Modeling Phenotypic Differentiation in B. Subtilis Biofilms Via an Environmentally-Regulated Genetic Network

by

Andira Putri

Department of Mathematics and Statistics
College of Arts and Sciences
Georgia State University
2018

Research Initiatives in Mathematics, Mathematics Education, and Statistics Program

Research Professor: Dr. Yi Jiang

Table of Contents

Section Title	**Page**
Abstract	1
Introduction	2
Network Motifs	3
Genetic Network	6
Dynamics of Type 2 Incoherent FFL	7
Mathematical Model	9
Methods of Dynamical Analysis	12
Qualitative Studies of Nested FFL	13
Main Python Code for Quantitative Analysis	14
Steady State Concentrations	15
Discussion	18
Parameter Analysis	19
Conclusion and Future Work	22
Dedications	24
Reflection of RIMMES Program	24
References	26

List of Figures

Figure	**Page**
1. Biofilm on the surface of a tooth.	3
2. Transcription network as directed graphs.	4
3. A simple feed-forward loop.	5
4. The eight feed-forward loops.	5
5. Pathway leading to RapA with AND-gate logic.	6
6. Pathway leading to Spo0A~P with AND-gate logic.	6
7. Complete genetic network leading to phenotypic differentiation in B. subtilis.	7
8. A simple transcription network.	8
9. Pulse-generating dynamics and response to S_Y.	9
10. Our simplified genetic network.	10
11. Qualitative study of our genetic network.	13
12. Nutrients ON and quorum sensing ON.	17
13. Nutrients ON and quorum sensing OFF.	17
14. Nutrients OFF and quorum sensing ON.	17
15. Nutrients OFF and quorum sensing OFF.	18
16. Graduated plot of S when varying parameter ‘a’ over an interval [0,5].	20
17. Visual of elasticity value.	21

List of Tables

Table	**Page**
1. Abbreviated parameters and variables.	9
2. Meanings and values of equation parameters.	12
3. Initial values of variables.	12
4. Expectations of final Spo0A~P concentrations.	16
5. Summary of steady state analysis focusing on Spo0A~P final concentration.	16

Abstract

Biofilms are colonies of bacteria bounded by a polymeric extracellular matrix, and they are prevalent in environmental, industrial, and health settings [12]. Much of the molecular mechanisms for biofilm growth and formation are still unclear, in particular how cells switch phenotypes, i.e. their cell type. To understand this switch, we focus on the Bacillus subtilis (B. subtilis) strain. Experimental evidence shows that phosphorylated Spo0A (Spo0A~P) is the master regulator determining cell phenotype [21]. The genetic network of B. subtilis leading to Spo0A~P production is affected by environmental factors, specifically nutrients and quorum sensing [17]. We construct a nested incoherent feed-forward loop to represent the genetic network of B. subtilis, integrating both nutrient and quorum sensing pathways. A system of coupled ordinary differential equations models our genetic network, and dynamical systems analysis reveals that starvation raises Spo0A~P levels and initiates sporulation. This simple model of the B. subtilis genetic network is the first step in understanding the growth mechanisms of biofilms in response to environmental conditions.

Introduction

When bacteria accumulate on a surface and enclose themselves in a polymeric extracellular matrix (ECM), they form a biofilm [12]. Biofilms can adhere to a variety of surfaces, such as water pipelines, medical tools, and body tissues. A plethora of illnesses caused by bacterial biofilms are difficult to treat with standard antibiotics because of their outer polymeric layer [11]. Biofilms are also of concern with regards to the environment and biotechnology development. Unfortunately, biofilms are hard to eradicate with antibiotics and regular cleaning agents alone since they are genetically equipped to adapt in harsh environments [13]. Biofilms adapt to adversity by changing their phenotype, i.e. their cell type [14]. Because of biofilms' presence in many aspects of our daily lives, understanding its phenotypic adaptations and growth mechanisms is of high importance.

Bacteria in B. subtilis biofilms react to nutrient availability and quorum sensing, i.e. surrounding cell population density, by switching their phenotype between sporulating or producing ECM polymers [18,20]. Phosphorylated Spo0A, abbreviated Spo0A~P, is the main regulator of phenotypic differentiation in B. subtilis [21]. When Spo0A~P levels are above a certain threshold level, the cells sporulate which is expected under high stress, i.e. low food availability and high competition for resources [17,21]. When the cells have ample nutrients and no competition, Spo0A~P concentrations are lower, and bacteria assume the matrix-producing phenotype [17,21]. This project seeks to understand the growth mechanisms of biofilms through sporulation by modeling the B. subtilis genetic network and observing Spo0A~P concentrations in response to environmental factors.

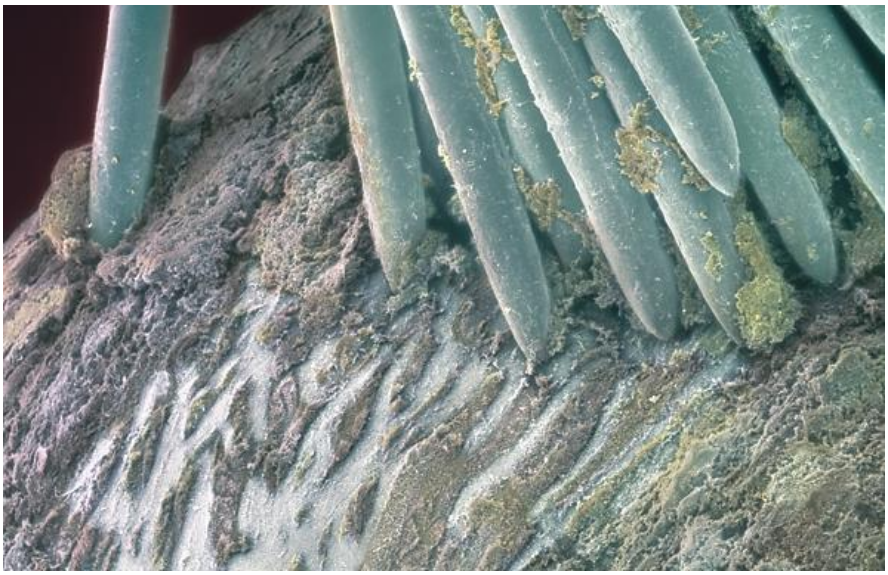

Figure 1. Biofilm on the surface of a tooth. Plaque is a common example of biofilms. Bacteria collect on the surface of a tooth and create a slimy, polymeric surface that is difficult to remove. Adapted from [9].

Network Motifs

Before presenting the genetic network of B. subtilis, we must understand network motifs. Cells contain transcription networks, and these networks act as genetic information-processing systems [2]. Transcription networks can be affected by nutrients and environmental stress. Within these networks are transcription factors which can either activate or repress (inhibit) gene expression. If a gene A activates a gene B, more of A increases the production of B. If a gene A represses a gene B, more of A decreases the production of B. One can visualize transcription networks as directed graphs, where nodes represent transcription factors and edges represent the interaction between molecules. The directed graphs form patterns, and some of them occur more frequently than others in a genetic network. Such patterns are called network motifs. Because of the significant frequency of these patterns, scientists believe network motifs give insight on the network dynamics of biological systems [2]. One of the most significant network motifs is the feed-forward loop (FFL) [1].

