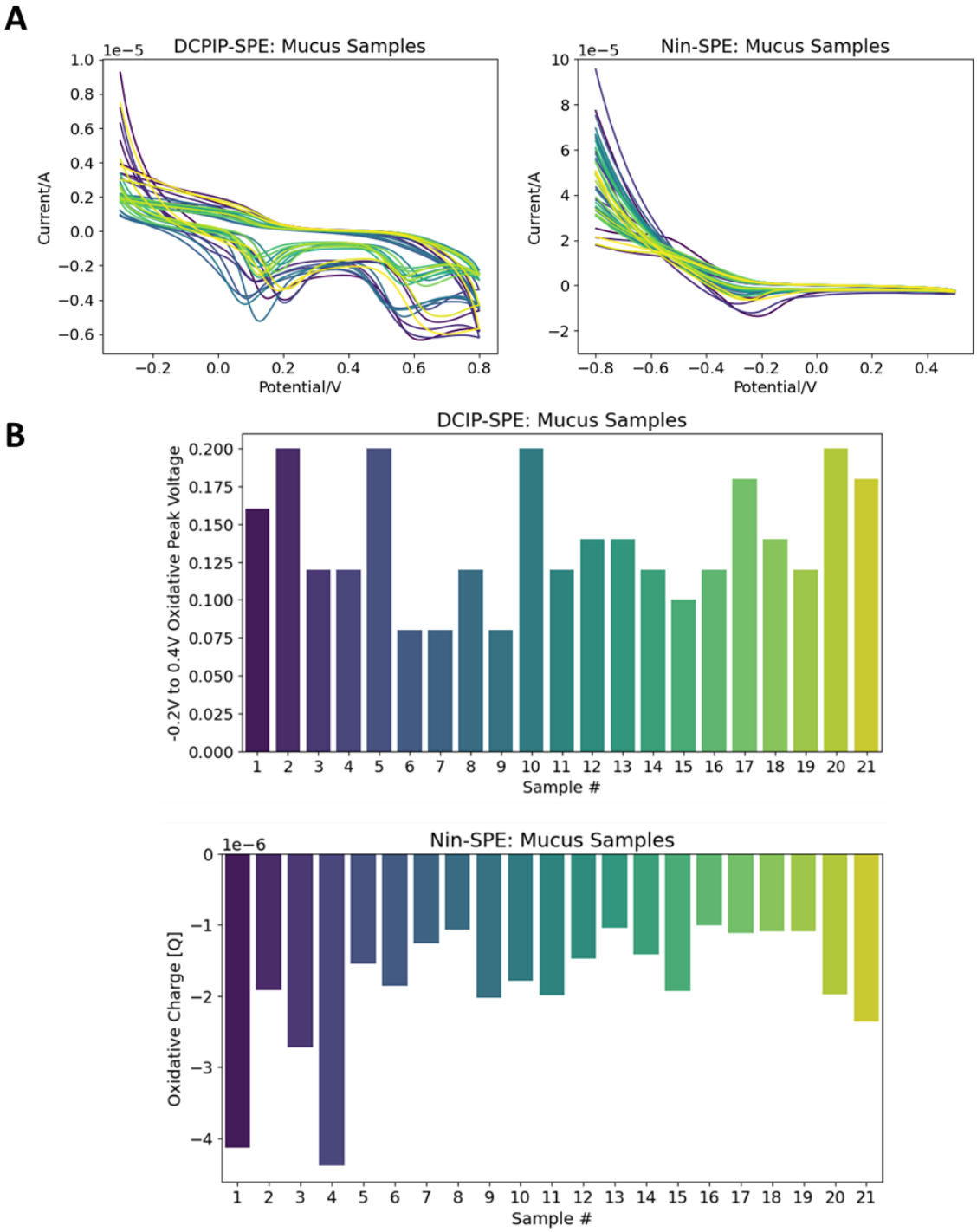


Figure 6.6 MEP characterization of human mucus.

(A) Mucus measured with modified SPEs. (B) Electrochemical metrics from mucus measurements.

Sample #	Prediction	qPCR Results (High to low abundance)	Average Nugent Score (n=5)
8	*L. crispatus*	*L. crispatus, L. jensenii*	2.2
13	*L. crispatus*	*L. crispatus, L. iners*	1.2
17	*L. crispatus*	*L. crispatus, L. iners*	1
2	*L. iners*	*L. iners, L. jensenii*	1.8
12	*L. iners*	*L. iners, L. jensenii*	4.6
20	*L. iners*	*L. iners, L. jensenii*	Not available
21	*L. iners*	*L. iners, L. crispatus, L. jensenii*	Not available
1	*L. iners*	*L. iners, L. crispatus, L. jensenii*	2.4
3	*L. iners*	*L. iners, L. crispatus*	4
4	*L. iners*	*L. gasseri*	3.2
15	*L. iners*	*L. gasseri*	6
16	*L. iners*	*L. gasseri*	6.4
5	*L. crispatus*	*L. iners, L. jensenii*	3.6
7	*L. crispatus*	*L. iners, L. jensenii*	3.4
10	*L. crispatus*	*L. iners, L. jensenii*	1.4
9	*L. crispatus*	*L. iners, L. crispatus, L. jensenii*	4
6	*L. crispatus*	*G. vaginalis*	10*
11	*L. iners*	*L. crispatus*	2.2
14	*L. iners*	*L. crispatus*	3.6
18	*L. iners*	*L. crispatus*	1
19	*L. iners*	*L. crispatus*	1.4

Table 6.1 Mucus species prediction results.

Predictions from a MLP classifier trained on electrochemical features and LDA components from monoculture measurements. qPCR results are listed from high to low abundance based on Cq, which can be found in **Appendix Figure 8.11**. The table is color coded based on prediction accuracy as follow: correct (green), not applicable (yellow), false negative (orange), and false positive (red).

6.3 Conclusions

In this chapter, we have developed novel disposable sensors for the determination of vaginal microbiome state. In conjunction with machine learning data pipelines, we demonstrated that these sensors could differentiate between vaginal related bacterial culture samples. This work aimed to create a simple and rapid method for predicting patient vaginal microbiome states that could feasibly be translated in clinic for point-of-care diagnosis. We developed two redox mediator modified screen printed electrodes, with the mediators DCPIP and ninhydrin respectively, that targeted two markers of vaginal microbiome health; pH and amine concentration. We first characterized these sensors using buffers at clinically relevant pHs and a biogenic diamine, cadaverine, commonly found in BV positive patients. The disposable sensors developed cost approximately $6 for a set of both modified SPEs and can be measured using available portable potentiostats within minutes, making our approach appropriate for low resource settings.

We then collected a dataset from *in vitro* cultures of three representative vaginal microbes (*L. crispatus, L. iners,* and *G. vaginalis*) to use for model development. We tested various data preprocessing methods used their outputs as model inputs. To fully test the pipelines, we performed a grid search across different classification models and their hyper parameters. The top performing pipeline consisted of using Linear Discriminant Analysis components and electrochemical metrics as the inputs of a Multi-layer Perceptron classifier. This pipeline yielded ROC-AUC metrics of greater than 0.9 for all prediction classes in a validation set containing 30% of the total dataset.

This trained pipeline was then used on human vaginal mucus samples to predict their dominant microbiome species. These samples were assessed using qPCR and Nugent scoring in parallel to validate the model predictions. Of the total 21 samples, 18 were dominated by species

the model was trained on. The model was able to accurately predict the dominant species of 9 of these samples. Samples which were incorrectly predicted as *L. crispatus* were all samples containing *L. iners* in addition to beneficial *Lactobacillus* species. Those which were incorrectly predicted as *L. iners* were all samples where only *L. crispatus* was detected with the exception of one *G. vaginalis* sample that was found to be also positive for *Mobiluncus spp.* These findings, indicate that the developed pipeline has promising potential to generate more accurate predictions on clinical samples. We suggest that collection of additional data consisting of other microbiome dominating species and various species compositions of *in vitro* cultures could improve model performance. Another potential approach to improve model performance would be collect datasets with modified culture conditions that more closely resemble the vaginal mucus microenvironment. Such cell culture models that aim to recapitulate the vaginal microenvironment have been developed through various avenues such as 3D cell culture [163-165], *ex vivo* tissue culture [166, 167], and microfluidic devices [168]. Lastly, we suggest that through the use of modern high throughput sequencing metagenomic analysis pipelines large volumes of patient samples could be processed and assigned CSTs which could be measured using our developed method in parallel to generate training data that would truly be observed clinically.

6.4 Methods

6.4.1 Colorimetric assays

For DCPIP characterization, citrate buffer was prepared at pH 3.0-5.0 by combining various ratios of 0.1M citric acid monohydrate and 0.2M sodium phosphate dibasic. 150 µL of buffer was mixed with 50 µL of 1mM DCPIP suspended in water. Samples were plated into a 96-well clear flat bottom plate and had their absorbance read at 600 nm using a TECAN Spark plate reader. For Ninhydrin characterization, 0-100 mM cadaverine was prepared from a liquid stock (Sigma-Aldrich). 180 µL of cadaverine sample was mixed with 20 µL of 100 mM of ninhydrin dissolved in water. These samples were read for absorbance at 570 nm in a 96-well clear flat bottom plate.

6.4.2 Preparation of mediator modified electrodes

Mediator modified electrodes were prepared using screen printed carbon electrodes (Zensor TE100s), which consist of a 3 mm graphitic carbon powder disk working electrode, graphitic carbon powder counter electrode, and Ag/AgCl pellet reference electrode. Mediator solutions were prepared in autoclaved distilled water at the following concentrations: 5 mM Ninhydrin (Sigma Aldrich) and 1 mM DCPIP (Sigma Aldrich). To prepare the mediator modified electrodes, 5 µL of mediator solution was drop casted onto the working electrode and then dried at 60°C for 30 minutes and cooled to room temperature before use. Modified electrodes were then stored at 4C, protected from light until ready for use.

6.4.3 Electrical measurements

Samples were measured by pipetting 50-100 µL onto the surface of the electrodes, ensuring to cover the working, reference, and counter electrode elements. Cyclic voltammograms were taken using a CHI1040c multichannel potentiostat. Custom housing for the screen-printed electrodes were used for ease of use. Housing consisted of laser cut acrylic boards to hold the SPEs in place

along with a three-pin adapter adhered using magnets for connecting the working, reference and counter electrodes.

For measurements using Ninhydrin modified electrodes the following cyclic voltammetry parameters were used: -0.8 to 0.5 V scan range, positive scanning polarity, 0.05 V/s scan rate, 0.001 V sample interval, and 1e-4 A/V sensitivity.

For measurements using DCPIP modified electrodes the following cyclic voltammetry parameters were used: -0.3 to 0.8 V scan range, positive scanning polarity, 0.05 V/s scan rate, 0.001 V sample interval, and 1e-5 A/V sensitivity.

6.4.4 Bacterial culture and media sampling

G. vaginalis, L. iners, and L. crispatus were plated on NYCIII agar plates for 48 hours at 37°C in anaerobic conditions. Liquid cultures were then seeded from colonies in NYCIII liquid media (cite ATCC) and incubated at 37C anaerobic for 48 hours. Seed cultures were then diluted 1:100 in NYCIII media and incubated overnight at 37C prior to co-culture experiments. Co-cultures were prepared by diluting these overnight cultures to ~0.1 OD600 in fresh NYCIII media and preparing cultures of various ratios. Anaerobic conditions were achieved by use of a vinyl anaerobic chamber (Coy Lab), which was maintained using 5% hydrogen gas mix and a palladium catalyst. G. vaginalis C0092E5 was kindly gifted from the Ravel Lab at the University of Maryland Baltimore Medical School, stocks were maintained in NYCIII media plus 25% glycerol. L. crispatus 33820 and L. iners were sourced from ATCC, stocks were maintained in MRS media plus 25% glycerol.

6.4.5 DNA extraction

DNA extractions were performed using Qiagen's DNeasy Tissue and Blood Extraction kits. A modified extraction protocol was utilized. Briefly, samples (10 mg for vaginal mucus and pellets

from 2 mL of bacterial cultures) were suspended in 180 µL enzyme lysis buffer (20mM tris-Cl, pH 8.0, 2mM sodium EDTA, 1.2% Triton X 100, 20 mg/mL lysozyme) and incubated at 37°C for 40 minutes. Then 20 µL proteinase K and 200 µL Buffer Al were added to the samples and vortexed to mix. Samples were then incubated at 56°C for 30 minutes, then 200 µL of ethanol was added to each sample and vortexed to mix. Samples were then transferred to a DNeasy spin column and washed two times with buffer AW1 and two times with AW2, then eluted in 50 µL buffer AE.

6.4.6 qPCR for CST approximation

Reaction mix was prepared using Power SYBR Master Mix (Thermofisher), the following components were combined per sample for a total of 10 µL per reaction: 5 µL Power SYBER 2X Master Mix, 0.5 µL 5 µM Forward Primer, 0.5 µL 5 µM Reverse Primer, 3.5 µL water, and 1 µL of template DNA. Extracted DNA samples were diluted 100X in autoclaved MilliQ water and used as template DNA per reaction. Each sample was plated in triplicates in a 384 well PCR plate with an optical adhesive cover (Thermofisher). An Applied Biosystems QuantStudio 7 was used to perform the following cycling program. Initial incubation at 95°C for 10 minutes then 40 cycles of 95°C for 15 seconds, 55°C for 35 seconds, and then 72°C for 35 seconds.

