LETTERS IN BIOMATHEMATICS
An International Journal

Intercollegiate Biomathematics Alliance

EDUCATION ARTICLE

OPEN ACCESS

Modeling and analysis of the firefly luciferase reaction and a G-protein coupled receptor signaling pathway with ordinary differential equations increases self confidence in mathematical cell biology for novice graduate students

Renee Dale[a], Naohiro Kato[b], Bill Wischusen[b]

[a]Donald Danforth Plant Science Center, St. Louis, MO; [b]Department of Biology, Louisiana State University, Baton Rouge, LA

ABSTRACT

The push for mathematics and computation in biology is greater than ever before. However, perceptions among biology students on the difficulty of mathematics and programming are a barrier to their implementation. A teaching module introducing quantitative skills to biology students is needed. We implemented modeling of firefly luciferase and G-protein coupled receptor signaling with ordinary differential equations in a course for novice graduate students. We assessed whether the course helped the students increase self confidence in application of mathematics in biology. Two concept inventories tracked learning gains in both general and cell biology knowledge. Pre- and post-semester surveys quantified changes in student confidence and their opinions of the usefulness of these techniques in cell biology. We found that the modeling and analysis activities appeared to improve self confidence in and appreciation of quantitative mathematical biology techniques. We describe our assessment methods to determine the suitability of our module.

ARTICLE HISTORY
Received May 6, 2019
Accepted January 7, 2020

KEYWORDS
Mathematical modeling, systems biology, cell biology, quantitative skills, computational biology, education

1 Introduction

Mathematics and quantitative skills are increasingly critical in the field of biology. Whether a student is looking to work in public health, education, or basic research, understanding mathematical models will be required. In particular, systems cell biology uses comprehensive and quantitative experimental methods that are interpreted using mathematical and statistical models (Mast et al., 2014).

Vision and Change calls for more quantitative training in the biological sciences (Woodin et al., 2010). However, the social perception of mathematics being for 'geniuses' serves to enhance the struggles of disadvantaged groups (Lamb, 2018, 2015; Leslie et al., 2015). Educators experiencing math anxiety are not always confident in their ability to teach students the quantitative skills necessary for their career (Chen et al., 2018). When biology students are exposed to new methods involving mathematics, they often 'freak out' (Chen et al., 2018). This is partially due to social perceptions of mathematics and to belief that new methods are inherently more difficult (Leslie et al., 2015; Lamb, 2018, 2015). Research has shown that when students engage in studying mathematics and technology, their attitudes improve (Uitto et al., 2006; Ravenel et al., 2014; Afari et al., 2012). These experiences have also led to increased efficacy in their studies and engagement when learning mathematics in context of their existing interests (Afari et al., 2013; Eaton and Highlander, 2017). However, post-secondary education is often not up to the challenge (Chen et al., 2018). Math class requirements for STEM undergraduates are a significant factor in their low retention in STEM (Chen, 2013; Belser et al., 2018). Mathematical anxiety or gender bias present in teachers may prevent them from demonstrating the utility or importance of mathematics to their students before they choose their majors in college.

Given the issues involved in obtaining the skills required to study systems cell biology, the ability to understand and quantify the effectiveness of an exercise or course is increasingly important. In the near future, the ultimate goal of a cell biology course will be for students to be able to model a cellular system. Due to the multidisciplinary nature of systems cell biology, this necessitates exposure to several knowledge bases: basic mathematics and programming skills in addition to specifics about the cellular system

CONTACT Renee Dale rdale825@gmail.com

being modeled. In some cases, the focus of such a class may be understanding non-linear kinetics of cellular systems. In our case, we wanted the students to understand the transferable concepts involved in signaling pathways, such as feedback loops, by modeling them with ordinary differential equations. This motivated the development of a teaching module that introduced modeling of cellular systems with ordinary differential equations to our students who are interested in mathematical cell biology, yet lacking the self confidence required to self-study. To measure student learning gains in the course, we developed a general biology concept inventory and two concept inventories on domain-specific knowledge (here, meaning the two biological systems modeled in class). We also measured student attitudinal changes toward mathematics and quantitative biology using a survey. In addition, we collected open-ended survey data to compare student comments with the overall improvement in the four areas. We find that our module allowed students to develop self confidence in their quantitative skills and their ability to utilize them in their future research. However, our sample size is very small. We hope to provide a starting point for future work on developing similar modules, including issues that we experienced during our work.

2 Methods

2.1 Study Population

The study population was a group of 7 students enrolled in graduate programs at Louisiana State University. Four students were from the department of biological sciences, and three students were from the department of engineering. The students took a course entitled mathematical modeling in cell biology, which was held in the fall semester of 2017. The students from the department of biological sciences had little to no prior experience with programming in MATLAB, while this subject is commonly introduced in undergraduate engineering curricula. The students from the school of engineering had little to no prior knowledge of the G-protein couple receptor signaling pathway, while this is commonly introduced in undergraduate biology curricula. All students had little or no prior experience with developing mathematical models. All students had taken at least one calculus course at the undergraduate level previously. This was determined by self-report on the part of the students. We received Institutional Review Board (IRB) approval for this study (E10547).