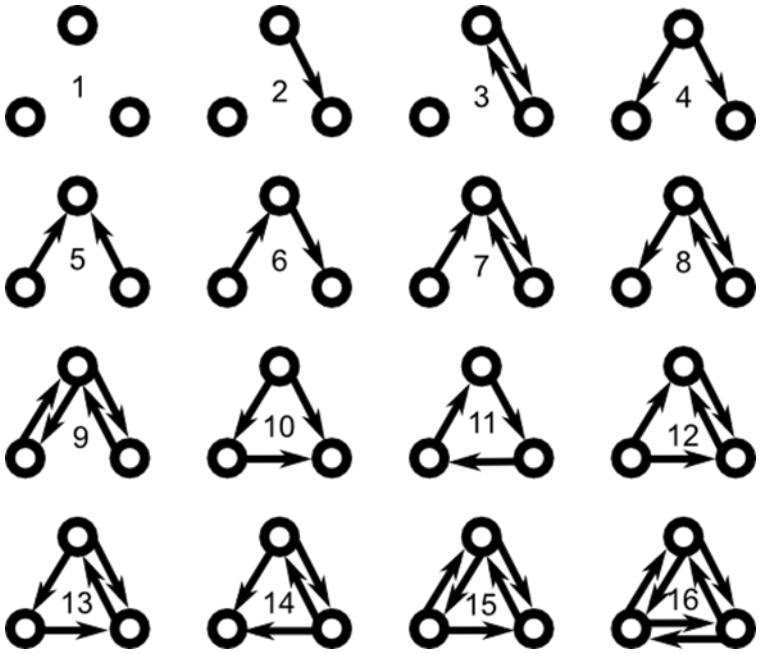

Figure 2. Transcription networks as directed graphs. The above directed graphs are some of the most frequently observed patterns in biological systems with three nodes, particularly in yeast and Escherichia coli (E. coli). Adapted from [10].

The simplest structure of a FFL, shown in Figure 3, consists of a target gene Z and two transcription factors X and Y [1,2]. X regulates Y, but both X and Y regulate Z. Environmental factors, depicted as S_X and S_Y in Figure 3, can also affect the expression of Z. A logic gate determines the outcome of Z based on input signals from X and Y. Such gates represent the cis-regulatory module functions on the gene Z within DNA [2]. A biological circuit can either have OR-gate logic or AND-gate logic. With OR logic, only one transcription factor, X or Y, needs to be expressed above a threshold value to have a significant effect on Z. However, AND logic requires both X and Y to be above a threshold concentration. There are two types of FFLs: coherent and incoherent. In coherent FFLs, the input from the direct path X → Z is the same as the indirect path X → Y → Z. Oppositely, the inputs from the direct and indirect paths are not the same for incoherent FFLs. A complete table of coherent and incoherent FFLs is shown in Figure 4.

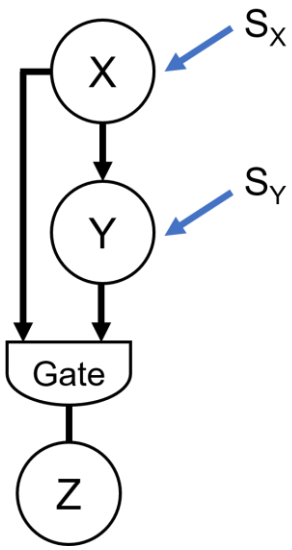

Figure 3. A simple feed-forward loop. In this circuit, X activates Y, but both X and Y activate Z. There are two environmental stimuli, S_X and S_Y, that may influence Z. The overall effect of the two pathways X → Y → Z and X → Z is modulated by a logical gate. Note that all relationships here are activations- however, inhibition relationships are possible.

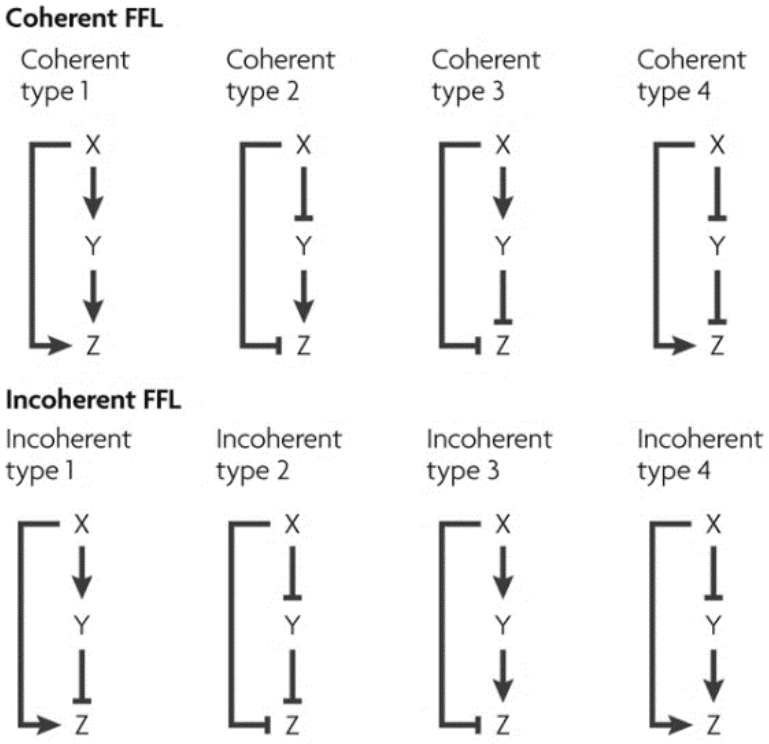

Figure 4. The eight feed-forward loops. Adapted from [2].

Genetic Network

We first consolidated several B. subtilis genetic networks from literature to create a unique system that is concise to our research problem [6,7,8]. Our genetic network has two main pathways with AND-gate logic, shown in Figures 5 and 6. The two pathways take the form of the Type 2 incoherent FFL shown above in Figure 4. To construct our main genetic network, we interlock the two pathways, and we get the nested incoherent FFL shown in Figure 7.

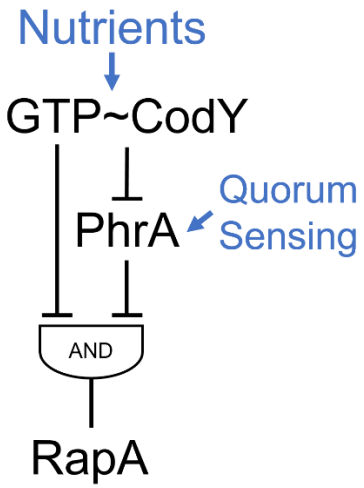

Figure 5. Pathway leading to RapA with AND-gate logic.