6.4.7 Cervicovaginal mucus collection and Nugent scoring

Human cervicovaginal mucus (CVM) was collected in accordance with protocol #2043110-3, as approved by the University of Maryland Institutional Review Board. Participants self-collected CVM using SoftDisc menstrual devices, as previously described.[169] Briefly, participants inserted the SoftDisc into the vagina for up to 1 min, and then removed the disc in a twisting motion to collect CVM. The SoftDisc was then placed into a 50 mL conical and spun at 300 x g for 5 min to collect CVM. Samples were characterized via Nugent score. Slides were imaged on a ZEISS

Axiovert 5 equipped with a 100x objective. Nugent scores were assigned based on Gram staining CVM smears.[169]

6.4.8 Model training and validation

The following electrochemical metrics were calculated per each cyclic voltammogram: Total Area Under the Curve (AUC), Reductive Charge (Q_red), Oxidative Charge (Q_ox), Redox charge ratio (Q_red/Q_ox), Max peak current (Ip_max), Minimum peak current (Ip_min, Initial Current (i_initial), Switching Current (i_switch), and Final Current (i_final). In addition to these metrics, the following dimensionality reduction techniques were used to extract features to be used in the training set: principal component analysis, linear discriminant analysis, and neighborhood component analysis. Data was scaled to its mean using standard scaling prior to these dimensionality reduction techniques. Lastly, the cycle numbers of the cyclic voltammogram were included as an input feature for model training.

Model development was performed using python's Scikit-Learn library [120]. Data cleaning, analysis, and visualization done using the following python libraries: NumPy[128], Pandas[129], SciPy[130], and Seaborn[90].

6.4.8.1 Model Descriptions

Below are brief technical descriptions of the classification models tested in this chapter.

6.4.8.1.1 Logistic Regression

Logistic Regression is a classification model where a logit transformation is applied to the input data and used to estimate the probability of the variables to belonging to each class [170].

6.4.8.1.2 Support Vector Machine

See 5.4.8.1.1

6.4.8.1.3 Decision Tree

See 5.4.8.1.2

6.4.8.1.4 Random Forest

See 5.4.8.1.3

6.4.8.1.5 K-Neighbors
See 5.4.8.1.4

6.4.8.1.6 Multi-Layer Perceptron
See 5.4.8.1.5

Chapter 7 Conclusions and Future Directions

Understanding the flow of information within biological systems allows for monitoring and ultimately outcome control through predictive design. This can be applied across all categories of biotechnology. In therapeutics, the molecular communication between drugs and their targets and cascading signals effected by the interaction are vital to safety and efficacy. Similarly in bioprocessing, optimal production is reliant on inducing proper metabolic conditions and gene expression at dialed in timings via molecular signals. In addition to using prior knowledge on these signaling pathways to understand molecular communication for therapeutic and bioprocessing development, extraction of information via modes of molecular communication is vital to providing diagnostics for evaluating system outputs. Synthetic biology has enabled microbial systems to be engineered for application in all of these contexts. For instance, bacterial hosts for live biotherapeutic delivery [168, 171, 172], industrial molecular production [63, 173], and as live biosensors. [174, 175] In this work, we explore signaling within microbial consortia as co-cultures provide expanded capabilities for complex function [173, 176].

The work presented in this dissertation focused on building experimental platforms and computational frameworks that enable the interpretation of molecular information in various scenarios. In **Chapters 3 and 4**, we developed multi-modal (i.e. QS, redox, electrogenetics) approaches to controlling bacterial consortia and a modeling framework to further study and design communication structures in multi-population systems. Our work to understand microbial consortia was also inspired by endogenous systems and the desire to understand how communication may function in their native environments. Naturally, microbial consortia live in communities by which their organization and subsequent function is dependent on their

interactions between niches. This has been studied across a variety of biomes, for example the rhizosphere [177, 178], the gastrointestinal tract [97, 179, 180], and human skin [181]. These environments are dynamic; influenced by spatial disturbances and fluctuating substrate availability causing shifts in intercommunity signaling and subsequent biome function. This work contributes to understanding the complexity of these community two-fold; (1) by studying complex bacterial co-cultures containing both redox and quorum sensing signaling modalities and (2) by considering the effect of various spatial conformations on communication. By studying synthetic consortia of various spatial organization and communication modalities, we provided insight into potential native interactions as well as established platforms for therapeutic development and bioprocessing optimization.

In **Chapters 5 and 6**, we leveraged the redox properties of biologics to electronically probe samples and subsequently employed ML methods to interpret electronic signals into biologic information. We applied these workflows to generate rapid diagnostics in two specific use cases: 1) in **Chapter 5** we quantified L-cysteine and antibody fragmentation for bioprocess monitoring, 2) in **Chapter 6** we classified vaginal microbiome species as an indicator of female reproductive health. The development of these electrochemical methods for monitoring biological systems expand capabilities for biologics to plug in to the "Internet of Things" [1, 2]. These types of tools provide additional methods of electronic diagnostics that expand the capacity for optimization of control systems in manufacturing environments [28, 182, 183]. In addition to industrial benefits, when applied clinically, these tools can be critical to the maintenance of chronic disease or tackling recurrent infection in susceptible populations [184-188].

In our future work, we plan to further explore the capabilities of our developed frameworks. With respect to investigating signaling in microbial communities, the potential to expand our network model to simulate consortia with more members and modes of communication is of interest. For our diagnostic workflows, we hope to improve prediction accuracy and develop hardware to facilitate even easier data collection. We suggest that increasing size and diversity of our training datasets in combination with increasing measurement throughput would be the next steps towards enabling technology transfer to relevant industrial and clinical sites. Eventual deployment of these diagnostic technologies would enable rapid insights to molecular information to work in conjunction with gold standard analytical approaches as quick screening tools.

Chapter 8 Appendices

8.1 Supplemental Information for Chapter 3

Strains	**Genotype**	**Source**
E. coli		
NEB10β	Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15e14- ϕ80*d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)*	New England Biolabs
PH04	W3110 Δ*lacU169 tna-2* Δ*luxS* Δ*ptsH*	[46]
ZK126	*E. coli* K-12 substr. W3110 Δ*lacU169 tna-2*	Laboratory Stock
LW7	W3110 Δ*lacU169 tna-2* Δ*luxS*::Kan	[189]
SW101	ZK126 Δ*soxRS*	This study
Plasmids	**Relevant Property**	**Source**
pZE12MCS	colEI origin, LlacO-1 promoter followed by multi-cloning site, Ap^r	ExpresSys
pZE-phzAG	pZE12MCS derivative, containing *phzA1-G1* under LlacO-1 promoter, Ap^r	This study
pZE-phzMS	pZE12MCS derivative, containing *phzM and phzS* under LlacO-1 promoter, Ap^r	This study
pZE-LacZα	pZE12MCS derivative, containing *lacZα* under LlacO-1 promoter, Ap^r	This study
pZE-phzAG-ptsH	pZE-phzAG derivative, containing growth control module (*ptsH* under *lasI* promoter, and dsRedExpress2 and LasR under constitutive T5 promoter)	This study
pZE-phzMS-ptsH	pZE-phzAG derivative, containing growth control module (*ptsH* under *lasI* promoter, and dsRedExpress2 and LasR under constitutive T5 promoter)	This study
pET-DsRed	pET200 derivative, containing *dsRed_Express_DR*	[190]
pCT10	pFZY1, *soxR*, *soxR* and *soxS* intergenic region, T7 RNA polymerase	This study
pTT01	pBR322, *soxR* gene and the overlapping divergent *soxR* and *soxS* promoters, phiLOV downstream of *soxS* promoter, Ap^r	[8]
pSox-LasI	pTT01 derivate, containing *lasI* under *soxS* promoter, Ap^r	[46]

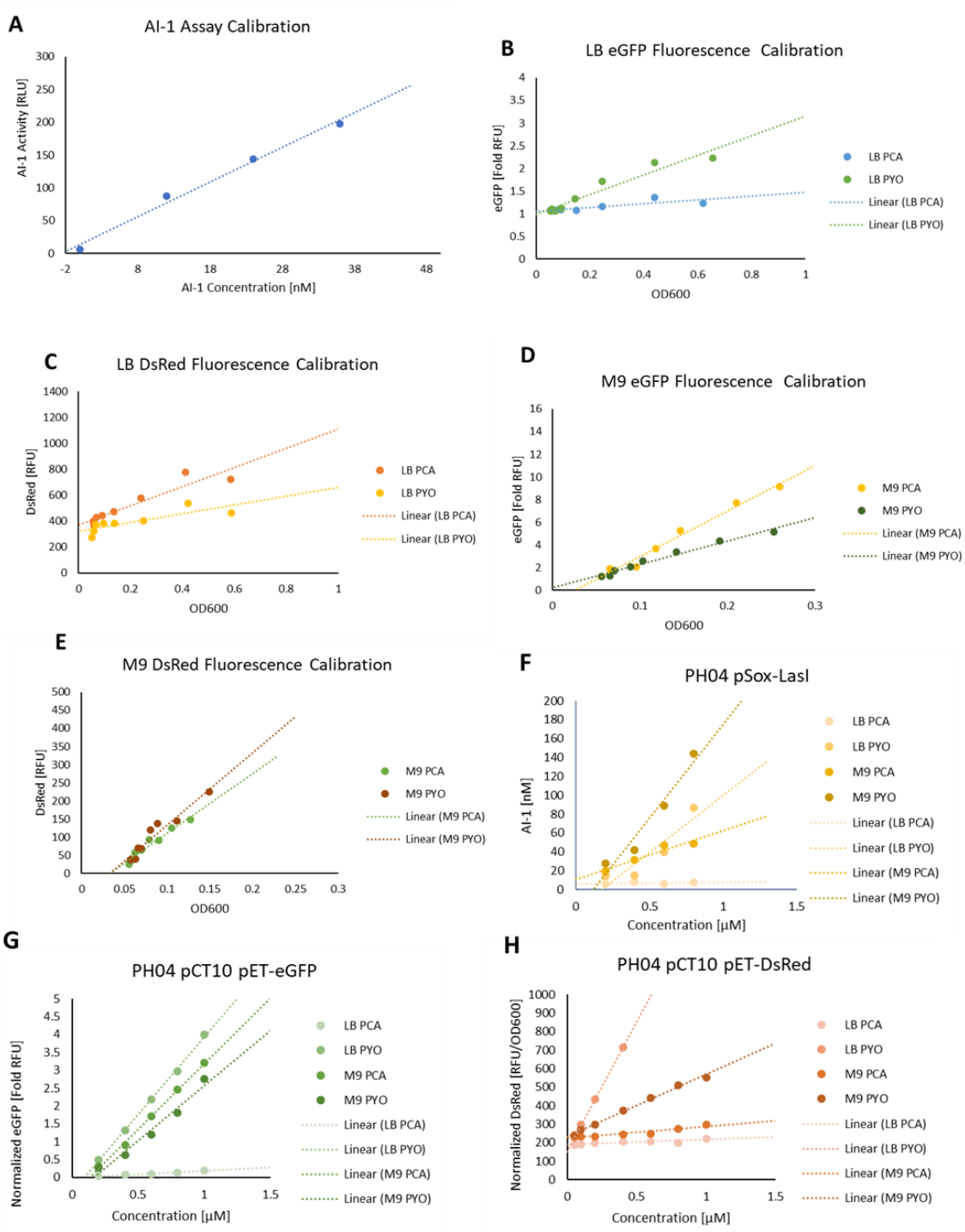
Appendix Table 8.1 Strains and plasmids used in Chapter 3.

Primer	Sequence
PciI-t7-term-RVS	TCGTACATGTCAAAAAACCCCTCAAGACCCGT
PciI-ptsH-RVS	ATCTACATGTTACTCGAGTTCCGCCATCAG
HindIII-phzM-RVS	CCTAGTAAGCTTTCATCAGGCCCTGGCA
HindIII-phzM-FWD	CACTGAAGCTTATGAATAATTCGAATCTTGCTGCTGC
KpnI-phzA-FWD	TGCATGGTACCATGAACGGTCAGCGGTAC
HindIII-phzG1-RVS	CATACAAGCTTCACGGTTGCAGGTAGC
KpnI-HindIII-RBS-phzS-FWD	TACGTGGTACCAAGCTTAAAGAGGAGAAAGCACCCATGAGCGAA
BamHI-phzS-RVS	AAGTAGGATCCTAGCGTGGCCGTTC
Fsoxp	GTTCTAGGATCCTTAGTTTTGTTC ATCTTCCAGCAAGCG
Rsoxp	GACTTTAAGCTTAAATCTGCCTCT TTTCAGTGTTCAGTTC
soxHP1	GTTCATCTTCCAGCAAGCGTGCGCCGGTACCTTCTTCTCCTAAGCGGTCGGTGTAGGCTGGAGCTGCTTC
soxHP2	CAATGGATGGAGCAATTACCCGCGCGGGAGTTAACGCGCGGGCAATAAAACATATGAATATCCTCCTTAG

Appendix Table 8.2 Primers used in Chapter 3.