2.2 Course Content

The objective of the course was for students to understand how to abstract a biological system to a conceptual model, describe this model using ordinary differential equations, write code with MATLAB to simulate this model, and gain a basic understanding of how to interpret model components, such as parameters. A total of 27 sessions including midterm and final examinations were held over the course of a semester, with each session lasting around 80 minutes. First we introduced students to MATLAB programming (2 sessions). We then discussed biological concepts using the textbooks Molecular Cell biology and Molecular Biology of The Cell (Lodish et al., 2016; Alberts et al., 2002). We focused on four main biological systems: 1) protein-protein interactions, 2) enzyme reactions, 3) gene expression, and 4) signaling pathways, a total of 8 sessions. We then described how these biological concepts could be described with ordinary differential equations using a textbook *Mathematical Modeling in Systems Biology: An Introduction* by Brian P. Ingalls over a total of 4 sessions (Ingalls, 2013).

Students were tasked with two modeling case studies: 1) conversion of Michaelis-Menten kinetics to ordinary differential equations, and 2) G-protein coupled signaling pathway. These case studies spanned 6 sessions. Finally, the students model for their own self-selected system as the final project over the remaining 5 sessions of the course. PowerPoint files used for the two case studies were provided as Supplemental materials.

2.2.1 Modeling Michaelis-Menten kinetics of the enzyme firefly luciferase

Our learning goals for the students in this module included conversion of a conceptual model to a system of ordinary differential equations. They were to complete this in MATLAB. This was the first time the students would be constructing a mathematical model, although some students had prior experience with MATLAB or Python. Because all students in the course had learned the basics of enzymatic reactions in high school and throughout introductory biology courses in college, we thought it would be a good system for the introduction of modeling regardless of their majors. It is also important to note that modeling Michaelis-Menten kinetics using a system of ordinary differential equations is a classical application of mathematics to biological systems. The equations describe how the enzyme, substrate, and product dynamics change over time. For an ordinary biology major student, however, K_M, the Michaelis constant, is the important parameter to understand the function of an enzyme quantitatively. The Michaelis constant can be obtained by simplifying the system by assuming that the amount of enzyme-substrate complex does not change significantly over time, so that the production rate depends on the amount of enzyme and substrate (quasi-steady-state approximation). The conceptional model of the Michaelis-Menten kinetics is described in Fig. 1. Here E stands for enzyme, S for substrate, $E \cdot S$ for the enzyme-substrate complex, and P for product. In this model, enzyme and substrate reversibly bind to form the enzyme-substrate complex before the enzyme catalytically releases the product. By assuming that

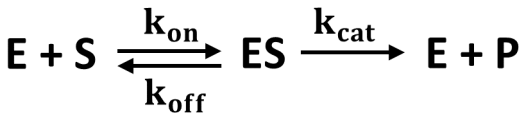

Figure 1: System of enzyme kinetics. E is the enzyme, S is the substrate, ES is the enzyme-substrate complex, and P is the product. Parameter names are shown near the arrows, which show the directionality of their effect. k_{on} is the substrate binding rate (Molar s^{-1}), k_{off} the substrate unbinding rate (s^{-1}), and k_{cat} the product formation rate (s^{-1}).

the concentration of $E \cdot S$ is not changed after burst formation of $E \cdot S$, we can obtain a mathematical model known as the Michaelis-Menten equation.

$$\frac{dP}{dt} = \frac{k_{\text{cat}} \cdot E_0 \cdot S}{K_{\text{M}} + S} \tag{1}$$

This model considers the concentration of unbound product P, total enzyme E_0, and substrate S. The catalytic rate k_{cat} is a constant (s^{-1}) that describes the catalytic release of P after catalysis occurs in the $E \cdot S$ complex. K_{M} represents the Michaelis constant, representing the amount of substrate required to reach half-maximum reaction velocity. The catalytic rate and total enzyme, $k_{\text{cat}} \cdot E_0$, is often alternatively expressed as V_{max}.

In the course, we used the non-steady state approximation modeling approach and consider both binding and unbinding of the substrate. This is because we emphasized the importance to translate a diagrammatic model into ordinary differential equations. The resulting system contains 4 components and 3 parameters (Fig. 1). Substrate binding is typically in terms of molar s^{-1}, while unbinding is s^{-1}, and amount of each component is molar.

This case study provides students with a simple system that is relatively intuitive. Product cannot be formed without the enzyme first binding the substrate. The amount of enzyme binding the substrate is described by the trade-off between the binding and unbinding rate, as well as the amount of substrate available. When the substrate is present in excess, the amount of enzyme-substrate complex reaches an equilibrium, and the Michaelis-Menten quasi-steady state model is equivalent to the non-steady state model. Simulations of the ordinal differential equation models illustrate these effects, helping the student understand quasi-steady state approximation, as well as how they might think more critically about their conceptual models of the enzymatic reaction.