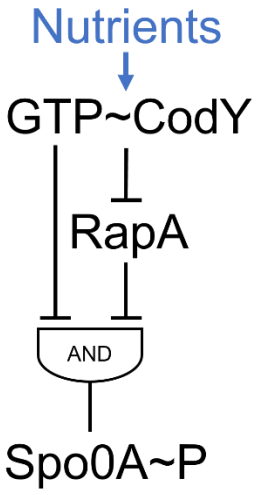

Figure 6. Pathway leading to Spo0A~P with AND-gate logic.

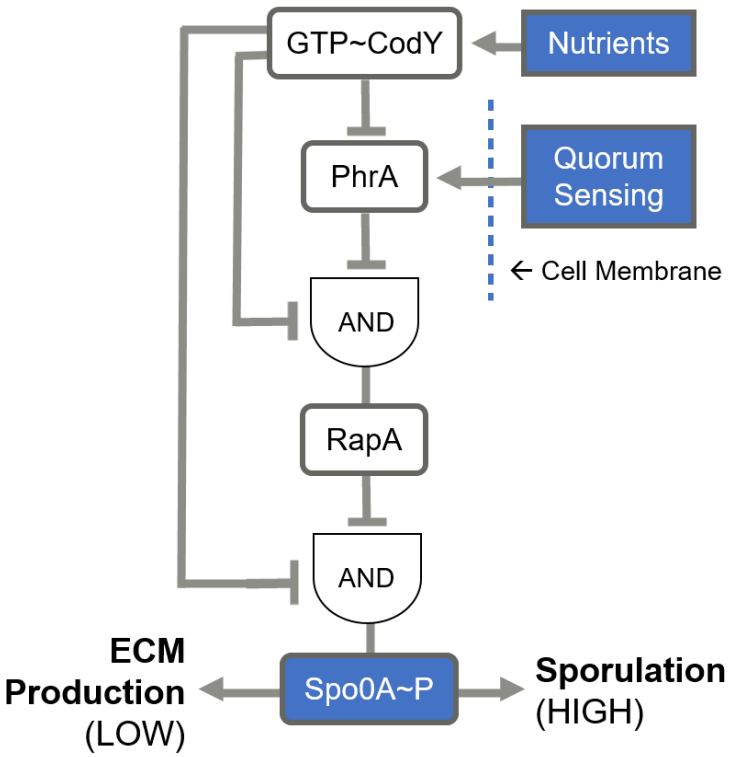

Figure 7. Complete genetic network leading to phenotypic differentiation in B. subtilis. The two main pathways from Figures 5 and 6 are joined together to create an environmentally-regulated nested incoherent FFL. Spo0A~P concentrations and, consequently, cell phenotype depend on the interactions of all transcription factors and environmental inputs.

Dynamics of Type 2 Incoherent FFL

Since our focus is on the Type 2 incoherent FFLs, understanding its elementary dynamics is critical for our study. The Type 2 incoherent FFL is notable for being pulse generators and sign sensitive accelerators [2]. In this section, we make comparisons between a Type 2 incoherent FFL and a simple transcription network, shown in Figure 8.

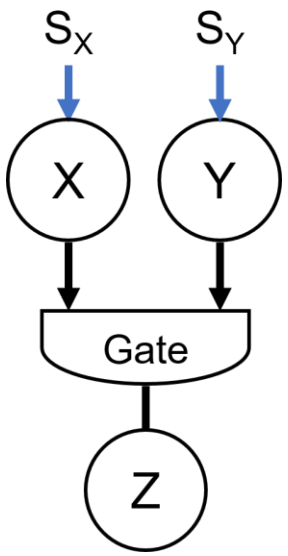

Figure 8. A simple transcription network. Like a typical FFL, there are three nodes in the biological system. However, there are two paths X → Z and Y → Z. X has no regulation on Y.

All incoherent FFLs are pulse generators in general [1,2]. When an environmental input S_X is ON, target gene Z is expressed by both X and Y at the beginning time step. However, in the Type 2 incoherent FFL, the direct repression relationship of Z from X decreases the production rate of Z over time. This is due to S_X activating X in the same time frame; with high expression of X, there is low expression of Z. The resulting dynamics is a pulse, as shown in Figure 9. The Type 2 incoherent FFL is considered a weak pulse generator compared to its Type 3 and 4 incoherent counterparts due to a lower amplitude when S_X and S_Y are both ON. The lower amplitude could be because the Type 2 incoherent FFL reacts to the environmental input S_Y. That is, as Y is activated by S_Y, a higher expression of Y inhibits Z at the same time X represses Z.

All incoherent FFLs are sign-sensitive accelerators; the Type 2 loop accelerates when the environmental input S_X is off [2]. The explanation is essentially the same as the pulse-generating dynamics. Without the expression X from S_X, Z rises quickly due to the direct repression relationship of X and Z. However, low expression of X causes a high expression of Y. As production of Y increases, the level of Z starts to decrease as well.

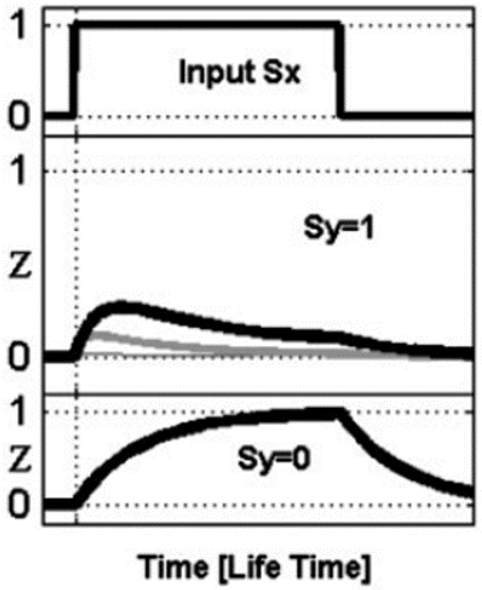

Figure 9. Pulse-generating dynamics and response to S_Y. The upper subplot shows the activation of S_X. 0 means S_X is OFF, and 1 means S_X is ON. In the middle sublot, a pulse is generated when S_X and S_Y are both ON. The lower subplot shows the Type 2 incoherent FFL's pulse-generating dynamics when S_Y is OFF. Adapted from [2].

Mathematical Model

Before we introduce the mathematical model, we revisit our overall genetic network. The network is shown in Figure 10, and the environmental signals and variables are renamed for simplicity. Table 1 shows the abbreviations.

Parameter or Variable	Network Component
N	Nutrient signal
Q	Quorum sensing signal
C	GTP~CodY complex
P	PhrA
R	RapA
S	Spo0A~P

Table 1. Abbreviated parameters and variables. The parameters and variables in this table are used for Figure 10, the equations, and dynamical analysis.