Sample #	AI-1 [nM]	Ratio [A:B]	M9 [hr^{-1}]	M9 Error [hr^{-1}]	LB [hr^{-1}]	LB Error [hr^{-1}]
1	0	[1:4]	0.0991	0.0179	0.2017	0.0010
2	400	[1:4]	0.0955	0.0162	0.2024	0.0017
3	1000	[1:4]	0.0853	0.0058	0.2165	0.0010
4	0	[1:1]	0.0955	0.0096	0.2099	0.0043
5	400	[1:1]	0.0895	0.0021	0.2137	0.0021
6	1000	[1:1]	0.0878	0.0045	0.2220	0.0003
7	0	[19:1]	0.0967	0.0013	0.2154	0.0015
8	400	[19:1]	0.0914	0.0052	0.2243	0.0028
9	1000	[19:1]	0.1022	0.0051	0.2089	0.0020
10	0	[99:1]	0.1027	0.0073	0.2124	0.0008
11	400	[99:1]	0.0964	0.0050	0.2098	0.0002
12	1000	[99:1]	0.1013	0.0120	0.2189	0.0029
13	0	[299:1]	0.0982	0.0131	0.2083	0.0013
14	400	[299:1]	0.0985	0.0037	0.2271	0.0013
15	1000	[299:1]	0.1043	0.0068	0.2315	0.0003
16	0	[1:0]	0.0876	0.0077	0.2153	0.0023
17	400	[1:0]	0.0953	0.0069	0.2245	0.0015
18	1000	[1:0]	0.0988	0.0053	0.2309	0.0011
19	0	[0:1]	0.0937	0.0097	0.2149	0.0020
20	400	[0:1]	0.0893	0.0011	0.2119	0.0005
21	1000	[0:1]	0.0902	0.0077	0.2045	0.0015

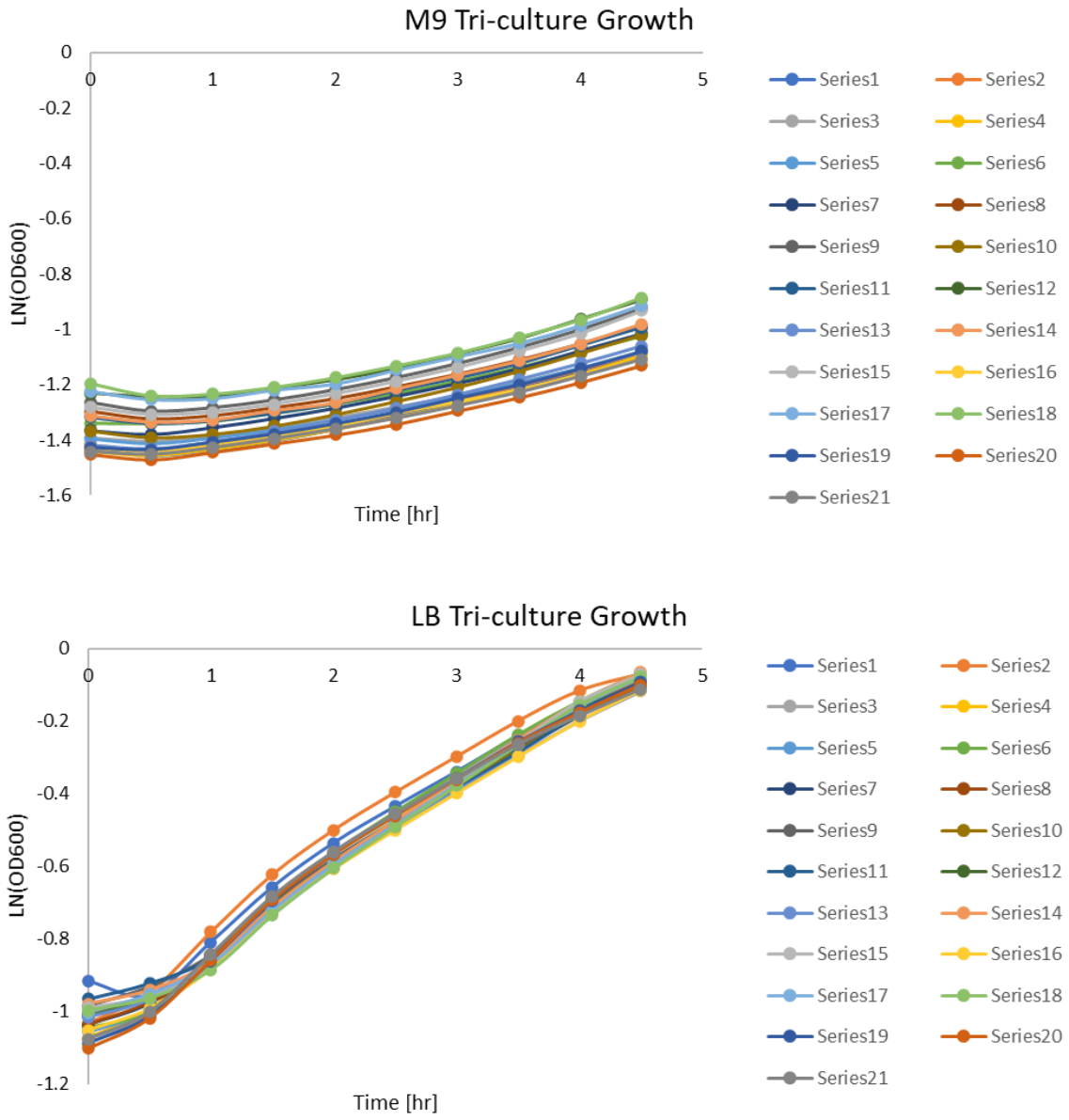
Appendix Table 8.3 Tri-culture growth rates from data depicted in Appendix Figure 8.2.



Appendix Figure 8.1 Product calibration and standardization.

Population C end products (AI-1, DsRed, and eGFP) were calibrated to standard measurements, then standardized in response to PYO and PCA in M9 and LB media. **A)** Linear region of standard curve from AI-1 luminescent reporter assay as described in Methods. **B-E)** Cultures of PH04 pCT10 pET-eGFP and PH04 pCT10-DsRed were inoculated from an overnight seed

culture at a 1:20 ratio in either M9 or LB media, with a concentration of 1µM PYO or PCA. These samples were incubated at 30ºC overnight, then measured for OD_{600} and fluorescence. **F-H)** PH04 pSox-LasI, PH04 pCT10 pET-eGFP, and PH04 pCT10 pET-DsRed were grown for 4 hours during log-phase in either LB or M9 with PYO or PCA at the concentrations indicated (ranging 0-1 µM), then eGFP, DsRed expression were measured using fluorescence and AI-1 was measured a cell based reporter assay (see Methods). eGFP measurements were normalized by subtracting background fluorescence from uninduced cells, then taking its fold increase over blank media.



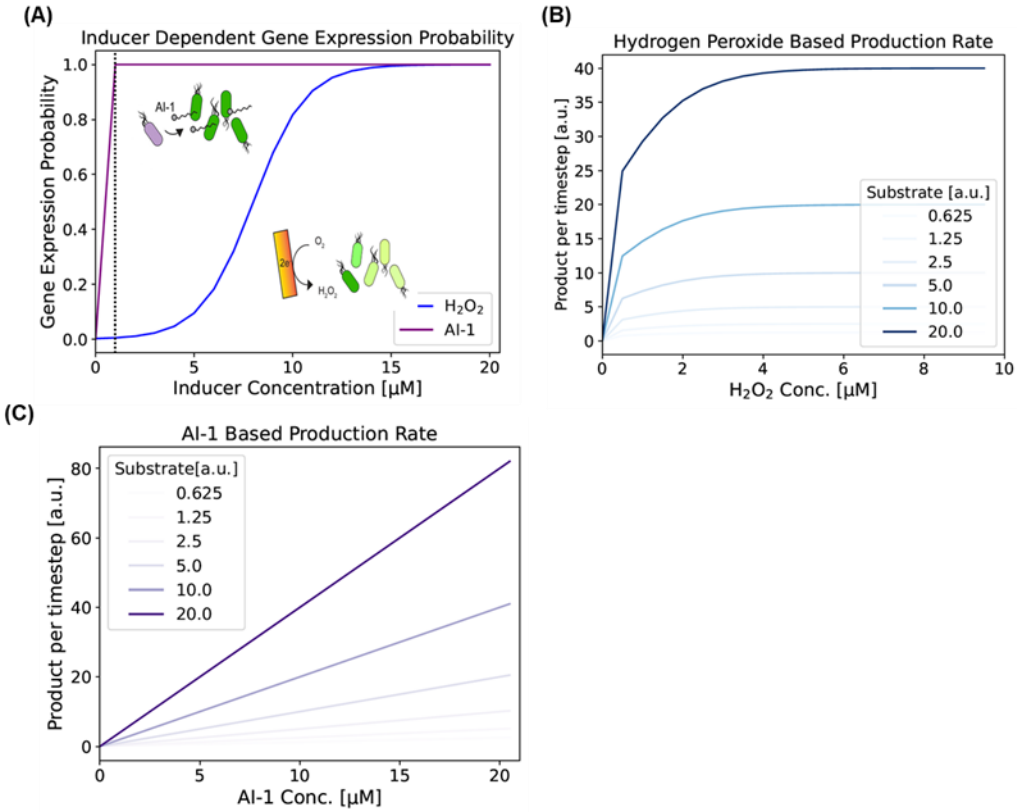
Appendix Figure 8.2 Tri-culture growth.

Population A (*E. coli* PH04 pZE-phzAG-ptsH-DsRed) and Population B (*E. coli* PH04 pZE-phzMS) were co-cultured together at initial inoculation ratios as indicated on the x-axis and with nanomolar AI-1 concentrations as indicated in the legend. After overnight growth, 20 µL of the PYO producing co-culture (~1.5 OD_{600}) was added to a culture of ~0.2 OD_{600} Population C (PH04 pCT10-pET-eGFP) with 200 µL final volume in fresh media. The tri-strain culture was then incubated for 4.5 hours, measuring OD_{600} every 30 minutes.

8.2 Supplemental Information for Chapter 4

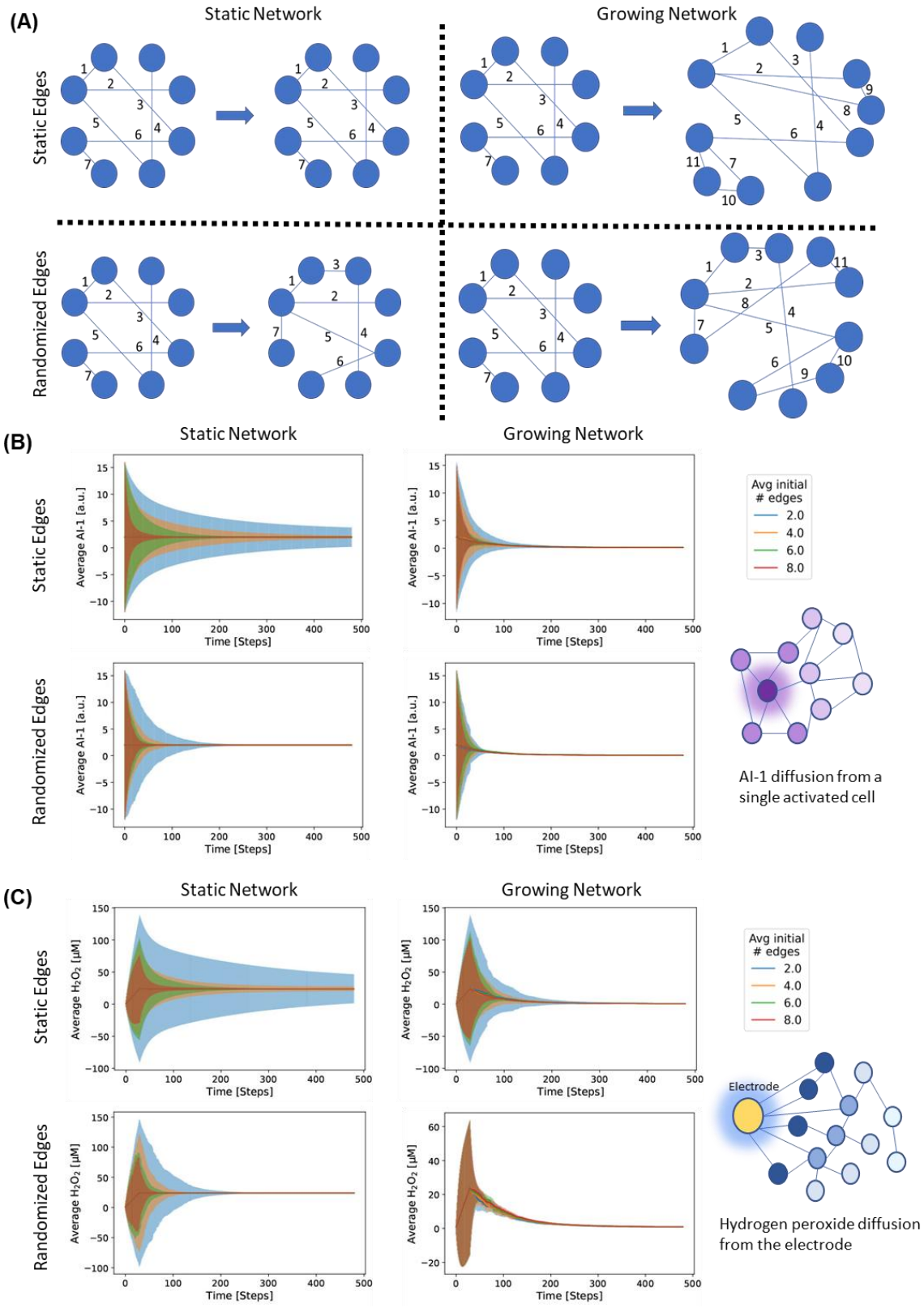
Parameters	Values	Descriptions
s_0	20	Initial substrate weight assigned to each node.
P_{div}	0.015	Probability of division at each timestep for each node.
α	1	Diffusion coefficient, generalized to apply to all signaling molecules in the network.
Δt	0.01	The timestep size by which diffusion across edges occurs.

Appendix Table 8.4 Simulation parameters with corresponding values used throughout Chapter 4 and descriptions.



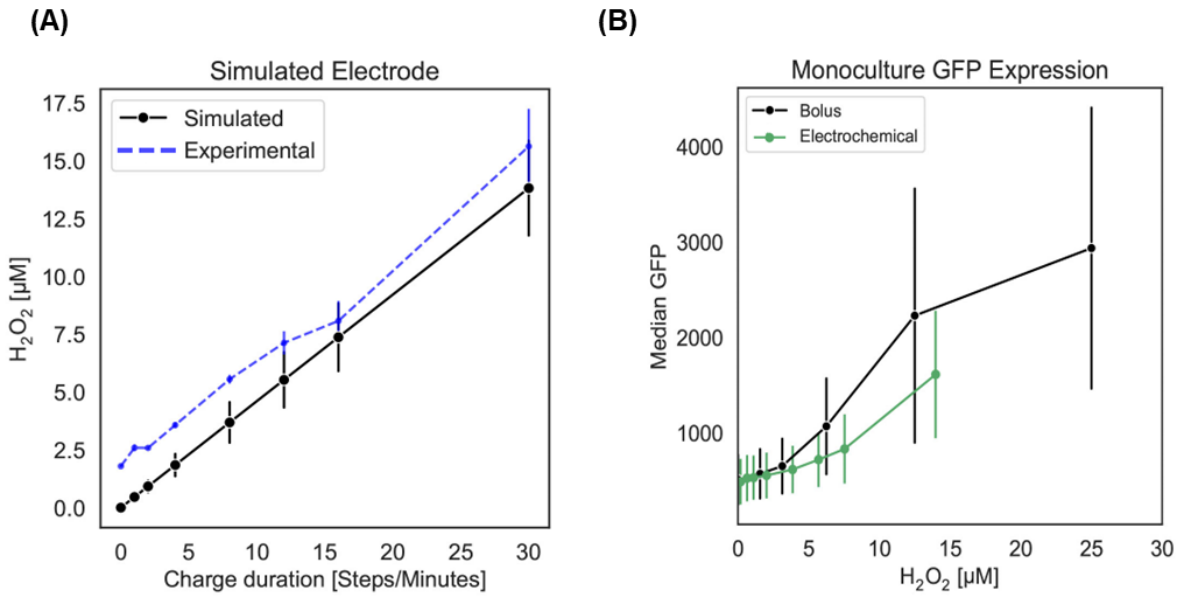
Appendix Figure 8.3 Gene activation and molecular production.