To model this system, students are instructed to use the states and parameters provided in Fig. 1 and describe as ordinary differential equations. Ordinary differential equations lend themselves well to the study of biological phenomena, as they describe how something changes over time. Many biological processes are described by how much they change over time.

$$\frac{\text{change in } X}{\text{time}} == \frac{dX}{dt} \tag{2}$$

Applying this logic to the enzyme reaction, students must identify how each of the 4 states is changing over time. Students are told that the amount of free enzyme decreases as it binds to the substrate. From the equation in Fig. 1,

$$E + S \longrightarrow E \cdot S \tag{3}$$

Mathematically, this translates to

$$\frac{dE}{dt} = -k_{\text{on}} \cdot E \cdot S \tag{4}$$

The binding event is dependent on the binding rate, as well as the amount of free enzyme and free substrate. In addition, the amount of free enzyme is increased when the enzyme-substrate complex breaks apart, or when catalysis occurs. Adding these two events into the model, the equation describing the change in free enzyme over time becomes

$$\frac{dE}{dt} = -k_{\text{on}} \cdot E \cdot S + k_{\text{off}} \cdot ES + k_{\text{cat}} \cdot ES \tag{5}$$

Applying these methods to the remaining states in the system, the students derive the following system of equations describing this system (Eqn. 6 to 9). Note that since enzyme is not used up when it catalyzes a reaction, when catalysis occurs to the enzyme-substrate complex, the enzyme involved in that catalysis is returned to the pool of free enzyme at the same rate that

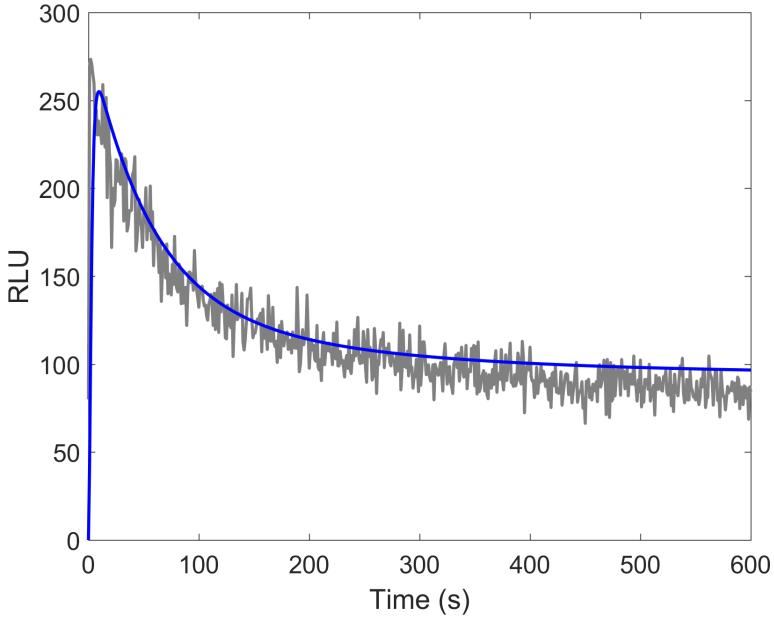

Figure 2: The dynamics of firefly luciferase show a short peak in light very quickly, followed by a slow decay as substrate is exhausted. The different parameters of luciferase affect different portions of the dynamics, illustrating their effects in an enzymatic system. Here, the model simulation (blue) has had its parameters altered to fit the data (grey).

catalysis occurs.

$$\frac{dE}{dt} = -k_{on} \cdot E \cdot S + k_{off} \cdot ES + k_{cat} \cdot ES \tag{6}$$

$$\frac{dS}{dt} = -k_{on} \cdot E \cdot S + k_{off} \cdot ES \tag{7}$$

$$\frac{dES}{dt} = k_{on} \cdot E \cdot S - k_{off} \cdot ES - k_{cat} \cdot ES \tag{8}$$

$$\frac{dP}{dt} = k_{cat} \cdot ES \tag{9}$$

Following this exercise, students were provided with real enzymatic data we collected using the firefly luciferase assay in the previously published research article (Dale et al., 2016). Firefly luciferase is an enzyme that oxidizes luciferin, the substrate to form oxyluciferin, which releases a photon. Since the radiative decay of excited oxyluciferin to the ground state is much faster (nanoseconds) than the catalytic reactions (sub-seconds), the production rate of luciferase can be determined by the photons emitted from oxyluciferin. This illustrates the dynamics of product formation when the light production is measured. In other words, the reaction model of firefly luciferase can be built mathematically and confirmed experimentally without the quasi-steady-state approximation. To this end, we added one more equation that describe light (λ) emission from oxyluciferin.

$$\frac{d\lambda}{dt} = \gamma \cdot k_{cat} \cdot ES - \lambda \tag{10}$$

Here γ represents a parameter for the light emission measured as relative luminescence units (RLU) recorded by luminometer, a light detector. The exact RLU value is unique to the equipment used. Students were tasked with modeling the firefly luciferase reaction and modifying the model to describe the dynamics in the data. They were also introduced to parameter estimation techniques to assist them in this task. In Fig. 2, we present the curve fit of the firefly luciferase reaction. This exercise allows us to illustrate how helpful mathematical models can be, in determining if your current understanding of the system is sufficient to explain the data you see. Also it is important to emphasize that "all models are wrong, but some are useful," as famously stated by statistician George Box. We feel this is best illustrated by seeing firsthand the subjectivity involved in attempting to produce a model that can reproduce the observed dynamics with a simulation.