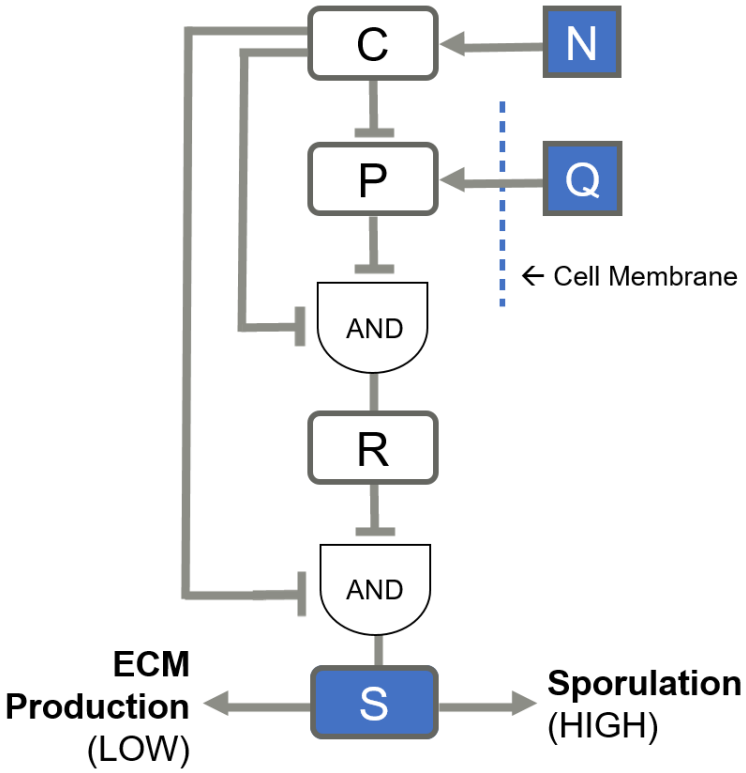

Figure 10. Our simplified genetic network. This genetic network serves as a basis for our mathematical model.

In the following model, we assume that the input of nutrients and quorum sensing signals are constant, and the environmental stimuli are inducers. To create the equations, we employ a simple template for some gene A:

$$\dot{A} = N \text{ or } Q + production\ of\ A - degradation\ of\ A \qquad (1).$$

In the above equation, we only include N and Q in the equations for C and P, respectively, since they are the transcription factors that are directly affected by the environmental stimuli.

All relationships between transcription factors are inhibition, so we use the Hill function to model the production of genes that are repressed. The Hill function is a notable equation in biochemistry and biophysics. Hill functions model the interaction between a ligand and a receptor,

and it quantitatively describes the binding abilities of the two molecules [3]. Repression of a gene A by some gene B is modeled by the following Hill function:

$$Repression = \frac{k^n}{k^n + B^n} \quad (2).$$

For a gene to be considered ON or at a high concentration, its levels must be above a threshold level 'k.' The parameter 'k' also represents the binding coefficient, and its value is unique to a certain ligand-receptor pair. The Hill coefficient, represented by 'n,' defines the binding cooperativity of molecules. When n = 1, binding is completely independent. That is, there is no interaction between ligand binding sites. In the case when n > 1, this denotes positive cooperativity, and the binding of one ligand accelerates overall affinity of subsequent ligands. When 0 < n < 1, this is negative cooperativity, and the binding of one ligand inhibits overall affinity of subsequent ligands [3,4].

Degradation is a naturally random and complex process. For simplicity, we generalize degradation by assuming that the decay of some gene A at a constant degradation rate α is:

$$Degradation = \alpha A \quad (3).$$

With all formulas, we can now create our system of differential equations. Note that $\dot{A}$ for some gene A is equivalent to $\frac{dA}{dt}$.

$$\dot{C} = N - \alpha_C C \quad (4),$$

$$\dot{P} = Q + \frac{k_1^n}{k_1^n + C^n} - \alpha_P P \quad (5),$$

$$\dot{R} = \frac{k_2^n}{k_2^n + C^n} + \frac{k_3^n}{k_3^n + P^n} - \alpha_R R \quad (6),$$

$$\dot{S} = \frac{k_4^n}{k_4^n + C^n} + \frac{k_5^n}{k_5^n + R^n} - \alpha_S S \quad (7).$$

All parameters and their corresponding values are defined below in Table 2. We also defined initial concentration of variables for quantitative analysis in Table 3.

Parameter	Meaning	Value
$\boldsymbol{\alpha_C}$	C degradation rate	0.7
$\boldsymbol{\alpha_P}$	P degradation rate	6
$\boldsymbol{\alpha_R}$	R degradation rate	4
$\boldsymbol{\alpha_S}$	S degradation rate	2.78
$\boldsymbol{k_1}$	C and P binding coefficient	1.67
$\boldsymbol{k_2}$	C and R binding coefficient	5
$\boldsymbol{k_3}$	P and R binding coefficient	3.3
$\boldsymbol{k_4}$	C and S binding coefficient	1
$\boldsymbol{k_5}$	R and S binding coefficient	0.58
$\boldsymbol{n}$	Hill coefficient	2
N	Nutrient signal	1 (OFF) or 10 (ON)
Q	Quorum sensing signal	1 (OFF) or 10 (ON)

Table 2. Meanings and values of equation parameters. The values of parameters remain the same throughout the dynamical analysis. Refer to Table 1 for variable definitions.

Variable	Initial Value
C	0.1
P	0.1
R	0.1
S	0.1

Table 3. Initial values of variables. N and Q initial values of 1 denote a low environmental signal. Otherwise, a value of 10 marks a high environmental impact on the genetic network from N and/or Q. Refer to Table 1 for variable definitions.

Methods of Dynamical Analysis

To determine which phenotype the cells switch to, we use two methods of analysis. First, we conduct a qualitative study of the biological circuit. Essentially, we infer the dynamics of all genes in the network over time given the relationships between transcription factors. By performing a qualitative study of the network, we have an idea of what dynamics to expect from

the mathematical model. Next, we perform a quantitative study, specifically dynamical systems analysis. We create a code for the model using Tellurium, a Python-based biochemical systems software.

Qualitative Studies of Nested Feed-forward Loop

We predict the dynamics of C, P, R, and S given constant inputs of nutrients and quorum sensing signals. The qualitative study holds for any combination of 1-OFF or 10-ON for N and Q. The graphs are shown in Figure 11.

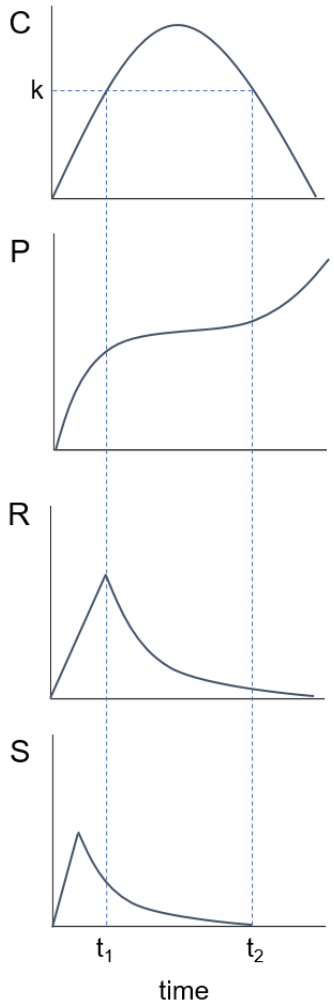

Figure 11. Qualitative study of our genetic network. Dynamics include biphasic behavior in P and pulse generation in R and S. Further elaboration is continued below.