(A) The gene activation probability for AI-1(purple) and hydrogen peroxide (blue) inducible targets dependent on inducer concentration. **(B)** Hydrogen peroxide and substrate dependent molecular production rate for various substrate concentrations. **(C)** AI-1 and substrate dependent molecular production rate for various substrate concentrations.



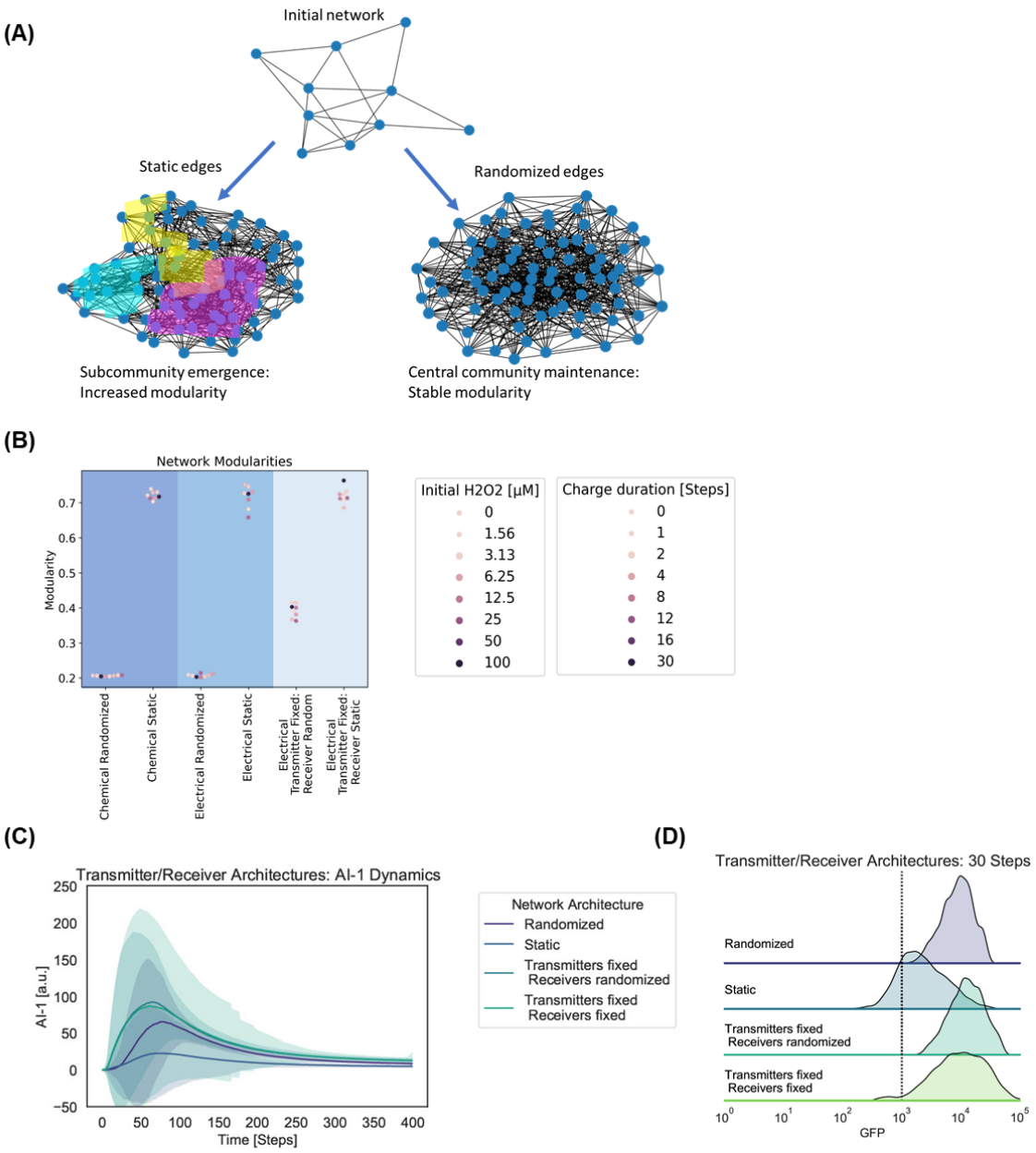
Appendix Figure 8.4 Diffusion schematic and dynamics.

(A) Schematic of edge dynamics for static and randomized edge properties for both static and growing networks. The transition arrows (blue) represent a time step. The static system has no change in nodes or edges. The growing network shows the doubling of one cell node and subsequent edge placement. Randomized edge networks demonstrate random edge rearrangement across time. **(B)** Average AI-1 of all nodes plotted over time for a static network, with static edges, a growing network with static edges, a static network with randomized edges, and a growing network with randomized edges. Shaded regions are error bars that indicate standard deviation of AI-1 for all nodes in the network. The schematic on the right depicts a graph with AI-1 diffusing from a single node in dark purple, with high saturation correlating to high concentration. **(C)** Average hydrogen peroxide level of all nodes plotted over time for a static network, with static edges, a growing network with static edges, a static network with randomized edges, and a growing network with randomized edges. Error bars (appear as shaded regions) indicate standard deviation of hydrogen peroxide for all nodes in the network. The schematic on the right depicts a graph with hydrogen peroxide diffusing from an electrode node with a high saturation of blue correlating to high concentration.



Appendix Figure 8.5 Electrode characteristics.

(A) Average hydrogen peroxide amongst an 100 node network dependent on charge duration (black) alongside experimentally measured hydrogen peroxide at various charge durations at -0.55V vs Ag/AgCl on a 2 mm gold disk electrode. Error bars in blue represent standard deviation of experimental results (n=3)[7]. Bars in black represent standard error for the simulated electrode (n=10). **(B)** Median GFP of the Monoculture Network at 180 timesteps post induction, with chemical induction (black) and electrochemical induction (green). Bars represent the standard deviation of a network data consisting of 10 simulation replicates.



Appendix Figure 8.6 Signal transmission measures for various Transmitter/Receiver network architectures.

(A) Graph schematics depicting network growth with and without edge randomization, showing subcommunity emergence and modularity increase in static edge networks compared to randomized edge networks. **(B)** Network modularity of differing node arrangements and edge dynamics, including fully static edges, at timestep 180. **(C)** AI-1 average across entire network over time for various Transmitter/Receiver network architectures, shadows represent standard deviation of aggregated network data from 10 simulation replicates. **(D)** The GFP distribution for each Transmitter/Receiver architecture tested at time 180 steps, induced with a charge duration of 30 steps. Distributions are representative of aggregated network data from 10 simulation replicates.

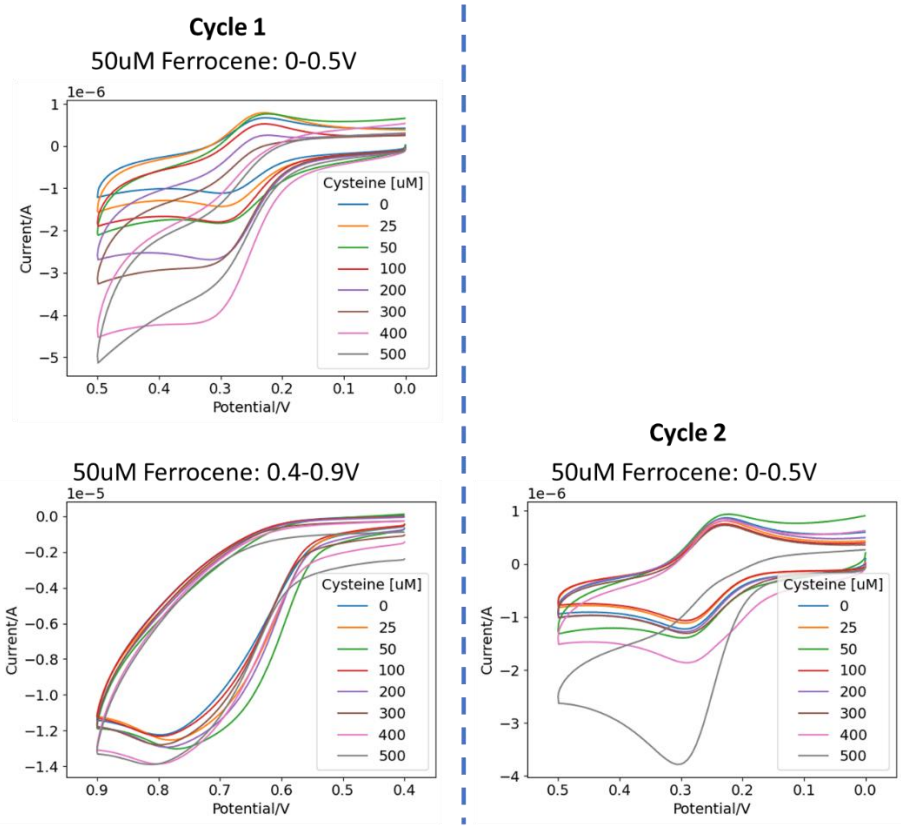
8.3 Supplemental Information for Chapter 5

Model	**Hyperparameters**	**Hyper parameter values**
Support Vector Machine	kernel	linear, poly, rbf, sigmoid
	C	1, 0.75, 0.5, 0.25, 0.1
Linear	None	
Random Forest	max_features	5, 10, 25, sqrt(features)
	n_estimators	100, 500, 1000
K-Neighbors	n_neighbors	1, 2, 3, 4, 5, 10
Decision Tree	max_features	5, 10, 25, sqrt(features)
Multi-Layer Perceptron	solver	sgd
	activation	logistic, relu, tanh
	alpha	0.01, 1e-3, 1e-4, 1e-5
	max_iter	10000
	hidden_layer_sizes	(5,) , (5,5), (15,15), (25,), (25,25), (100,)

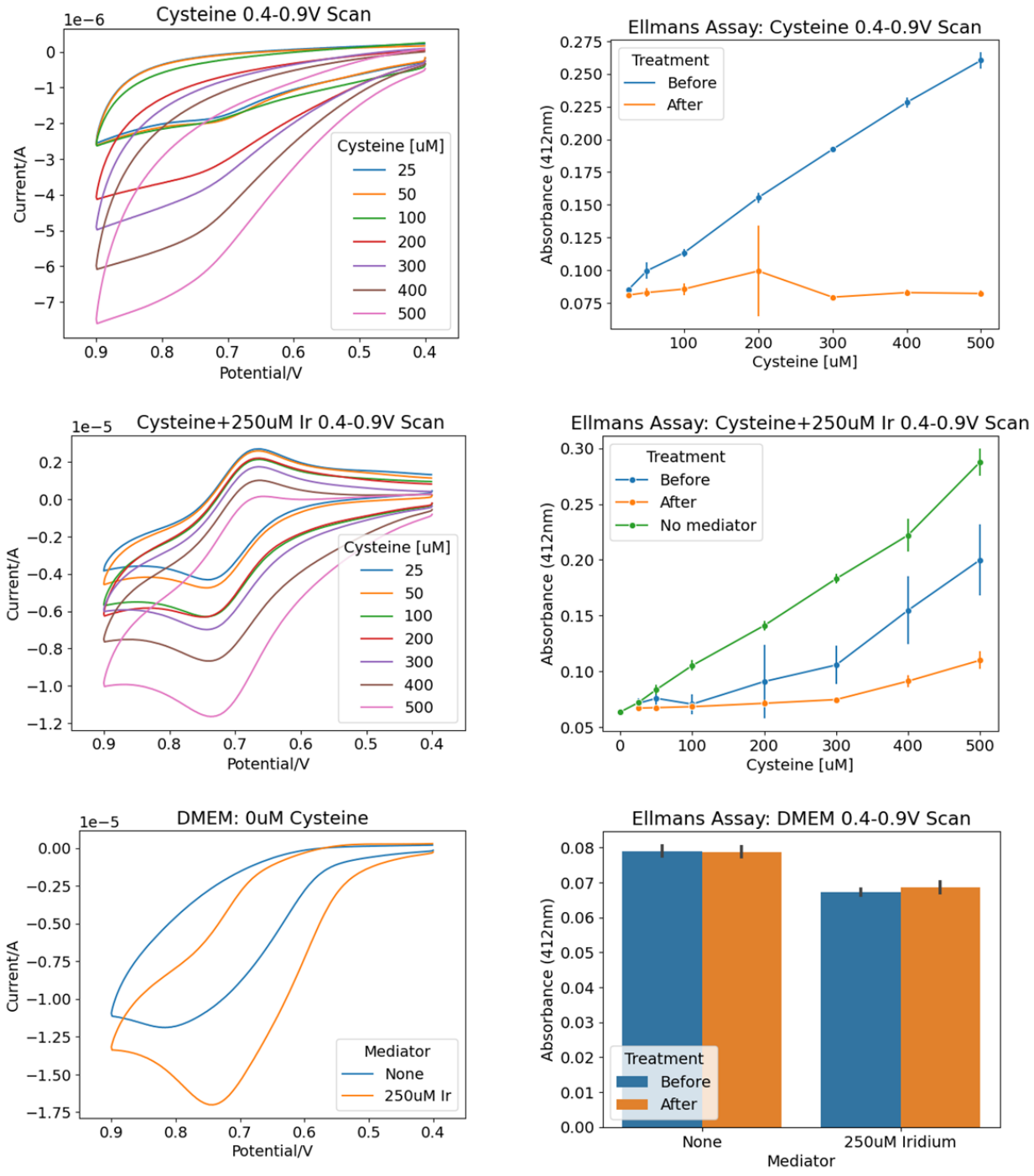
Appendix Table 8.5 Grid search models and hyperparameters.