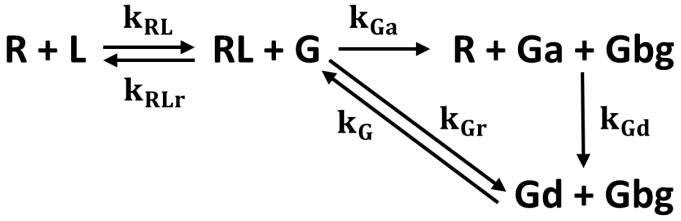

Figure 3: G-protein signaling pathway. G is free G-protein trimer, Ga is Gα bound to GTP, Gd is Gα bound to GDP, Gbg is G$\beta\gamma$ subunit complex, R is the receptor, and L is the ligand. Parameter names are shown near the arrows, which show the directionality of their effect. The ligand binding rate is indicated by k_{RL}, k_{RLr} the ligand unbinding rate, k_{Ga} the rate of Ga dissociating and binding GTP, k_{Gd} is the rate of dephosphorylation of the GTP bound to Ga, k_{Gr} is the dissociation rate and k_{G} the association rate of the G-protein trimer.

2.2.2 Modeling the G-protein coupled receptor signaling pathway

G-protein coupled receptors are membrane-bound proteins involved in many biological signaling processes. These receptors bind signaling ligands produced by other cells, and the signals produce changes inside the cell. When the receptor is unbound, the G-protein Gα (here, Ga) subunit is bound to GDP, and is associated with the Gγ (Gg) and Gβ (Gb) subunits, forming a trimer. After the ligand reversibly binds the receptor, the Ga subunit reversibly binds to GTP and propagates the signal by dephosphorylating the GTP to form GDP (represented as Gd). The Gb and the Gg complexes reversibly separate. This system contains 7 components and 6 parameters (Fig. 3). We chose this system to model because G-protein coupled receptors are described in all major textbooks of cell biology for college students. We used the model and analysis that were previously described in the textbook, *Mathematical Modeling in Systems Biology: An Introduction* by Brian P. Ingalls (Ingalls, 2013).

The system of equations describing this system is shown in equations 11 to 16. Eqn. 11 describes the change in receptor-ligand complex over time, as the ligand binds the receptor reversibly. Eqn. 12 describes the change in free G-protein over time, as it binds the receptor-ligand complex and GDP. Eqn. 13 describes the change in Ga complex, which increases when G-protein binds the receptor-ligand complex and decreases when GTP is dephosphorylated. Again, binding is molar s^{-1}, unbinding s^{-1}, and dephosphorylation is in terms of s^{-1}. The model components are expressed in terms of molarity.

$$\frac{dRL}{dt} = k_{RL} \cdot L \cdot R - k_{RLr} \cdot RL \tag{11}$$

$$\frac{dG}{dt} = -k_{Ga} \cdot G \cdot RL + k_G \cdot Gd \cdot Gbg - k_{Gr} \cdot G \tag{12}$$

$$\frac{dGa}{dt} = k_{Ga} \cdot RL \cdot G - k_{Gd} \cdot Ga \tag{13}$$

$$R = R_{\mathrm{Total}} - RL \tag{14}$$

$$Gd = G_{\mathrm{Total}} - Ga - G \tag{15}$$

$$Gbg = Gd + Ga \tag{16}$$

This system also contains some algebraic terms, since we consider the total number of existing receptors (Eqn. 14), Ga-GDP (Eqn. 15) and Gbg (Eqn. 16) complexes in the system. Algebraic terms can be useful when logical constraints exist on the system. For example, if we have some total receptor existing in the cell, then we can determine free receptor from the receptor-ligand dynamics algebraically.

This system is not as intuitive as enzyme kinetics, and it illustrates the benefit of mathematical modeling well. There are more complex transitions, with multiple things occurring at a single step (for example, dephosphorlyation). In addition, some components not explicitly modeled, such as GTP and GDP. Although the students might be able to predict the effect of changes in a component or parameter on the system accurately, they likely would not be able to predict the magnitude of the effect without a mathematical model. In Fig. 4, we present the effect of reducing the dephosphorylation rate of Gα by 0.11 units per second to zero that was described in the textbook and discussed in the course (Ingalls, 2013).