C exhibits negative parabolic behavior because the constant input of nutrients catalyzes C's production. However, as C levels build, its decay increases which causes C to start decreasing. Now, we consider P. Initially, P levels rise because Q activates its production. When C surpasses a threshold concentration 'k' at time t_1, P production slows because C inhibits P. At time t_2, C concentrations fall below the threshold concentration, so P begins to rise rapidly again. The switch between high and low production suggests that P exhibits biphasic behaviors. Next, we study R. In the beginning, R rises rapidly because there is not enough of C to repress its production. However, as C (and eventually P) levels rise above the threshold concentration, R's production decreases rapidly; this creates a pulse. S behaves similarly to R in that it is also exhibits pulsing behavior. However, the initial below-threshold concentrations of C, P, and R accelerate the production S, which explains why S peaks earlier than R. This qualitative study provides a basis for comparison with the quantitative dynamical systems analysis.

Main Python Code for Quantitative Analysis

The master code is shown below; it contains the equations of the mathematical model as well as the initial concentrations of the genes and the parameter values. The primary function of the master code is to initialize the model and plot all the variables in one graph. Additional lines of codes are added throughout the paper for specific methods of analysis; they are appended to the end of the main code.

```
# -*- coding: utf-8 -*-
"""
Created on Wed Jun 28 13:22:15 2017

@author: Andira

"""
import tellurium as te
import roadrunner as r
from tellurium import ParameterScan as p

r = te.loada('''
```

```
    model pathway()

  $X -> C; N - a*C
  $X -> P; Q + (k1^n)/(k1^n+C^n) - b*P
  $X -> R; (k2^n)/(k2^n+P^n) * (k3^n)/(k3^n+C^n) - c*R
  $X -> S; (k4^n)/(k4^n+R^n) * (k5^n)/(k5^n+C^n) - d*S

C=0.1
P=0.1
R=0.1
S=0.1
a=0.7
k1=1.67
c=6
k2=5
k3=3.3
f=4
k4=1
k5=0.58
i=2.78
n=2
N=1
Q=10 end
''')

res = r.simulate(0, 5, 200, ['TIME','C','P','R', 'S'])
r.plot()
```

Steady State Concentrations

Our main method from dynamical systems is steady state analysis. A steady state is reached when the rate of change of gene expression is 0, i.e. $\dot{A} = 0$ for some gene A [3]. We consider the steady state value of a gene to be its final concentration in the network, and this is important for Spo0A~P specifically. By observing the final concentration of Spo0A~P, we can determine which phenotype the cells switch to depending on the threshold value. We also graph the variables over time to visually observe the dynamics. Finally, we calculate eigenvalues to ensure system stability. We are mainly checking that the eigenvalues are negative for each variable.

In Table 3, we set the following expectations for Spo0A~P steady state values based on biological background information and qualitative studies.

Nutrient	Quorum Sensing	Spo0A~P
ON	ON	Low
ON	OFF	Low
OFF	ON	High
OFF	OFF	High

Table 4. Expectations of final Spo0A~P concentrations. When Spo0A~P is low, this denotes a concentration below the threshold value, and the cells have the matrix-producing phenotype. Otherwise, Spo0A~P is above the threshold, i.e. high, and the cells sporulate.

To perform steady state analysis in Tellurium, we add the following code to the master:

```
print r.getSteadyStateValues()
print r.getFullEigenValues()
```

For all observations, we have the eigenvalue matrix [-0.7, -6, -4, -2.78], which shows the eigenvalues for C, P, R, and S, respectively. Because all eigenvalues are negative, we can assure that our biological system is stable and any small perturbations cannot drastically change the dynamics.

The results of the steady state analysis are shown below. Table 5 summarizes the results of the analysis, and Figures 12, 13, 14, and 15 visualize the network dynamics with the y-axis representing concentrations. In the following four figures, the right sublot removes the curves for C and P. The concentrations of C and P overpowers R and S in some cases, and it causes the Tellurium program to zoom out too much on the graph. By removing the C and P curves, we get a better observation of R and S dynamics.

Nutrient	Quorum Sensing	Spo0A~P Concentration
ON	ON	5.9188e-04
ON	OFF	5.9187e-04
OFF	ON	0.049178
OFF	OFF	0.048754

Table 5. Summary of steady state analysis focusing on Spo0A~P final concentration. The final concentration is calculated when the rate of change of S is zero and concentration can no longer change.

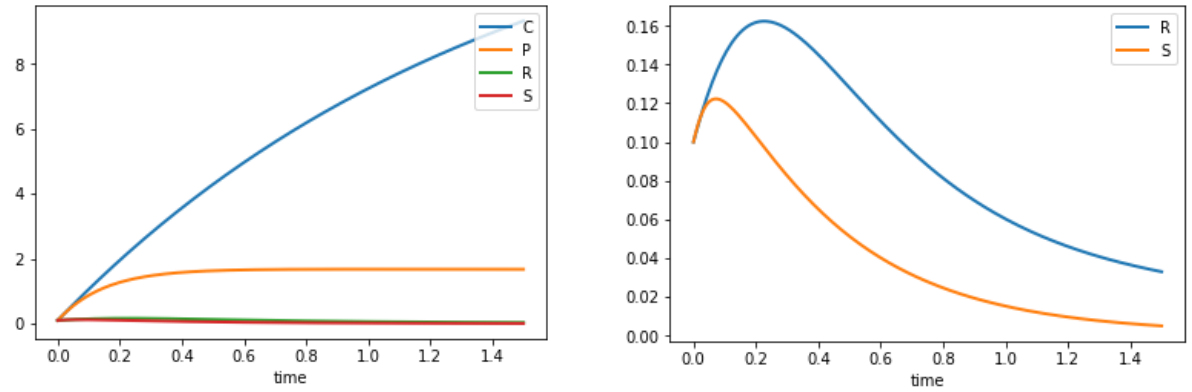

Figure 12. Nutrients ON and quorum sensing ON.