Due to size constraints this table was uploaded to the following data repository: Chun, K. (2025). Chapter 5 and 6 Grid search results [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15305483

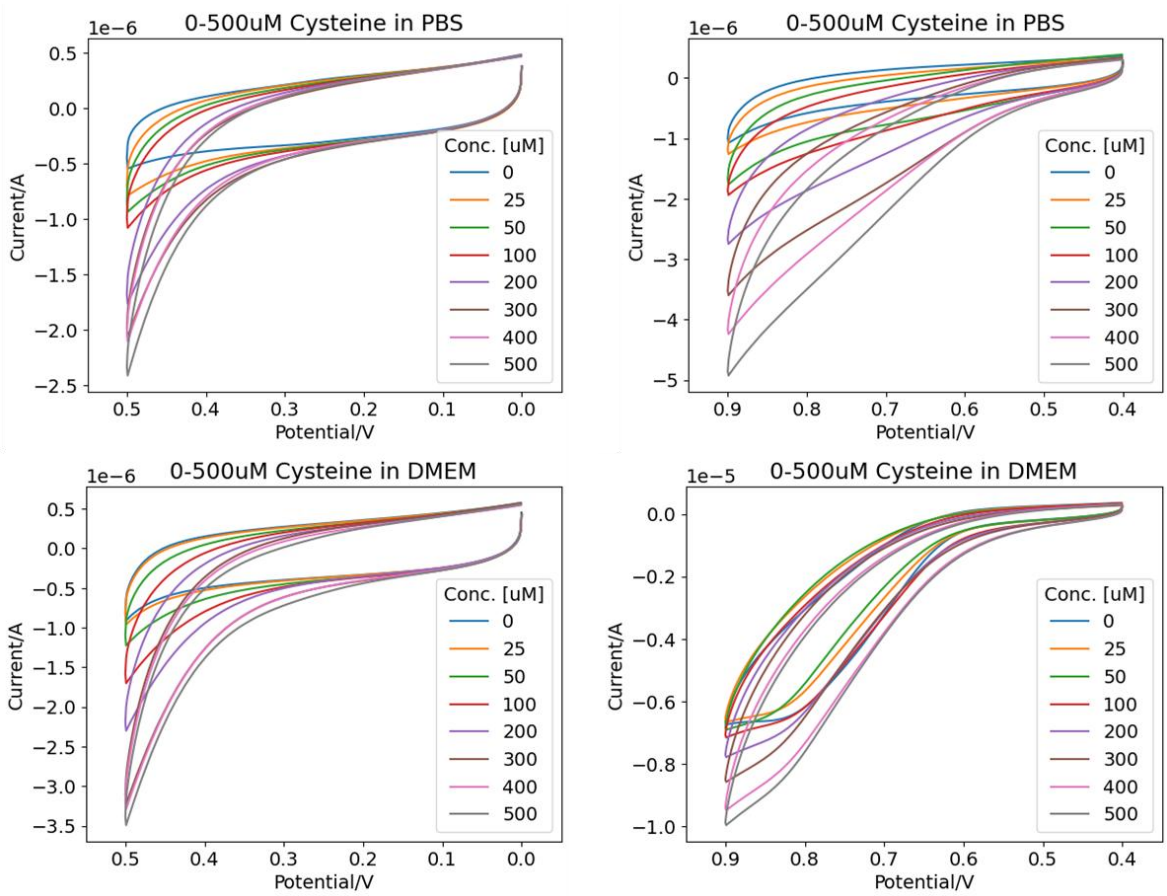
Appendix Table 8.6 Cysteine and antibody grid search results.



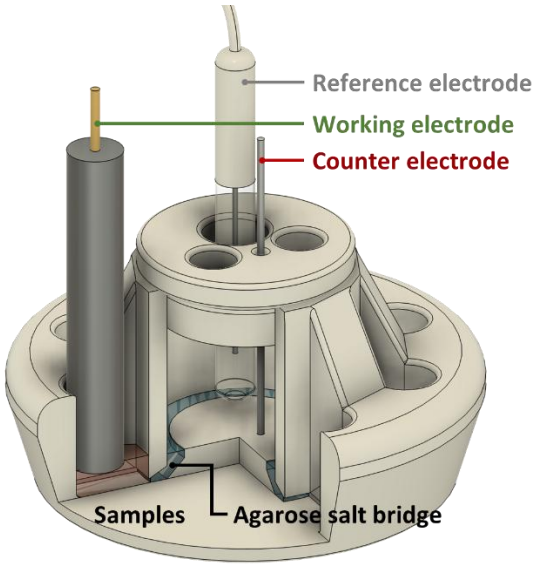
Appendix Figure 8.7 Effect of oxidative scan range on Ferrocene probing.



Appendix Figure 8.8 Thiol oxidation at Iridium scan ranges.



Appendix Figure 8.9 Cysteine cyclic voltammograms without mediators.



Appendix Figure 8.10 Diagram of 8-well 3D printed sampling device.

8.4 Supplemental Information for Chapter 6

Dimensionality Reduction Method	Electrochemical Metrics	Sensor
Principal Component Analysis	Included	DCPIP
Principal Component Analysis	None	DCPIP
Neighborhood Component Analysis	Included	DCPIP
Neighborhood Component Analysis	None	DCPIP
Linear Discriminant Analysis	Included	DCPIP
Linear Discriminant Analysis	None	DCPIP
None	Included	DCPIP
Principal Component Analysis	Included	Ninhydrin
Principal Component Analysis	None	Ninhydrin
Neighborhood Component Analysis	Included	Ninhydrin
Neighborhood Component Analysis	None	Ninhydrin
Linear Discriminant Analysis	Included	Ninhydrin
Linear Discriminant Analysis	None	Ninhydrin
None	Included	Ninhydrin
Principal Component Analysis	Included	Both
Principal Component Analysis	None	Both
Neighborhood Component Analysis	Included	Both
Neighborhood Component Analysis	None	Both
Linear Discriminant Analysis	Included	Both
Linear Discriminant Analysis	None	Both
None	Included	Both

Appendix Table 8.7 Data preprocessing approaches.

Classifier Model	**Hyperparameters**	**Hyper parameter values**
Support Vector Machine	kernel	linear, poly, rbf, sigmoid
	C	1, 0.1, 0.01
Logistic Regression	None	
Random Forest	max_features	5, 10, 25, sqrt(features)
	n_estimators	100, 500, 1000
K Neighbors	n_neighbors	1, 2, 3, 4, 5, 10
Decision Tree	max_features	5, 10, 25, sqrt(features)
Multi-Layer Perceptron	solver	sgd
	activation	logistic, relu, tanh
	alpha	0.01, 1e-3, 1e-4, 1e-5
	max_iter	10000
	hidden_layer_sizes	(5,) , (5,5), (15,15), (25,), (25,25), (100,)

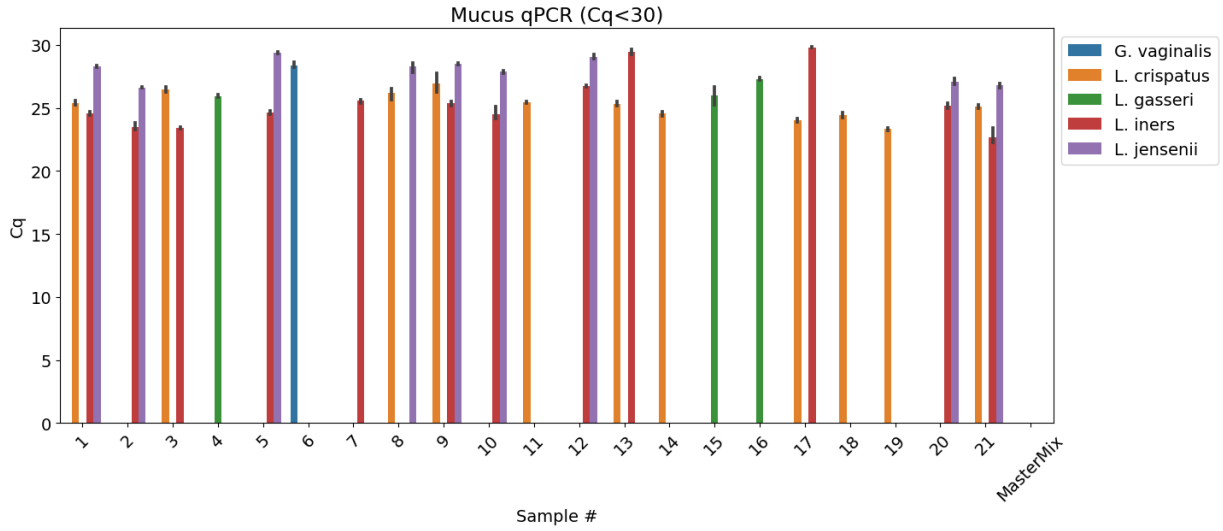
Appendix Table 8.8 Grid search hyperparameters.

Due to size constraints this table was uploaded to the following data repository: Chun, K. (2025). Chapter 5 and 6 Grid search results [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15305483

Appendix Table 8.9 Grid search results.

Bacteria	Forward Primer	Reverse Primer
L. crispatus	AGCGAGCGGAACTAACAGATTTAC	AGCTGATCATGCGATCTGCTT
L. iners	GTCTGCCTTGAAGATCGG	ACAGTTGATAGGCATCATC
L. jensenii	AAGTCGAGCGAGCTTGCCTATAGA	CTTCTTTCATGCGAAAGTAGC
L. gasseri	AGCGAGCTTGCCTAGATGAATTTG	TCTTTTAAACTCTAGACATGCGTC
G. vaginalis	TTACTGGTGTATCACTGTAAGG	CCGTCACAGGCTGAACAGT

Appendix Table 8.10 Primers used in Chapter 6.



Appendix Figure 8.11 Mucus qPCR results.

Bibliography

1. Kumar, S.; Tiwari, P.; Zymbler, M., Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big Data* **2019,** *6* (1).
2. Nizetic, S.; Solic, P.; Lopez-de-Ipina Gonzalez-de-Artaza, D.; Patrono, L., Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *J Clean Prod* **2020,** *274*, 122877.
3. Pesci, E. C.; Pearson, J. P.; Seed, P. C.; Iglewski, B. H., Regulation of las and rhl quorum sensing in Pseudomonas aeruginosa. *J Bacteriol* **1997,** *179* (10), 3127-32.
4. Waters, C. M.; Bassler, B. L., Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **2005,** *21*, 319-46.
5. Wang, S.; Payne, G. F.; Bentley, W. E., Quorum Sensing Communication: Molecularly Connecting Cells, Their Neighbors, and Even Devices. *Annu Rev Chem Biomol Eng* **2020,** *11*, 447-468.
6. Kim, E.; Zhao, Z.; Rzasa, J. R.; Glassman, M.; Bentley, W. E.; Chen, S.; Kelly, D. L.; Payne, G. F., Association of acute psychosocial stress with oxidative stress: Evidence from serum analysis. *Redox Biol* **2021,** *47*, 102138.
7. VanArsdale, E.; Pitzer, J.; Wang, S.; Stephens, K.; Chen, C. Y.; Payne, G. F.; Bentley, W. E., Electrogenetic Signal Transmission and Propagation in Coculture to Guide Production of a Small Molecule, Tyrosine. *ACS Synth Biol* **2022,** *11* (2), 877-887.
8. Tschirhart, T.; Kim, E.; McKay, R.; Ueda, H.; Wu, H. C.; Pottash, A. E.; Zargar, A.; Negrete, A.; Shiloach, J.; Payne, G. F.; Bentley, W. E., Electronic control of gene expression and cell behaviour in Escherichia coli through redox signalling. *Nat Commun* **2017,** *8*, 14030.
9. Terrell, J. L.; Tschirhart, T.; Jahnke, J. P.; Stephens, K.; Liu, Y.; Dong, H.; Hurley, M. M.; Pozo, M.; McKay, R.; Tsao, C. Y.; Wu, H.-C.; Vora, G.; Payne, G. F.; Stratis-Cullum, D. N.; Bentley, W. E., Bioelectronic control of a microbial community using surface-assembled electrogenetic cells to route signals. *Nature Nanotechnology* **2021,** *16* (6), 688-697.
10. Wu, F.; Menn, D. J.; Wang, X., Quorum-sensing crosstalk-driven synthetic circuits: from unimodality to trimodality. *Chem Biol* **2014,** *21* (12), 1629-38.
11. Scott, S. R.; Hasty, J., Quorum Sensing Communication Modules for Microbial Consortia. *ACS Synth Biol* **2016,** *5* (9), 969-77.
12. Lee, J.; Zhang, L., The hierarchy quorum sensing network in Pseudomonas aeruginosa. *Protein Cell* **2015,** *6* (1), 26-41.
13. Turkina, M. V.; Vikstrom, E., Bacteria-Host Crosstalk: Sensing of the Quorum in the Context of Pseudomonas aeruginosa Infections. *J Innate Immun* **2019,** *11* (3), 263-279.
14. Mavrodi, D. V.; Bonsall, R. F.; Delaney, S. M.; Soule, M. J.; Phillips, G.; Thomashow, L. S., Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from Pseudomonas aeruginosa PAO1. *J Bacteriol* **2001,** *183* (21), 6454-65.
15. Schiessl, K. T.; Hu, F.; Jo, J.; Nazia, S. Z.; Wang, B.; Price-Whelan, A.; Min, W.; Dietrich, L. E. P., Phenazine production promotes antibiotic tolerance and metabolic heterogeneity in Pseudomonas aeruginosa biofilms. *Nat Commun* **2019,** *10* (1), 762.
16. Evans, C. R.; Smiley, M. K.; Asahara Thio, S.; Wei, M.; Florek, L. C.; Dayton, H.; Price-Whelan, A.; Min, W.; Dietrich, L. E. P., Spatial heterogeneity in biofilm metabolism elicited by local control of phenazine methylation. *Proc Natl Acad Sci U S A* **2023,** *120* (43), e2313208120.