The mathematical model forces the students to understand the system quantitatively, which helps them identify a portion that is not involved in the dynamic response of the cell or whose function is wrongly assumed, given the quantitative relationships between the system components. For example, if we have data showing that a lack of dephosphorylation does not cause a 10-fold increase in the amount of Gα bound to GTP, maybe another dephosphorylation enzyme is present, or another component is providing negative feedback on Gα. Both of these hypotheses can be rapidly and quantitatively tested by modifying the model.

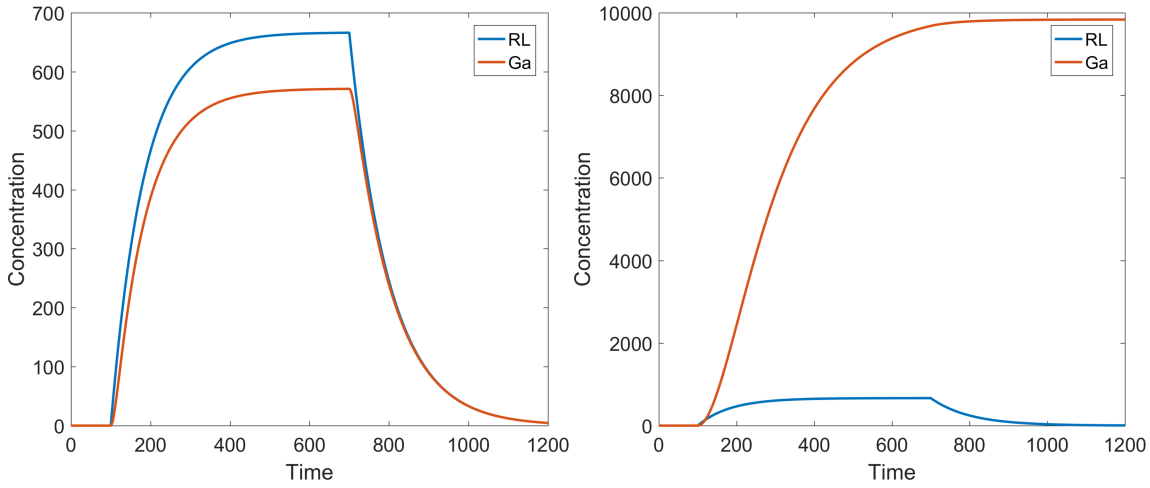

Figure 4: The mathematical model illustrates the effect of the dephosphorylation rate on Gα bound to GTP (blue) and the receptor-ligand complex (orange). Dynamic response is described. At time t = 100 sec, 1 nM of ligand is introduced. This signal is removed at time t = 700 sec, causing the response to decay. (Figure left) the dephosphorylation rate *kgd1* is set at 0.11 units per second (assuming wild type cell). (Figure right) the rate sets at zero (assuming the Gα protein is mutated in the cell).

2.3 Opinion and confidence survey

We developed a 5-point Likert scale survey to capture attitudes of the students on topics related to mathematical modeling. We sought to measure student confidence changes in four areas: technical skill, synthesis, domain knowledge, and in general.

- "Technical skill": These questions ask students about their confidence in performing quantitative techniques such as programming and math. The survey contained 5 Technical skill questions.
- "Synthesis": These questions ask students about their confidence and opinion on the value of systems-thinking. In this case, we consider systems-thinking to be synthesis between biological and quantitative understanding. The survey contained 5 Synthesis questions.
- "Domain knowledge": These questions ask students about their confidence as an experimental biologist. The survey contained 2 Domain questions.
- "General": confidence to perform basic functions of a scientist - experimental design, collaboration, and communication. The survey contained 3 General questions.

The survey had four apparent categories: techniques, experimental design, knowledge, and confidence. The grouping was intended to mask the true intentions of the survey. This is to avoid "social desirability" response bias, since students are aware that mathematical biologists would be viewing their responses (van de Mortel, 2008; Willis and Artino, 2013). This tendency is of direct concern to us, due to the social perceptions of math and programming. Questions given to the students and their areas are

Techniques

1. I am confident in my ability to write computer code. *Tech*
2. I am confident in my ability to write differential equations to describe a process. *Tech*
3. I am confident in my ability to design biological experiments. *Domain*
4. I am confident in my ability to construct conceptual models. *Synth*
5. I am confident in my ability to draw a diagram describing a system. *Synth*
6. I am confident in my ability to write a mathematical model. *Tech*

Experimental Design

1. When designing an experiment I think about a conceptual model. *Synth*

2. When designing an experiment I think about a diagram of the system. *Synth*
3. When designing an experiment I think about a mathematical model of the system. *Tech*

Knowledge

1. Knowledge of mathematics is important as a researcher in Biology. *Tech*
2. Knowledge of systems is important as a researcher in Biology. *Synth*
3. Knowledge of experimental techniques is important as a researcher in Biology. *Domain*

Confidence

1. I am confident in my ability to design an interdisciplinary research project. *General*
2. I am confident in my ability to collaborate with someone outside my field. *General*
3. I am confident in my ability to communicate with someone outside my field. *General*

2.4 Concept inventory

Because this course consisted of new biological, mathematical, and programming content, we wanted to determine if any benefit might be had in terms of general biology knowledge, particularly due to the presence of students outside of the field (engineering). We curated a set of 24 concept inventory questions on general biological knowledge from the literature (Fisher et al., 2011; ECI, ECI; Klymkowsky et al., 2010; Shi et al., 2010; Bretz and Linenberger, 2012; Couch et al., 2015). As the course consisted of 4 general biological topics (enzymes, gene expression, diffusion, and signaling), we selected a set of 6 questions per course topic. The 24 concept questions are provided as Supplemental Materials.