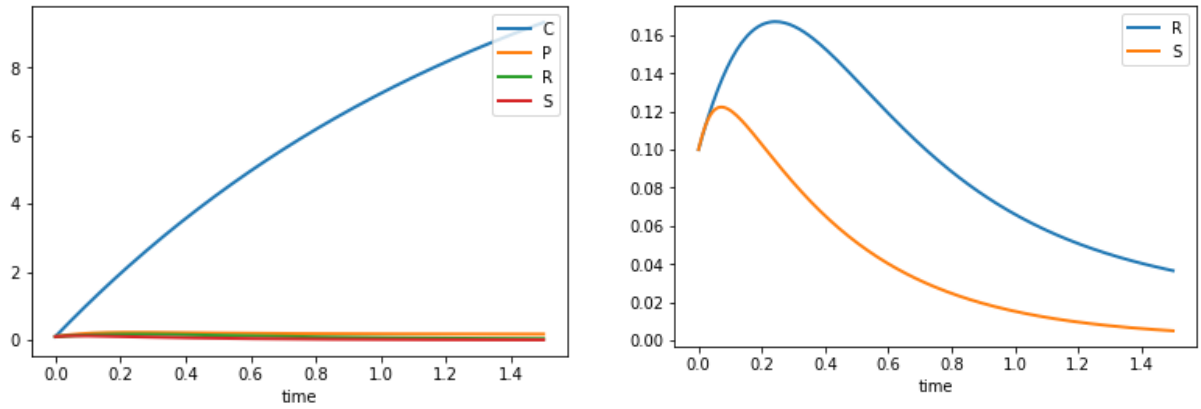

Figure 13. Nutrients ON and quorum sensing OFF.

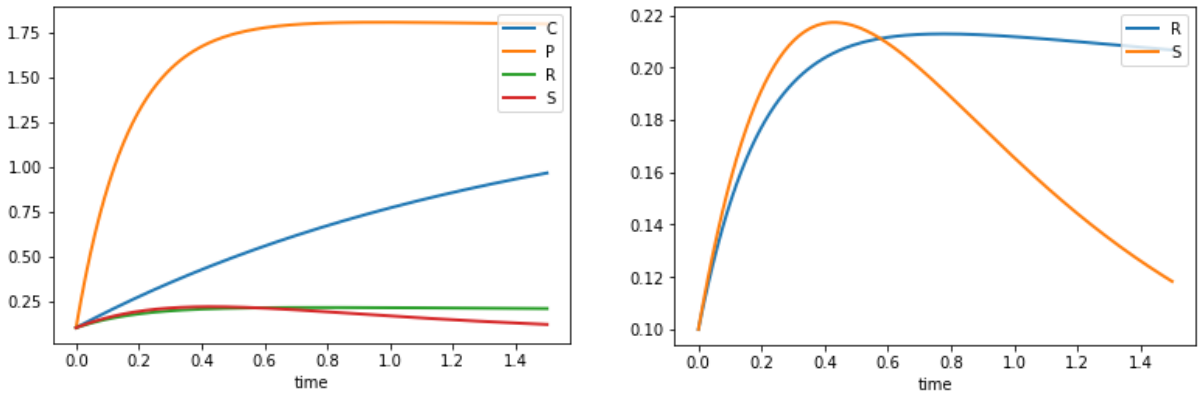

Figure 14. Nutrients OFF and quorum sensing ON.

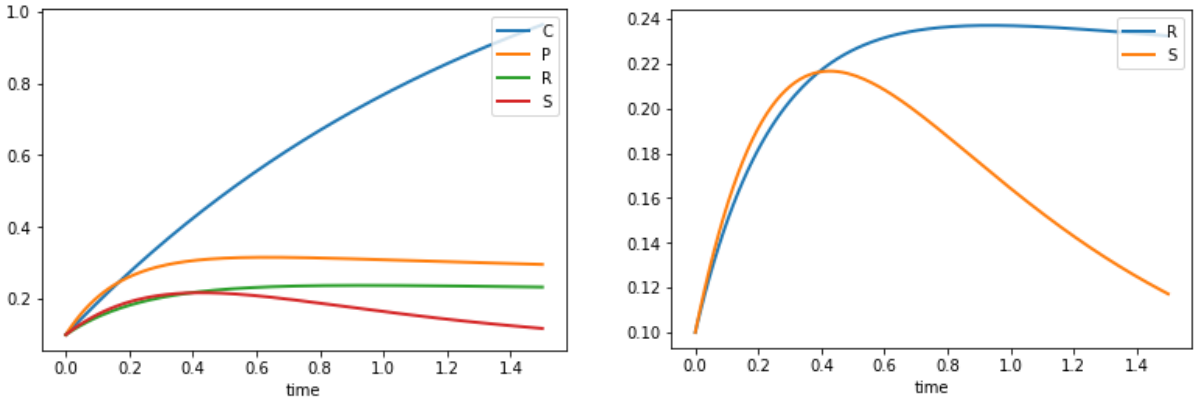

Figure 15. Nutrients OFF and quorum sensing OFF.

Discussion

Based on the results of the steady state analysis, we have that S concentrations are higher when nutrients are OFF. The highest S concentration, 0.049178, occurs when nutrients are OFF and quorum sensing is ON. In this scenario, the cells undergo the highest stress because of competition for resources, i.e. low nutrient availability and a high surrounding cell population density. The result suggests that under high stress, cells in B. subtilis biofilms produce high levels of Spo0A~P, and they switch to the sporulation phenotype.

Now, we discuss the dynamics of our nested Type 2 incoherent FFL based on the steady state analysis. The right sublots of Figures 12-15 show that R and S exhibit most dynamics of an elementary Type 2 incoherent FFL. R and S show pulsing dynamics in all cases. With regards to sign-sensitive acceleration, R and S seem to react quicker when nutrients are ON. When nutrients are ON, R and S reach their peak concentrations at 0.3 and 0.1 time steps, respectively. At the OFF step, R and S reach their peak levels at 0.6 and 0.2 time steps, respectively. The results suggest that with our given parameter set, reaction rate doubles when nutrients are ON; this contradicts the dynamics of an elementary Type 2 incoherent FFL.

Finally, we compare the qualitative study and steady state analysis dynamics. R and S are similar in both studies because they are both pulsing. In addition, S peaks earlier than R in both

studies. The main difference between the qualitative and quantitative studies is the behavior of C and P. With our steady state analysis plots, C does not begin to decrease at any time step, and P does not show biphasic behavior. However, given the dynamics of C and P in the steady state analysis, R and S are still pulsing. We considered the time constraints of the graph (0-1.5 time steps) to be a reason why we were not able to see C and P exhibiting behaviors from Figure 11. When we change the constraint to 0-100 time steps, the qualitative study dynamics were still not showing. This could potentially be an issue involving our parameters, so further testing on parameters may be necessary.

Parameter Analysis

Parameter determination proves to be the most tedious and time-consuming task of the biofilm project. To find the best parameters to use, we perform parameter scans and metabolic control analysis. In the following codes, we have nutrients OFF and quorum sensing ON, and we use the same parameter and variable values defined in Tables 2 and 3.