17. Demple, P. J. P. a. B., Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends in Biotechnology* **2001**.
18. Imlay, J. A., Transcription Factors That Defend Bacteria Against Reactive Oxygen Species. *Annual Review of Microbiology* **2015,** *69* (1), 93-108.
19. Chiang, S. M.; Schellhorn, H. E., Regulators of oxidative stress response genes in Escherichia coli and their functional conservation in bacteria. *Arch Biochem Biophys* **2012,** *525* (2), 161-9.
20. Bhokisham, N.; VanArsdale, E.; Stephens, K. T.; Hauk, P.; Payne, G. F.; Bentley, W. E., A redox-based electrogenetic CRISPR system to connect with and control biological information networks. *Nature Communications* **2020,** *11* (1).
21. Karbelkar, A. A.; Furst, A. L., Electrochemical Diagnostics for Bacterial Infectious Diseases. *ACS Infect Dis* **2020,** *6* (7), 1567-1571.
22. Dorothee Grieshaber, R. M., Janos V¨or¨os, Erik Reimhult, Electrochemical Biosensors - Sensor Principles and Architectures. *Sensors* **2008**.
23. Cesewski, E.; Johnson, B. N., Electrochemical biosensors for pathogen detection. *Biosens Bioelectron* **2020,** *159*, 112214.
24. Kim, E.; Li, J.; Kang, M.; Kelly, D. L.; Chen, S.; Napolitano, A.; Panzella, L.; Shi, X.; Yan, K.; Wu, S.; Shen, J.; Bentley, W. E.; Payne, G. F., Redox Is a Global Biodevice Information Processing Modality. *Proc IEEE Inst Electr Electron Eng* **2019,** *107* (7), 1402-1424.
25. Elgrishi, N.; Rountree, K. J.; McCarthy, B. D.; Rountree, E. S.; Eisenhart, T. T.; Dempsey, J. L., A Practical Beginner's Guide to Cyclic Voltammetry. *J Chem Educ* **2018,** *95* (2), 197-206.
26. Li, J.; Kim, E.; Gray, K. M.; Conrad, C.; Tsao, C. Y.; Wang, S. P.; Zong, G.; Scarcelli, G.; Stroka, K. M.; Wang, L. X.; Bentley, W. E.; Payne, G. F., Mediated Electrochemistry to Mimic Biology's Oxidative Assembly of Functional Matrices. *Advanced Functional Materials* **2020,** *30* (30).
27. Sajda, P., Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng* **2006,** *8*, 537-65.
28. Walsh, I.; Myint, M.; Nguyen-Khuong, T.; Ho, Y. S.; Ng, S. K.; Lakshmanan, M., Harnessing the potential of machine learning for advancing "Quality by Design" in biomanufacturing. *MAbs* **2022,** *14* (1), 2013593.
29. Saberi-Karimian, M.; Khorasanchi, Z.; Ghazizadeh, H.; Tayefi, M.; Saffar, S.; Ferns, G. A.; Ghayour-Mobarhan, M., Potential value and impact of data mining and machine learning in clinical diagnostics. *Critical Reviews in Clinical Laboratory Sciences* **2021,** *58* (4), 275-296.
30. Chen, X.; Lu, Y.; Chen, T.; Li, R., The Female Vaginal Microbiome in Health and Bacterial Vaginosis. *Front Cell Infect Microbiol* **2021,** *11*, 631972.
31. Javed, A.; Parvaiz, F.; Manzoor, S., Bacterial Vaginosis: An insight into the prevalence, alternative regimen treatments and it's associated resistance patterns. *Microbial Pathogenesis* **2019,** *127* (November 2018), 21-30.
32. Kumar, N.; Behera, B.; Sagiri, S. S.; Pal, K.; Ray, S. S.; Roy, S., Bacterial vaginosis: Etiology and modalities of treatment - A brief note. *Journal of Pharmacy and Bioallied Sciences* **2011,** *3* (4), 496-503.
33. Lu, H.; Jiang, P.-C.; Zhang, X.-D.; Hou, W.-J.; Wei, h.-H.; Lu, J.-Q.; Xu, G.-X.; Chen, Y.-P.; Ren, Y.; Wang, L.; Zhang, R.; Han, Y., Characteristics of bacterial vaginosis infection in cervical lesions with high risk human papillomavirus infection. *Int J Clin Exp Med* **2015**.

34. Gillet, E.; Meys, J. F.; Verstraelen, H.; Bosire, C.; Sutter, P. D.; Temmerman, M.; Broeck, D. V., Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect Dis* **2011,** *11* (10).
35. Petrova, M. I.; Reid, G.; Vaneechoutte, M.; Lebeer, S., Lactobacillus iners: Friend or Foe? *Trends Microbiol* **2017,** *25* (3), 182-191.
36. Money, D., The laboratory diagnosis of bacterial vaginosis. *Can J Infect Dis Med Microbiol* **2005**.
37. Coleman, J. S.; Gaydos, C. A., Molecular Diagnosis of Bacterial Vaginosis: an Update. *J Clin Microbiol* **2018,** *56* (9).
38. Fredricks, D. N.; Fiedler, T. L.; Thomas, K. K.; Oakley, B. B.; Marrazzo, J. M., Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. *J Clin Microbiol* **2007,** *45* (10), 3270-6.
39. Johnson, P. J.; Rosenbluth, E. M.; Nygaard, I. E.; Parikh, M. K.; Hitchcock, R. W., Development of a novel intra-vaginal transducer with improved dynamic response. *Biomedical Microdevices* **2009,** *11* (6), 1213-1221.
40. Jacek Lewandowski, K.-M. C., Wojciech Ni a ski, Design concepts and operation principles for Smart Intra-vaginal Health Monitoring Devices. *IEEE International Conference on e-Business Engineering* **2016**.
41. Crociati, M.; Sylla, L.; Stradaioli, G.; Monaci, M.; Zecconi, A., Assessment of Sensitivity and Profitability of an Intravaginal Sensor for Remote Calving Prediction in Dairy Cattle. *Sensors (Basel)* **2021,** *21* (24).
42. Choi, K. R.; Jang, W. D.; Yang, D.; Cho, J. S.; Park, D.; Lee, S. Y., Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends Biotechnol* **2019,** *37* (8), 817-837.
43. Shong, J.; Jimenez Diaz, M. R.; Collins, C. H., Towards synthetic microbial consortia for bioprocessing. *Curr Opin Biotechnol* **2012,** *23* (5), 798-802.
44. Jawed, K.; Yazdani, S. S.; Koffas, M. A., Advances in the development and application of microbial consortia for metabolic engineering. *Metabolic engineering communications* **2019,** *9*, e00095.
45. Roell, G. W.; Zha, J.; Carr, R. R.; Koffas, M. A.; Fong, S. S.; Tang, Y. J., Engineering microbial consortia by division of labor. *Microb Cell Fact* **2019,** *18* (1), 35.
46. Stephens, K.; Pozo, M.; Tsao, C. Y.; Hauk, P.; Bentley, W. E., Bacterial co-culture with cell signaling translator and growth controller modules for autonomously regulated culture composition. *Nat Commun* **2019,** *10* (1), 4129.
47. Dinh, C. V.; Chen, X.; Prather, K. L. J., Development of a Quorum-Sensing Based Circuit for Control of Coculture Population Composition in a Naringenin Production System. *ACS Synth Biol* **2020**.
48. Balagadde, F. K.; Song, H.; Ozaki, J.; Collins, C. H.; Barnet, M.; Arnold, F. H.; Quake, S. R.; You, L., A synthetic Escherichia coli predator-prey ecosystem. *Molecular systems biology* **2008,** *4*, 187.
49. Din, M. O.; Danino, T.; Prindle, A.; Skalak, M.; Selimkhanov, J.; Allen, K.; Julio, E.; Atolia, E.; Tsimring, L. S.; Bhatia, S. N.; Hasty, J., Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **2016,** *536* (7614), 81-85.
50. Jones, J. A.; Vernacchio, V. R.; Sinkoe, A. L.; Collins, S. M.; Ibrahim, M. H. A.; Lachance, D. M.; Hahn, J.; Koffas, M. A. G., Experimental and computational optimization of

an Escherichia coli co-culture for the efficient production of flavonoids. *Metab Eng* **2016,** *35*, 55-63.
51. Zhang, H.; Pereira, B.; Li, Z.; Stephanopoulos, G., Engineering Escherichia coli coculture systems for the production of biochemical products. *Proc Natl Acad Sci U S A* **2015,** *112* (27), 8266-71.
52. Zhang, H.; Stephanopoulos, G., Co-culture engineering for microbial biosynthesis of 3-amino-benzoic acid in Escherichia coli. *Biotechnology journal* **2016,** *11* (7), 981-7.
53. Zhou, K.; Qiao, K.; Edgar, S.; Stephanopoulos, G., Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat Biotech* **2015,** *33* (4), 377-383.
54. Shin, H. D.; McClendon, S.; Vo, T.; Chen, R. R., Escherichia coli binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Applied and environmental microbiology* **2010,** *76* (24), 8150-9.
55. Li, Z.; Wang, X.; Zhang, H., Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering. *Metab Eng* **2019,** *54*, 1-11.
56. Wang;, S.; Payne;, G. F.; Bentley, W. E., Repurposing E. coli by Engineering Quorum Sensing and Redox Genetic Circuits. *InTechOpen* **2018**.
57. Kylilis, N.; Tuza, Z. A.; Stan, G. B.; Polizzi, K. M., Tools for engineering coordinated system behaviour in synthetic microbial consortia. *Nat Commun* **2018,** *9* (1), 2677.
58. Tsao, C. Y.; Hooshangi, S.; Wu, H. C.; Valdes, J. J.; Bentley, W. E., Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of E. coli. *Metab Eng* **2010,** *12* (3), 291-7.
59. Bhokisham, N.; VanArsdale, E.; Stephens, K. T.; Hauk, P.; Payne, G. F.; Bentley, W. E., A redox-based electrogenetic CRISPR system to connect with and control biological information networks. *Nature communications* **2020,** *11* (1), 2427.
60. Ha, J.-H., Evidence of link between quorum sensing and sugar metabolism in *Escherichia coli* revealed cocrystal structures of LsrK and HPr. *Science Advances* **2018,** *4*.
61. Lindsay, A.; Ahmer, B. M., Effect of sdiA on biosensors of N-acylhomoserine lactones. *J Bacteriol* **2005,** *187* (14), 5054-8.
62. Jimenez, P. N.; Koch, G.; Thompson, J. A.; Xavier, K. B.; Cool, R. H.; Quax, W. J., The multiple signaling systems regulating virulence in Pseudomonas aeruginosa. *Microbiology and molecular biology reviews : MMBR* **2012,** *76* (1), 46-65.
63. da Silva, A. J.; Cunha, J. S.; Hreha, T.; Micocci, K. C.; Selistre-de-Araujo, H. S.; Barquera, B.; Koffas, M. A. G., Metabolic engineering of E. coli for pyocyanin production. *Metab Eng* **2021,** *64*, 15-25.
64. Cao, M.; Gao, M.; Suastegui, M.; Mei, Y.; Shao, Z., Building microbial factories for the production of aromatic amino acid pathway derivatives: From commodity chemicals to plant-sourced natural products. *Metab Eng* **2019**.
65. Marr, T. S. A., Effect of Nutrient Concentration on the Growth of Escherichia coli. *Journal of Bacteriology* **1971,** *107*.
66. Yang, Q.; Liang, Q.; Balakrishnan, B.; Belobrajdic, D. P.; Feng, Q. J.; Zhang, W., Role of Dietary Nutrients in the Modulation of Gut Microbiota: A Narrative Review. *Nutrients* **2020,** *12* (2).
67. Koop, A. H.; Hartley, M. E.; Bourgeois, S., A low-copy-number vector utilizing beta-galactosidase for the analysis of gene control elements. *Gene* **1987,** *52* (2-3), 245-56.

68. Datsenko, K. A.; Wanner, B. L., One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **2000,** *97* (12), 6640-5.
69. Mimee, M.; Tucker, A. C.; Voigt, C. A.; Lu, T. K., Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst* **2015,** *1* (1), 62-71.
70. Jiang, M.; Zhang, H., Engineering the shikimate pathway for biosynthesis of molecules with pharmaceutical activities in E. coli. *Curr Opin Biotechnol* **2016,** *42*, 1-6.
71. Cao, M.; Gao, M.; Suastegui, M.; Mei, Y.; Shao, Z., Building microbial factories for the production of aromatic amino acid pathway derivatives: From commodity chemicals to plant-sourced natural products. *Metab Eng* **2020,** *58*, 94-132.
72. Hwang, I. Y.; Koh, E.; Wong, A.; March, J. C.; Bentley, W. E.; Lee, Y. S.; Chang, M. W., Engineered probiotic Escherichia coli can eliminate and prevent Pseudomonas aeruginosa gut infection in animal models. *Nat Commun* **2017,** *8*, 15028.
73. Dinh, C. V.; Chen, X.; Prather, K. L. J., Development of a Quorum-Sensing Based Circuit for Control of Coculture Population Composition in a Naringenin Production System. *ACS Synth Biol* **2020,** *9* (3), 590-597.
74. Zhao, S.; Li, F.; Yang, F.; Ma, Q.; Liu, L.; Huang, Z.; Fan, X.; Li, Q.; Liu, X.; Gu, P., Microbial production of valuable chemicals by modular co-culture strategy. *World J Microbiol Biotechnol* **2022,** *39* (1), 6.
75. Gwon, D.-a.; Seo, E.; Lee, J. W., Construction of Synthetic Microbial Consortium for Violacein Production. *Biotechnology and Bioprocess Engineering* **2023**.
76. Quan, D. N.; Tsao, C. Y.; Wu, H. C.; Bentley, W. E., Quorum Sensing Desynchronization Leads to Bimodality and Patterned Behaviors. *PLoS Comput Biol* **2016,** *12* (4), e1004781.
77. Servinsky, M. D.; Terrell, J. L.; Tsao, C. Y.; Wu, H. C.; Quan, D. N.; Zargar, A.; Allen, P. C.; Byrd, C. M.; Sund, C. J.; Bentley, W. E., Directed assembly of a bacterial quorum. *ISME J* **2016,** *10* (1), 158-69.
78. Virgile, C.; Hauk, P.; Wu, H. C.; Shang, W.; Tsao, C. Y.; Payne, G. F.; Bentley, W. E., Engineering bacterial motility towards hydrogen-peroxide. *PLoS One* **2018,** *13* (5), e0196999.
79. VanArsdale, E.; Navid, A.; Chu, M. J.; Halvorsen, T. M.; Payne, G. F.; Jiao, Y.; Bentley, W. E.; Yung, M. C., Electrogenetic signaling and information propagation for controlling microbial consortia via programmed lysis. *Biotechnol Bioeng* **2023,** *120* (5), 1366-1381.
80. Barabasi, A. L., Network science. *Philos Trans A Math Phys Eng Sci* **2013,** *371* (1987), 20120375.
81. Gosak, M.; Markovic, R.; Dolensek, J.; Slak Rupnik, M.; Marhl, M.; Stozer, A.; Perc, M., Network science of biological systems at different scales: A review. *Phys Life Rev* **2018,** *24*, 118-135.
82. Sayama, H., The modeling and analysis of complex systems. *Math LibreTexts* **2023**.
83. Newman, M. E. J.; Girvan, M., Finding and evaluating community structure in networks. *Physical review E* **2004**.
84. Chun, K.; Stephens, K.; Wang, S.; Tsao, C. Y.; Payne, G. F.; Bentley, W. E., Parsed synthesis of pyocyanin via co-culture enables context-dependent intercellular redox communication. *Microb Cell Fact* **2021,** *20* (1), 215.