2.5 Specific inventory

Since many of the concept inventory questions we found contained information from introductory biology courses, which was many years ago for our study population of graduate students, we developed two specific concept inventory surveys based on specific modeling case studies: the G-protein coupled signaling pathway and firefly luciferase enzymatic reactions. Students took the pre-test immediately before applying their modeling skills on the case study, and post-test after. The questions covered broad concepts that required "mathematical" or "systems" thinking, such as the factor that drives the dynamics of the system. Questions are provided as a Supplement. Some questions are only available by contacting the authors due to the stipulation of the original authors.

2.6 Open-ended opinion survey

We designed the open-ended opinion survey to validate our previous surveys. The open-ended survey included questions about math confidence, utility, and student opinion of the value of the course. All questions are available in the supplemental materials.

2.7 Analysis

To determine how the course affected student perceptions related to mathematical modeling in biology, we used a survey to track changes in opinion and confidence, a general biology concept inventory to track general biology knowledge, and two concept inventories specific to the two modeling case studies as summarized in Fig. 5.

To quantify gains in learning, confidence, or opinion we normalized for values greater than zero using $\frac{\text{posttest} - \text{pretest}}{100 - \text{pretest}}$ (Hake, 1998). We used a cutoff of zero since we assume that a negative learning gain represents student guessing. Negative learning gains were observed fairly commonly in the concept inventories. For the opinion and confidence survey, only one question by one respondent dropped by one point in the post-survey. We considered this to be a non-significant zero change.

To identify potential changes in student responses we used dependent-samples t-tests. We consider a significant difference to be at the $p < 0.05$ level. In the opinion and confidence survey, we analyzed each question across all respondents to determine which questions had significantly different responses from the initial survey (provided as Supplemental File 1). These t-tests had 6 respondents and 5 degrees of freedom. To determine changes in confidence across the 4 categories described earlier in the Methods section, we averaged the responses across each category, and took the degrees of freedom to be the number of questions less 1.

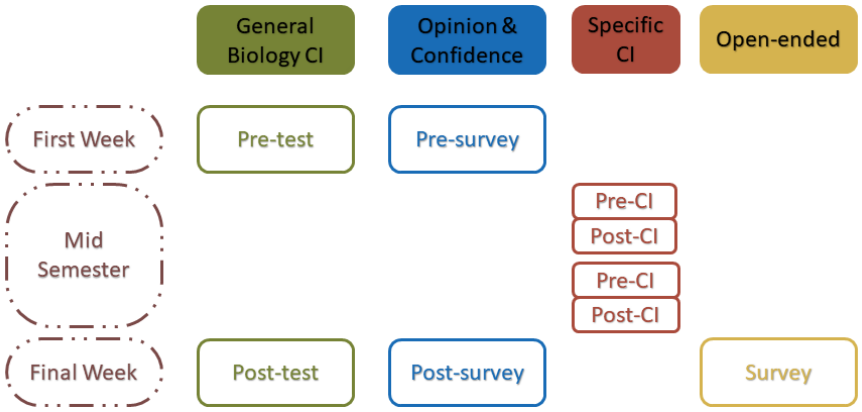

Figure 5: Student gains were measured by the four assessment methods spread out over the semester.

Caveat. Due to the very low sample size, we urge caution with regard to interpretation of statistical significance. Without replication of our results, no conclusive inference can be made, particularly for the concept inventories. For the survey responses, our use of the t-test is valid since the Likert scale produces ordinal data (Sullivan and Ar, 2013). When analyzing Likert scale data, its considered better to use the median rather than the mean. For our data, almost all of the mean changes are very close to the median, so it appears our data would be close to normally distributed.

2.8 Results

We measured changes in student understanding of biological concepts using general biology questions and questions specific to course content. A dependent-samples t-test was conducted to compare student performance on the initial and final concept inventories. We ask that you please keep in mind the sample size when interpreting our results. We suggest that our module and assessments be stepping stones for others, as suggestions or cautions. Student learning gains as measured by the general biology concept inventory showed an average gain of 15% compared to the initial student scores (Fig. 6). There was not a significant difference between student scores on the pre- and post-concept inventories ($t(4) = 1.3, p > 0.2$). Our two specific inventories covering material implicitly covered during the course (i.e., quizzes on conceptual materials) showed larger learning gains. The inventory covering the G-protein coupled receptor signaling pathway had an average gain of 13%, and the firefly luciferase enzyme kinetics showed an average gain of 28% (Fig. 6). There was a significant difference between student scores on the luciferase inventory ($t(5) = 2.7, p < 0.05$) but not the G-protein coupled receptor inventory ($t(4) = 1.6, p > 0.1$).