Parameter scanning is a quick way to graphically see the effects of changing a parameter over a defined interval. To perform a parameter scan, we append the following lines to the master code:

```
p = ps(r)

p.endTime = 1.5
p.numberOfPoints = 200
p.polyNumber = 10
p.startValue = 0
p.endValue = 5
#parameter to test
p.value = 'a'
#variable you want to see affected by changing parameter
p.selection = ['S']
p.plotGraduatedArray()
```

The code above varies parameter ‘a,’ the degradation rate of C, over an interval [0,5]. 10 values of ‘a’ are generated, and the values are linearly-spaced over the specified interval. Then, the code

plots 10 curves of S from 0 to 1.5 time steps, with each curve corresponding to a different value of ‘a.’ The output of the parameter scan is shown below in Figure 16.

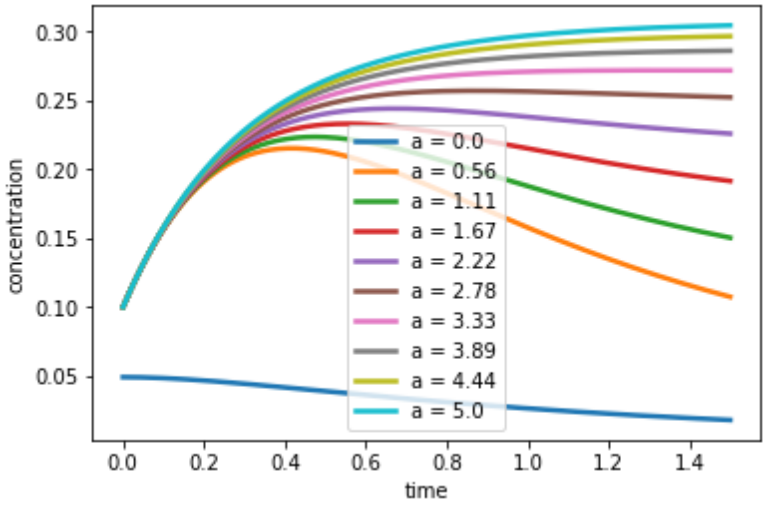

Figure 16. Graduated plot of S when varying parameter ‘a’ over an interval [0,5]. The parameter scan plot suggests that a higher degradation rate of C would increase the steady state concentration of S.

Of course, the code can be changed to test any parameter and variable in the system. Parameter scans are quick and efficient for determining values to use for the model. However, parameter scans are sometimes not informative enough because they just give us a visual of the network dynamics when we change one parameter.

Metabolic control analysis, a form of sensitivity analysis, is a more technical and elaborate method of assessing parameter values. Sensitivity analysis observes how small perturbations in a parameter set can affect an entire system [5]. For local sensitivity, we analyze elasticity matrices. For global sensitivity, we analyze concentration coefficient matrices.

Elasticity can be described by:

$$\varepsilon_S^v = \frac{\partial \ln v}{\partial \ln S} = \frac{\partial v}{\partial F}\frac{S}{v} \quad (8).$$

Elasticity describes how a transcription factor S changes a reaction rate v within one time step. Elasticity can be viewed as a slope, as depicted in Figure 17. Equation 8 defines elasticity as a partial derivative, so the only transcription factor in the system that is changing is the one being perturbed [5]. Elasticities with the greatest absolute value are the most interesting since the transcription factor associated with it can lead to the greatest system change.

Concentration coefficients are like elasticities. However, while elasticities consider an entire reaction, concentration coefficients describe how a transcription factor S changes when just one parameter P is perturbed. Concentration coefficients are given by:

$$C_P^S = \frac{dS}{dP}\frac{P}{S} = \frac{d\ln S}{d\ln P} \quad (8).$$

Larger concentration coefficients (in magnitude) suggest that a transcription factor is sensitive to a certain parameter, and changing the parameter may have a significant global effect on the network dynamics.

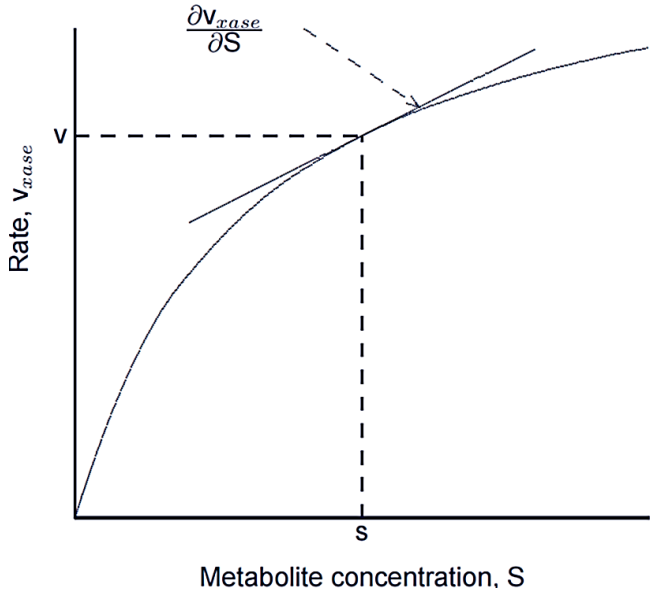

Figure 17. Visual of elasticity value. Elasticity is a partial derivative of reaction rate with respect to metabolite, i.e. transcription factor, concentration. The slope of the tangent line represents the elasticity value.

The sensitivity analysis can be easily implemented in Tellurium. To perform the procedure, we append the following line to the master code:

```
print r.getScaledElasticityMatrix()
print r.getCC('P','Q')
print r.getCC('P','c')
```

Our output is:

```
                C,        P,        R,        S
_J0 [[ -3606.39,        0,        0,        0],
_J1  [ -22.0291, -8684.14,        0,        0],
_J2  [ -18.3692, -593.714, -2510.08,        0],
_J3  [ -27.2804,        0, -11.8109, -127.418]]

0.910260993448
-0.999974688762
```

In the elasticity matrix, J0, J1, J2, and J3 are synonymous to Equations 4, 5, 6, and 7, respectively. We notice that the effect of P in Equation 5, which describes the rate of change of P, is significantly high. The rate of change of P may be changing quickly due to the quorum sensing signal or the degradation rate of P. We check the sensitivity, or concentration coefficient, for the quorum sensing signal, 'Q.' The command "print r.getCC('P','Q')" returns a single numerical value; this is the sensitivity of P when 'Q,' the quorum sensing signal, is perturbed. Then, we compare this value with the concentration coefficient of P with respect to the degradation rate of P, which is 'c.' The concentration coefficient corresponding to the degradation rate has a higher magnitude, so it is probably the one we would want to change.

Conclusion and Future Work

By studying a simple genetic network qualitatively and quantitatively, we find that B. subtilis biofilms have a sporulating phenotype when nutrients are not readily available. A special case is when there are low nutrients and high population density; S levels here are the highest, so there is a greater probability of cells sporulating. Otherwise, B. subtilis cells switch to the ECM-

producing phenotype with low nutrients. We also find that R and S have pulsing behaviors, and S rises quickly in comparison to R. When nutrients are low, the reaction rate of R and S doubles. We would like to further study the C and P dynamics since the qualitative and quantitative studies did not give any conclusion about these transcription factors. Also, we would like to draw connections between our theoretical mathematical model and qualitative laboratory experiments. Particularly, we hope to create an SBML model and simulation that replicate the spatiotemporal growth of biofilms as observed in a wet-lab setting.