85. Newman, M.; Barabási, A.-L.; Watts, D. J., *The Structure and Dynamics of Networks*. Princeton University Press: Princeton, 2011.
86. Shiloach, J.; Kaufman, J.; Guillard, A. S.; Fass, R., Effect of Glucose Supply Strategy on Acetate Accumulation, Growth, and Recombinant Protein Production by Escherichia coli BL21 (DE3) and Escherichia coli JM 109. *Biotechnol Bioeng* **1996,** *49*.
87. Hagberg, A. A.; Schult, D. A.; Swart, P. J., Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science Conference* **2008**.
88. Menczer, F.; Fortunato, S.; Davis, C. A., *A First Course in Network Science*. 2020.
89. Hunter, J. D., Matplotlib: A 2D graphics environment. *Comput Sci Eng* **2007,** *9* (3), 90-95.
90. Waskom, M., seaborn: statistical data visualization. *Journal of Open Source Software* **2021,** *6* (60).
91. Zargar, A.; Quan, D. N.; Emamian, M.; Tsao, C. Y.; Wu, H. C.; Virgile, C. R.; Bentley, W. E., Rational design of 'controller cells' to manipulate protein and phenotype expression. *Metab Eng* **2015,** *30*, 61-68.
92. Li, J.; Attila, C.; Wang, L.; Wood, T. K.; Valdes, J. J.; Bentley, W. E., Quorum sensing in Escherichia coli is signaled by AI-2/LsrR: effects on small RNA and biofilm architecture. *J Bacteriol* **2007,** *189* (16), 6011-20.
93. Cornell, W. C.; Zhang, Y.; Bendebury, A.; Hartel, A. J. W.; Shepard, K. L.; Dietrich, L. E. P., Phenazine oxidation by a distal electrode modulates biofilm morphogenesis. *Biofilm* **2020,** *2*, 100025.
94. Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E., Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008,** *2008* (10).
95. Luo, Y.; Shoichet, M. S., A photolabile hydrogel for guided three-dimensional cell growth and migration. *Nat Mater* **2004,** *3* (4), 249-53.
96. Duraj-Thatte, A. M.; Manjula-Basavanna, A.; Rutledge, J.; Xia, J.; Hassan, S.; Sourlis, A.; Rubio, A. G.; Lesha, A.; Zenkl, M.; Kan, A.; Weitz, D. A.; Zhang, Y. S.; Joshi, N. S., Programmable microbial ink for 3D printing of living materials produced from genetically engineered protein nanofibers. *Nat Commun* **2021,** *12* (1), 6600.
97. Ciccarese, D.; Micali, G.; Borer, B.; Ruan, C.; Or, D.; Johnson, D. R., Rare and localized events stabilize microbial community composition and patterns of spatial self-organization in a fluctuating environment. *ISME J* **2022,** *16* (5), 1453-1463.
98. Wang, L.; Zhang, X.; Tang, C.; Li, P.; Zhu, R.; Sun, J.; Zhang, Y.; Cui, H.; Ma, J.; Song, X.; Zhang, W.; Gao, X.; Luo, X.; You, L.; Chen, Y.; Dai, Z., Engineering consortia by polymeric microbial swarmbots. *Nat Commun* **2022,** *13* (1), 3879.
99. Dsouza, A.; Constantinidou, C.; Arvanitis, T. N.; Haddleton, D. M.; Charmet, J.; Hand, R. A., Multifunctional Composite Hydrogels for Bacterial Capture, Growth/Elimination, and Sensing Applications. *ACS Appl Mater Interfaces* **2022,** *14* (42), 47323-47344.
100. Yanamandra, A. K.; Bhusari, S.; Del Campo, A.; Sankaran, S.; Qu, B., In vitro evaluation of immune responses to bacterial hydrogels for the development of living therapeutic materials. *Biomater Adv* **2023,** *153*, 213554.
101. Molinari, S.; Tesoriero, R. F., Jr.; Li, D.; Sridhar, S.; Cai, R.; Soman, J.; Ryan, K. R.; Ashby, P. D.; Ajo-Franklin, C. M., A de novo matrix for macroscopic living materials from bacteria. *Nat Commun* **2022,** *13* (1), 5544.

102. Ritacco, F. V.; Wu, Y.; Khetan, A., Cell culture media for recombinant protein expression in Chinese hamster ovary (CHO) cells: History, key components, and optimization strategies. *Biotechnol Prog* **2018,** *34* (6), 1407-1426.
103. Chung, W. K.; Russell, B.; Yang, Y.; Handlogten, M.; Hudak, S.; Cao, M.; Wang, J.; Robbins, D.; Ahuja, S.; Zhu, M., Effects of antibody disulfide bond reduction on purification process performance and final drug substance stability. *Biotechnol Bioeng* **2017,** *114* (6), 1264-1274.
104. Ulrich, K.; Jakob, U., The role of thiols in antioxidant systems. *Free Radical Bio Med* **2019,** *140*, 14-27.
105. Averill-Bates, D. A., The antioxidant glutathione. *Vitam Horm* **2023,** *121*, 109-141.
106. Mari, M.; Morales, A.; Colell, A.; Garcia-Ruiz, C.; Fernandez-Checa, J. C., Mitochondrial glutathione, a key survival antioxidant. *Antioxid Redox Signal* **2009,** *11* (11), 2685-700.
107. Komuczki, D.; Stadermann, A.; Bentele, M.; Unsoeld, A.; Grillari, J.; Mueller, M. M.; Paul, A.; Fischer, S., High cysteine concentrations in cell culture media lead to oxidative stress and reduced bioprocess performance of recombinant CHO cells. *Biotechnology journal* **2022,** *17* (11), e2200029.
108. Ali, A. S.; Chen, R.; Raju, R.; Kshirsagar, R.; Gilbert, A.; Zang, L.; Karger, B. L.; Ivanov, A. R., Multi-Omics Reveals Impact of Cysteine Feed Concentration and Resulting Redox Imbalance on Cellular Energy Metabolism and Specific Productivity in CHO Cell Bioprocessing. *Biotechnology journal* **2020,** *15* (8), e1900565.
109. Ali, A. S.; Raju, R.; Kshirsagar, R.; Ivanov, A. R.; Gilbert, A.; Zang, L.; Karger, B. L., Multi-Omics Study on the Impact of Cysteine Feed Level on Cell Viability and mAb Production in a CHO Bioprocess. *Biotechnology journal* **2019,** *14* (4), e1800352.
110. Vlasak, J.; Ionescu, R., Fragmentation of monoclonal antibodies. *MAbs* **2011,** *3* (3), 253-63.
111. Liu, H.; May, K., Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* **2012,** *4* (1), 17-23.
112. Moritz, B.; Stracke, J. O., Assessment of disulfide and hinge modifications in monoclonal antibodies. *Electrophoresis* **2017,** *38* (6), 769-785.
113. Lacy, E. R.; Baker, M.; Brigham-Burke, M., Free sulfhydryl measurement as an indicator of antibody stability. *Anal Biochem* **2008,** *382* (1), 66-8.
114. Trexler-Schmidt, M.; Sargis, S.; Chiu, J.; Sze-Khoo, S.; Mun, M.; Kao, Y. H.; Laird, M. W., Identification and prevention of antibody disulfide bond reduction during cell culture manufacturing. *Biotechnol Bioeng* **2010,** *106* (3), 452-61.
115. Liu, H.; Gaza-Bulseco, G.; Chumsae, C., Analysis of reduced monoclonal antibodies using size exclusion chromatography coupled with mass spectrometry. *J Am Soc Mass Spectrom* **2009,** *20* (12), 2258-64.
116. Motabar, D.; Kim, E.; Li, J.; Zhao, Z.; Mouchahoir, T.; Gallagher, D. T.; Schiel, J. E.; Garige, M.; Sourbier, C.; Payne, G. F.; Bentley, W. E., Detecting features of antibody structure through their mediator-accessible redox activities. *Nat Chem Biol* **2025,** *21* (2), 291-299.
117. Zhao, Z.; Ozcan, E. E.; VanArsdale, E.; Li, J.; Kim, E.; Sandler, A. D.; Kelly, D. L.; Bentley, W. E.; Payne, G. F., Mediated Electrochemical Probing: A Systems-Level Tool for Redox Biology. *ACS Chem Biol* **2021,** *16* (7), 1099-1110.

118. Dodero, G.; De Michieli, L.; Cavalleri, O.; Rolandi, R.; Oliveri, L.; Daccà, A.; Parodi, R., L-cysteine chemisorption on gold:: an XPS and STM study. *Colloid Surface A* **2000,** *175* (1-2), 121-128.
119. Falk, L. N.; Bentley, W. E.; Kelly, D. L.; Payne, G. F.; Kim, E., Measuring oxidative stress by the iridium reducing capacity assay (Ir-RCA). *Advances in Redox Research* **2025,** *15*.
120. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011,** *12*, 2825-2830.
121. Ying, X., An Overview of Overfitting and its Solutions. *J Phys Conf Ser* **2019,** *1168*.
122. Vidarsson, G.; Dekkers, G.; Rispens, T., IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* **2014,** *5*, 520.
123. Song, Y.; Cai, H.; Tan, Z.; Mussa, N.; Li, Z. J., Mechanistic insights into inter-chain disulfide bond reduction of IgG1 and IgG4 antibodies. *Appl Microbiol Biotechnol* **2022,** *106* (3), 1057-1066.
124. Natesan, R.; Dykstra, A. B.; Banerjee, A.; Agrawal, N. J., Heterogeneity in Disulfide Bond Reduction in IgG1 Antibodies Is Governed by Solvent Accessibility of the Cysteines. *Antibodies (Basel)* **2023,** *12* (4).
125. Li, F.; Vijayasankaran, N.; Shen, A. Y.; Kiss, R.; Amanullah, A., Cell culture processes for monoclonal antibody production. *MAbs* **2010,** *2* (5), 466-79.
126. Xu, J.; Rehmann, M. S.; Xu, M.; Zheng, S.; Hill, C.; He, Q.; Borys, M. C.; Li, Z. J., Development of an intensified fed-batch production platform with doubled titers using N-1 perfusion seed for cell culture manufacturing. *Bioresources and Bioprocessing* **2020,** *7* (1).
127. Habeeb, A. F., Reaction of protein sulfhydryl groups with Ellman's reagent. *Methods Enzymol* **1972,** *25*, 457-64.
128. Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; Del Rio, J. F.; Wiebe, M.; Peterson, P.; Gerard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E., Array programming with NumPy. *Nature* **2020,** *585* (7825), 357-362.
129. team, T. p. d., pandas-dev/pandas: Pandas. *Zenodo* **2024**.
130. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy, C., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **2020,** *17* (3), 261-272.
131. Schneider, A.; Hommel, G.; Blettner, M., Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* **2010,** *107* (44), 776-82.
132. Aiassa, S.; Ny Hanitra, I.; Sandri, G.; Totu, T.; Grassi, F.; Criscuolo, F.; De Micheli, G.; Carrara, S.; Demarchi, D., Continuous monitoring of propofol in human serum with fouling compensation by support vector classifier. *Biosens Bioelectron* **2021,** *171*, 112666.
133. Chih-Wei Hsu, C.-C. C., and Chih-Jen Lin, A Practical Guide to Support Vector Classification. *2010*.