We measured changes in student confidence and opinion related to quantitative biology over four categories (Fig. 6, see Methods for description). We saw the largest improvement in the respondents' confidence in their mathematical and programming skills (*Technical skill*, *Mean* = 61%, *SD* = 11%). This was a statistically significant difference compared to initial student responses ($t(5) = 12.4, p < 0.001$). Their confidence in collaboration and communication improved by 53% on average (*General*, *SD* = 7%). This was a statistically significant difference compared to initial student responses ($t(2) = 13.8, p < 0.01$). The reported confidence in synthesizing mathematical and biological concepts increased by 52% (*Synth*, *SD* = 31%). This was a statistically significant difference compared to initial student responses ($t(4) = 3.7, p < 0.05$). The respondents' confidence in their field of biology, unrelated to mathematical biology, increased by 37% (*Domain*, *SD* = 4%). This was a statistically significant difference compared to initial student responses ($t(1) = 13.8, p < 0.05$). There was a significant difference in student score for 11 out of the 15 survey questions (shown in Supplemental File 1). None of the survey questions in the "Knowledge" section had a significant difference across the semester, most likely due to high agreement on the importance of knowledge of mathematics, systems, and experimental techniques in the pre-survey.

The open-ended survey included questions that were redundant with the survey intended to understand confidence and opinion changes. Respondents answered positively, and generally in agreement with the observed effects. In answer to the question "Do you see yourself modeling in the future?" a student answered: *I can see myself modeling in the future to determine what can be the best initial conditions in a biological experiments.*

We found the most interesting and promising answers to the question "Do you think this course will have helped you form interdisciplinary collaboration in the future?" *Yes, it has showed how important and helpful a mathematical model can be to a biological system.* The final question "What were your previous methods to understand or explain a biological system in an educational or publication context? Do you think this will change after you have done some mathematical modeling?"

- *Wet lab, trial and error. I will use mathematics for a better planification of experiments.*

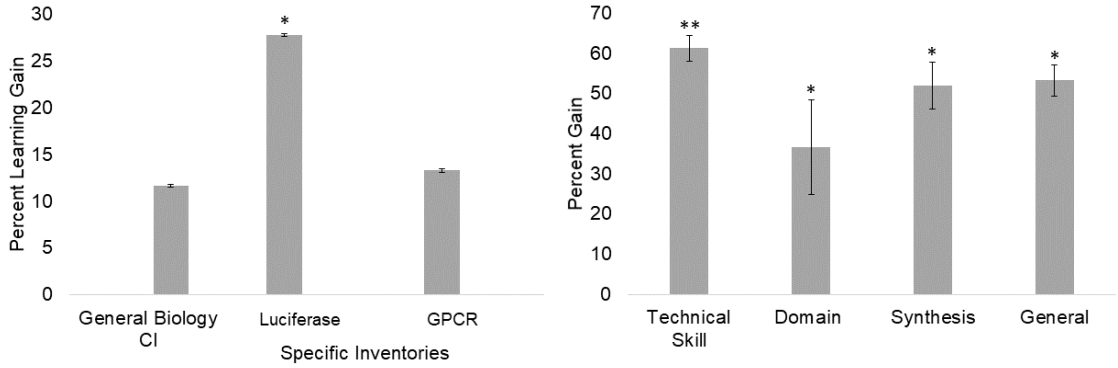

Figure 6: Outcomes of a mathematical modeling course measured by 4 different areas and two inventories. (Left) We measured learning gains in a general biology concept inventory (*Mean* = 12%) as well as learning gains specific to two modeling scenarios (*Mean* = 28% luciferase, 13% GPCR). The learning gains were significant for the luciferase inventory ($p < 0.05$), but not the general concept or GPCR inventories. (Right) We measured changes in students' opinions using a survey. The students reported large gains in their mathematics and programming confidence (technical skill, *Mean* = 61%) and their opinion of the importance and utility of quantitative methods in biology (synthesis, *Mean* = 52%, and general, *Mean* = 53%). All categories of questions had statistically significant differences compared to initial responses at the $p < 0.05$ (*) or $p < 0.001$ (**) levels, as indicated in the figure.

- *I would dig in literature and take online courses on introductory levels of systems biology.*
- *Previous methods only included understanding steps in a diagram. Now, one must consider quantities and speed and each step.*

3 Discussion

Our results indicate that having novice graduate students apply mathematical methods to topics indirectly related to their fields improves their confidence in quantitative methods as well as their opinion of the usefulness of those methods. We did not expect to observe improvement in their confidence to design a biological experiment. However, student responses to the open-ended survey shed light on why we observed this improvement. The answers provided above illustrate why the course may have had the side-effect of improving student confidence in biological experimental design. The students saw the value of considering quantitative analysis prior to designing and carrying out an experiment. This is likely also why we saw improvements in the general confidence category, specifically the ability to design a project, collaborate, and communicate outside the field. We believe that the results show that presenting multiple biological systems helps reinforce the transferability of the quantitative techniques. However, more research would be needed to determine if different results would be found in a course which focused on one system.