Bacterial biofilms are becoming more prevalent in our daily lives. Biofilms can help us; for example, they have a major role in wastewater treatment. Biofilms can also harm us. Bacteria in biofilms are more resistant to antibiotics due to their polymeric ECM, and the diseases they cause are more difficult to eradicate than ordinary bacterial infections. Studying the growth mechanisms of biofilms from a molecular standpoint is critical to our understanding of such bacterial communities.

Dedications

I want to thank my Professor and Primary Investigator **Dr. Yi Jiang** for mentoring me as I grow my roots in bioinformatics research. Without taking Dr. Jiang's Mathematical Biology class and getting her help in modeling, coding, etc. for two years, I would have never discovered my niche in the quantitative sciences. I give thanks to **Howard Smith**, my Ph.D. student mentor, for helping me get started with the biofilm research project and giving his time to train me. Any research lab in the country would be lucky to have a dedicated bioinformatician like Howard. Thank you to my best friend/sister **Christella Dhammaputri**, a Chemistry Ph.D. student at Emory University, for her continued support and inspiration. Christella's love of research motivated me to find a field I am passionate about and pursue my own project in it. Finally, I thank my partner **Naman Kanwar**, a Computer Science undergraduate at Georgia State University, for always pushing me to challenge myself and go beyond my goals. Naman's encouragement and advice is largely a reason why my research project came to fruition and why I plan to pursue a Ph.D. in Bioinformatics.

Reflection of RIMMES Program

When I first started research with the Jiang Lab, my peers were working on projects that were significantly more advanced than mine. The projects involved image analysis, machine learning, and many other mathematical methods I have never seen before. In comparison, my work seemed like a preschooler's finger painting. It was easy to feel discouraged in my situation, and I imagine many other math undergraduates would feel the same way. Math research is special in that it requires more working knowledge than other fields; to illustrate, I could not perform dynamical systems analysis on my genetic network without understanding differential equations or learning how to implement it with a code. Seeing the advanced math in my peers' research was

intimidating, but I did not want to give up on the biofilm project. I had to build my skillset up to work on my project. The process was difficult for sure, but without taking on the challenge, I would not have learned Python and MATLAB. I would not have known as much as I do now about systems biology, bioinformatics, etc. I would not have gotten the opportunity to share with the academic community a topic that I believed was so important and underrated. The obstacles and challenges I endured were all worth the effort because I saw my research project come to life, and that is more rewarding than anything I can imagine. I am very grateful for the opportunity to research with Dr. Jiang and the Math Department at Georgia State, and I hope more undergraduates will take advantage of the RIMMES program. Now, I know the big (quantitative science) questions I want to answer in my life, but first, I need a higher degree...

References

[1] Alon, Uri. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall/CRC, 2007.

[2] Alon, U., and Mangan, S. "Structure and Function of the Feed-Forward Loop Motif." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, 2003.

[3] Edelstein-Keshet, Leah. *Mathematical Models in Biology*. Random House, 1988.

[4] Voit, Eberhard O. *A First Course in Systems Biology*. Garland Science, 2018.

[5] Sauro, Herbert M. *Systems Biology: Basic Pathway Modeling Techniques.* 1st ed., 2015.

[6] Jabbari, Sara, et al. "Mathematical Modelling of the Sporulation-Initiation Network in Bacillus Subtilis Revealing the Dual Role of the Putative Quorum-Sensing Signal Molecule PhrA."*SpringerLink*, Springer-Verlag, 18 Mar. 2010.

[7] Boguslawski, Kristina M., et al. "Novel Mechanisms of Controlling the Activities of the Transcription Factors Spo0A and ComA by the Plasmid-Encoded Quorum Sensing Regulators Rap60-Phr60 in Bacillus Subtilis." *Molecular Microbiology*, Wiley/Blackwell (10.1111), 18 Feb. 2015, onlinelibrary.wiley.com/doi/10.1111/mmi.12939/abstract.

[8] Mura, Ivan, et al. "Computational Modelling and Analysis of the Molecular Network Regulating Sporulation Initiation in Bacillus Subtilis." *BMC Systems Biology*, BioMed Central, 24 Oct. 2014, bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-014-0119-x.

[9] The Gum Disease Information Bureau. Retrieved March 15, 2018 from http://gumdiseaseinfo.org.uk/toothbrushes/.

[10] Math Insight. Retrieved March 13, 2018 from mathinsight.org/image/three_node_motifs/.

[11] Bjarnsholt, T. “The role of bacterial biofilms in chronic infections.” *APMIS Supplementum.* PubMed, May 2013.

[12] Donlan, R. “Biofilm Formation: A Clinically Relevant Microbiological Process.” *Clinical Infectious Diseases*, Volume 33, Issue 8, 15 October 2001, 1387–1392.

[13] Mah, T., and O’Toole, G. “Mechanisms of biofilm resistance to antimicrobial agents.” *TRENDS in Microbiology,* Volume 9, No.1, January 2001.

[14] Costerton, J.W., Stewart, P., and Greenberg, E. P. “Bacterial Biofilms: A Common Cause of Persistent Infections.” *Microbes, Immunity, and Disease.* Science Magazine. Volume 284, May 1999.

[15] Kim, D., Kwon, Y., and Cho, K. “The biphasic behavior of incoherent feed-forward loops in biomolecular regulatory networks.” *BioEssays.* Wiley Periodicals. Volume 30, 2008.

[16] Donlan, R. “Biofilms: Microbial Life on Surfaces.” *Emerging Infectious Diseases.* Volume 8, No. 9, September 2002.

[17] Vlamakis, H., Aguilar, C., et. al. “Control of cell fate by the formation of an architecturally complex bacterial community.” *Genes & Development.* Cold Spring Harbor Laboratory Press. Volume 22. 2008.

[18] Cairns, L., Hobley, L., and Stanley-Wall, N. “Biofilm formation by Bacillus subtilis: new insights into regulatory strategies and assembly mechanisms.” *Molecular Microbiology.* Volume 93. 2014.

[19] Wang, X., et. al. “Probing phenotypic growth in expanding Bacillus subtilis biofilms.” *Applied Microbiology Biotechnology.* Springer. Volume 100. 2016.

[20] Lopez, D., Vlamakis, H., and Kolter, R. “Generation of multiple cell types in Bacillus subtilis.” *Federation of European Microbiological Societies Reviews*. Blackwell Publishing. 2009.

[21] Fujita, M., et. al. “High- and Low-Threshold Genes in the Spo0A Regulon of Bacillus subtilis.” *Journal of Bacteriology.* Volume 187, No. 4. 2005.