134. Austin, P. C.; Tu, J. V.; Ho, J. E.; Levy, D.; Lee, D. S., Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol* **2013,** *66* (4), 398-407.
135. Breiman, L., Random forests. *Machine Learning* **2001,** *45* (1), 5-32.
136. Manandhar, I.; Alimadadi, A.; Aryal, S.; Munroe, P. B.; Joe, B.; Cheng, X., Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. *Am J Physiol Gastrointest Liver Physiol* **2021,** *320* (3), G328-G337.
137. Zhang, Z., Introduction to machine learning: k-nearest neighbors. *Ann Transl Med* **2016,** *4* (11), 218.
138. Murtagh, F., Multilayer perceptrons and regression for classification. *Neurocomputing* **1991**.
139. Cleveland, M. H.; Karageorgos, I. L.; Marino, J. P.; Tarlov, M. J.; Yandrofski, K. S.; Zangmeister, R. A.; Kelman, Z., Recommended nomenclature convention for the NISTCHO cell line and its product monoclonal antibody, cNISTmAb. *MAbs* **2025,** *17* (1), 2490789.
140. Greenbaum, S.; Greenbaum, G.; Moran-Gilad, J.; Weintraub, A. Y., Ecological dynamics of the vaginal microbiome in relation to health and disease. *Am J Obstet Gynecol* **2019,** *220* (4), 324-335.
141. Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G. M.; Koenig, S. S.; McCulle, S. L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C. O.; Brotman, R. M.; Davis, C. C.; Ault, K.; Peralta, L.; Forney, L. J., Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* **2011,** *108 Suppl 1* (Suppl 1), 4680-7.
142. Gajer, P.; Brotman, R. M.; Bai, G.; Sakamoto, J.; Schutte, U. M.; Zhong, X.; Koenig, S. S.; Fu, L.; Ma, Z. S.; Zhou, X.; Abdo, Z.; Forney, L. J.; Ravel, J., Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* **2012,** *4* (132), 132ra52.
143. Javed, A.; Parvaiz, F.; Manzoor, S., Bacterial vaginosis: An insight into the prevalence, alternative treatments regimen and it's associated resistance patterns. *Microb Pathog* **2019,** *127*, 21-30.
144. Ma, B.; Forney, L. J.; Ravel, J., Vaginal microbiome: rethinking health and disease. *Annu Rev Microbiol* **2012,** *66*, 371-89.
145. Elovitz, M. A.; Gajer, P.; Riis, V.; Brown, A. G.; Humphrys, M. S.; Holm, J. B.; Ravel, J., Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. *Nat Commun* **2019,** *10* (1), 1305.
146. Ravel, J.; Moreno, I.; Simon, C., Bacterial vaginosis and its association with infertility, endometritis, and pelvic inflammatory disease. *Am J Obstet Gynecol* **2021,** *224* (3), 251-257.
147. Kumar, N.; Behera, B.; Sagiri, S. S.; Pal, K.; Ray, S. S.; Roy, S., Bacterial vaginosis: Etiology and modalities of treatment-A brief note. *J Pharm Bioallied Sci* **2011,** *3* (4), 496-503.
148. Zheng, N.; Guo, R.; Yao, Y.; Jin, M.; Cheng, Y.; Ling, Z., Lactobacillus iners Is Associated with Vaginal Dysbiosis in Healthy Pregnant Women: A Preliminary Study. *Biomed Res Int* **2019,** *2019*, 6079734.
149. Holm, J. B.; Carter, K. A.; Ravel, J.; Brotman, R. M., Lactobacillus iners and genital health: molecular clues to an enigmatic vaginal species. *Curr Infect Dis Rep* **2023,** *25* (4), 67-75.
150. Zheng, N.; Guo, R.; Wang, J.; Zhou, W.; Ling, Z., Contribution of Lactobacillus iners to Vaginal Health and Diseases: A Systematic Review. *Front Cell Infect Microbiol* **2021,** *11*, 792787.

151. Salam, M. A.; Al-Amin, M. Y.; Salam, M. T.; Pawar, J. S.; Akhter, N.; Rabaan, A. A.; Alqumber, M. A. A., Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. *Healthcare (Basel)* **2023,** *11* (13).
152. Liu, L.; Xu, H. J.; Chen, J. L.; Chen, Z.; Zhan, H. Y.; Xu, D. X.; Chen, Y.; Xu, Z. F.; Chen, D. Z., Detection of Vaginal Metabolite Changes in Premature Rupture of Membrane Patients in Third Trimester Pregnancy: a Prospective Cohort Study. *Reprod Sci* **2021,** *28* (2), 585-594.
153. Srinivasan, S.; Morgan, M. T.; Fiedler, T. L.; Djukovic, D.; Hoffman, N. G.; Raftery, D.; Marrazzo, J. M.; Fredricks, D. N., Metabolic signatures of bacterial vaginosis. *mBio* **2015,** *6* (2).
154. Shalaby, E. A.; Aboul-Enein, A. M.; Sayed, N. M.; Said, M. E.; Hussain, A. R.; Shehata, Y. M.; Sayed, S. R.; Saeed, M. M. H.; Hanafy, E. A., Validation of an alternative quantitative method for determination of antioxidant potential in Rosmarinus officinalis L. and Chilli pepper. *Chemical Papers* **2023,** *78* (1), 275-282.
155. MacFadyen, D. A., On the Mechanism of the Reaction of Ninhydrin with α-Amino Acids. *Journal of Biological Chemistry* **1950,** *186* (1), 1-12.
156. Lamothe, P. J.; McCormick, P. G., Role of hydrindantin in the determination of amino acids using ninhydrin. *Anal Chem* **1973,** *45* (11), 1906-11.
157. Sun, S.-W.; Lin, Y.-C.; Weng, Y.-M.; Chen, M.-J., Efficiency improvements on ninhydrin method for amino acid quantification. *Journal of Food Composition and Analysis* **2006,** *19* (2-3), 112-117.
158. Bottom, C. B.; Hanna, S. S.; Siehr, D. J., Mechanism of the ninhydrin reaction. *Biochemical Education* **1978,** *6* (1), 4-5.
159. Guedes, P. H. G.; Brussasco, J. G.; Moco, A. C. R.; Moraes, D. D.; Flauzino, J. M. R.; Luz, L. F. G.; Almeida, M. T. G.; Soares, M.; Oliveira, R. J.; Madurro, J. M.; Brito-Madurro, A. G., Ninhydrin as a novel DNA hybridization indicator applied to a highly reusable electrochemical genosensor for Candida auris. *Talanta* **2021,** *235*, 122694.
160. Gavrilenko, N. A.; Sukhanov, A. V.; Mokhova, O. V., Redox and acid-base properties of 2,6-dichlorophenolindophenol immobilized on a polymethacrylate matrix. *Journal of Analytical Chemistry* **2010,** *65* (1), 17-20.
161. Borgogna, J. C.; Shardell, M. D.; Grace, S. G.; Santori, E. K.; Americus, B.; Li, Z.; Ulanov, A.; Forney, L.; Nelson, T. M.; Brotman, R. M.; Ravel, J.; Yeoman, C. J., Biogenic Amines Increase the Odds of Bacterial Vaginosis and Affect the Growth of and Lactic Acid Production by Vaginal Lactobacillus spp. *Applied and environmental microbiology* **2021,** *87* (10).
162. McKenzie, R.; Maarsingh, J. D.; Laniewski, P.; Herbst-Kralovetz, M. M., Immunometabolic Analysis of Mobiluncus mulieris and Eggerthella sp. Reveals Novel Insights Into Their Pathogenic Contributions to the Hallmarks of Bacterial Vaginosis. *Front Cell Infect Microbiol* **2021,** *11*, 759697.
163. Hjelm, B. E.; Berta, A. N.; Nickerson, C. A.; Arntzen, C. J.; Herbst-Kralovetz, M. M., Development and characterization of a three-dimensional organotypic human vaginal epithelial cell model. *Biol Reprod* **2010,** *82* (3), 617-27.
164. Laniewski, P.; Gomez, A.; Hire, G.; So, M.; Herbst-Kralovetz, M. M., Human Three-Dimensional Endometrial Epithelial Cell Model To Study Host Interactions with Vaginal Bacteria and Neisseria gonorrhoeae. *Infect Immun* **2017,** *85* (3).

165. Ayehunie, S.; Cannon, C.; Lamore, S.; Kubilus, J.; Anderson, D. J.; Pudney, J.; Klausner, M., Organotypic human vaginal-ectocervical tissue model for irritation studies of spermicides, microbicides, and feminine-care products. *Toxicol In Vitro* **2006,** *20* (5), 689-98.
166. Merbah, M.; Introini, A.; Fitzgerald, W.; Grivel, J. C.; Lisco, A.; Vanpouille, C.; Margolis, L., Cervico-vaginal tissue ex vivo as a model to study early events in HIV-1 infection. *Am J Reprod Immunol* **2011,** *65* (3), 268-78.
167. Breshears, L. M.; Edwards, V. L.; Ravel, J.; Peterson, M. L., Lactobacillus crispatus inhibits growth of Gardnerella vaginalis and Neisseria gonorrhoeae on a porcine vaginal mucosa model. *BMC Microbiol* **2015,** *15*, 276.
168. Mahajan, G.; Doherty, E.; To, T.; Sutherland, A.; Grant, J.; Junaid, A.; Gulati, A.; LoGrande, N.; Izadifar, Z.; Timilsina, S. S.; Horvath, V.; Plebani, R.; France, M.; Hood-Pishchany, I.; Rakoff-Nahoum, S.; Kwon, D. S.; Goyal, G.; Prantil-Baun, R.; Ravel, J.; Ingber, D. E., Vaginal microbiome-host interactions modeled in a human vagina-on-a-chip. *Microbiome* **2022,** *10* (1), 201.
169. Zierden, H. C.; DeLong, K.; Zulfiqar, F.; Ortiz, J. O.; Laney, V.; Bensouda, S.; Hernandez, N.; Hoang, T. M.; Lai, S. K.; Hanes, J.; Burke, A. E.; Ensign, L. M., Cervicovaginal mucus barrier properties during pregnancy are impacted by the vaginal microbiome. *Front Cell Infect Microbiol* **2023,** *13*, 1015625.
170. Bewick, V.; Cheek, L.; Ball, J., Statistics review 14: Logistic regression. *Crit Care* **2005,** *9* (1), 112-8.
171. Atassi, F.; Pho Viet Ahn, D. L.; Lievin-Le Moal, V., Diverse Expression of Antimicrobial Activities Against Bacterial Vaginosis and Urinary Tract Infection Pathogens by Cervicovaginal Microbiota Strains of Lactobacillus gasseri and Lactobacillus crispatus. *Front Microbiol* **2019,** *10*, 2900.
172. Lagenaur, L. A.; Hemmerling, A.; Chiu, C.; Miller, S.; Lee, P. P.; Cohen, C. R.; Parks, T. P., Connecting the Dots: Translating the Vaginal Microbiome Into a Drug. *J Infect Dis* **2021,** *223* (12 Suppl 2), S296-S306.
173. Thuan, N. H.; Tatipamula, V. B.; Canh, N. X.; Van Giang, N., Recent advances in microbial co-culture for production of value-added compounds. *3 Biotech* **2022,** *12* (5), 115.
174. Jeong, Y.; Kong, W.; Lu, T.; Irudayaraj, J., Soft hydrogel-shell confinement systems as bacteria-based bioactuators and biosensors. *Biosensors and Bioelectronics* **2023,** *219*.
175. Moya-Ramirez, I.; Kotidis, P.; Marbiah, M.; Kim, J.; Kontoravdi, C.; Polizzi, K., Polymer Encapsulation of Bacterial Biosensors Enables Coculture with Mammalian Cells. *ACS Synth Biol* **2022,** *11* (3), 1303-1312.
176. Che, S.; Men, Y., Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. *J Ind Microbiol Biotechnol* **2019,** *46* (9-10), 1343-1358.
177. Brunel, C.; Pouteau, R.; Dawson, W.; Pester, M.; Ramirez, K. S.; van Kleunen, M., Towards Unraveling Macroecological Patterns in Rhizosphere Microbiomes. *Trends Plant Sci* **2020,** *25* (10), 1017-1029.
178. Ajilogba, C. F.; Olanrewaju, O. S.; Babalola, O. O., Plant Growth Stage Drives the Temporal and Spatial Dynamics of the Bacterial Microbiome in the Rhizosphere of Vigna subterranea. *Front Microbiol* **2022,** *13*, 825377.
179. Duncan, K.; Carey-Ewend, K.; Vaishnava, S., Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer. *Gut Microbes* **2021,** *13* (1), 1874815.

180. Liu, W.; Jacquiod, S.; Brejnrod, A.; Russel, J.; Burmolle, M.; Sorensen, S. J., Deciphering links between bacterial interactions and spatial organization in multispecies biofilms. *ISME J* **2019,** *13* (12), 3054-3066.
181. Hernandez-Valdes, J. A.; Zhou, L.; de Vries, M. P.; Kuipers, O. P., Impact of spatial proximity on territoriality among human skin bacteria. *NPJ Biofilms Microbiomes* **2020,** *6* (1), 30.
182. Motabar, D.; Li, J.; Payne, G. F.; Bentley, W. E., Mediated electrochemistry for redox-based biological targeting: entangling sensing and actuation for maximizing information transfer. *Curr Opin Biotechnol* **2021,** *71*, 137-144.
183. Rathore, A. S.; Nikita, S.; Thakur, G.; Mishra, S., Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. *Trends Biotechnol* **2023,** *41* (4), 497-510.
184. Kim, J. H.; Suh, Y. J.; Park, D.; Yim, H.; Kim, H.; Kim, H. J.; Yoon, D. S.; Hwang, K. S., Technological advances in electrochemical biosensors for the detection of disease biomarkers. *Biomed Eng Lett* **2021,** *11* (4), 309-334.
185. Oliver, N. S.; Toumazou, C.; Cass, A. E.; Johnston, D. G., Glucose sensors: a review of current and emerging technology. *Diabet Med* **2009,** *26* (3), 197-210.
186. Teymourian, H.; Barfidokht, A.; Wang, J., Electrochemical glucose sensors in diabetes management: an updated review (2010-2020). *Chem Soc Rev* **2020,** *49* (21), 7671-7709.
187. Aziz, S.; Ahmed, S.; Alouini, M. S., ECG-based machine-learning algorithms for heartbeat classification. *Sci Rep* **2021,** *11* (1), 18738.
188. Prajapati, H. B.; Dabhi, V. K.; Jambukia, S. H., Classification of ECG signals using Machine Learning Techniques: A Survey. *International Conference on Advances in Computer Engineering and Applications (ICACEA)* **2015**.
189. Wang, L.; Hashimoto, Y.; Tsao, C. Y.; Valdes, J. J.; Bentley, W. E., Cyclic AMP (cAMP) and cAMP receptor protein influence both synthesis and uptake of extracellular autoinducer 2 in Escherichia coli. *Journal of bacteriology* **2005,** *187* (6), 2066-76.
190. Wang, S., Tsao, C.Y., Motabar, D., Li, J., Payne, G.F., Bentley, W.E., A Redox-based Autoinduction Strategy to Facilitate Expression of 5xCys-tagged Proteins for Electrobiofabrication. *Frontiers in Microbiology* **2021,** *12*, 1473.