Our concept inventories did not show strong learning gains. This may be due to question formulation, or a lack of emphasis on specific biological concepts. Survey respondents misunderstanding a question is another possible issue in assessment development (Willis and Artino, 2013). We attempted to ask highly conceptual questions in assessing student understanding of the two systems, enzyme kinetics and G-protein coupled receptor signaling pathway. This is because conceptualization of a system is an important first step, prior to writing a mathematical model. Conceptual understanding is more difficult to assess, and poor wording is likely in the first iteration of usage. For example, in the open-ended survey, we noticed at least one incidence of misunderstood question. In answer to the question, "Do you think experimental design is affected by knowledge of mathematical modeling?" a student answered: *No. I don't think modeling can overpower the benefits of experiments on any day. Each is needed in its own way. Once you learn about the connection, it will be affected and possibly improved.* The student interpreted this question as forcing a choice between either experimental or mathematical methods, whereas our intention was to understand how the students might apply their new skills in realistic research scenarios. Improvement in questionnaires would be necessary to avoid such misunderstanding among students.

The luciferase enzyme system is much simpler than the GPCR system, both in terms of the conceptual and mathematical complexity. This may explain the difference in learning gains between the two specific concept inventories. It may be prudent to spend more time on the GPCR system before asking conceptual questions. Finally, setting specific learning goals for each

example system in the module, in addition to the module as a whole, may help construction of conceptual systems questions. We set goals for the module, which focused on student confidence and technical skill, but not for the two systems.

Our module could be incorporated into existing course materials in a cell biology course, and is especially well-suited as a laboratory component. The students will first need to be introduced to the basics of modeling, programming, and parameter optimization. Then the students will be prepared to engage in the two modeling systems, particularly when provided with example code (see Supplemental for our materials and code). Due to the emphasis on technical skill, and not on biological concepts per se, we think that the module could be utilized with undergraduate students as well. More testing is needed to determine its suitability and effectiveness at the undergraduate course in which larger numbers of students are enrolled and more diverse population regarding mathematical and computation skills is expected.

References

Afari, E., J. M. Aldridge, and B. J. Fraser (2012). Effectiveness of using games in tertiary-level mathematics classrooms. *International Journal of Science and Mathematics Education 10*, 1369–1392. 3

Afari, E., J. M. Aldridge, B. J. Fraser, and M. S. Khine (2013). Students' perceptions of the learning environment and attitudes in game-based mathematics classrooms. *Learning Environments Research 16*, 131–150. 3

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2002). *Molecular Biology of the Cell, Sixth Edition*. New York: Garland Science. 4

Belser, C. T., M. A. Shillingford, A. P. Daire, D. J. Prescod, and M. A. Dagley (2018). Predicting undergraduate retention in stem majors based on demographics, math ability, and career development factors. *The Professional Counselor 8*(3), 262–276. 3

Bretz, S. L. and K. J. Linenberger (2012). Development of the enzyme-substrate interactions concept inventory. *Biochemistry and Molecular Biology Education 2012*(40), 4. 9

Chen, M. M., S. M. Scott, and J. D. Stevens (2018). Technology as a tool in teaching quantitative biology at the secondary and undergraduate levels: a review. *Letters in Biomathematics 5*, 30–48. 3

Chen, X. (2013). *STEM Attrition: College Students' Paths Into and Out of STEM Fields (NCES 2014-001)*. Institute of Education Sciences, U.S. Department of Education. Washington, DC: National Center for Education Statistics. 3

Couch, B. A., W. B. Wood, and J. K. Knight (2015). The molecular biology capstone assessment: A concept assessment for upper-division molecular biology students. *CBE—Life Sciences Education 14*. 9

Dale, R., Y. Ohmuro-Matsuyama, H. Ueda, and N. Kato (2016). Mathematical model of the firefly luciferase complementation assay reveals a non-linear relationship between the detected luminescence and the affinity of the protein pair being analyzed. *PLOS ONE 11*. 6

Eaton, C. D. and H. C. Highlander (2017). The case for biocalculus: Design, retention, and student performance. *Cell Biology Education 16*. 3

ECI. Energy concept inventory. 9

Fisher, K. M., K. S. Williams, and J. E. Lineback (2011). Osmosis and diffusion conceptual assessment. *Cell Biology Education 10*, 418–429. 9

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics 66*, 64–74. 9

Ingalls, B. P. (2013). *Mathematical modeling in systems biology: an introduction*. Mass, MIT Press: Cambridge. 4, 7

Klymkowsky, M. W., S. M. Underwood, and R. K. Garvin-Doxas (2010). Biological concepts instrument (bci): A diagnostic tool for revealing student thinking. ArXiv Preprint. 9

Lamb, E. (2015). Math and the genius myth. 3

Lamb, E. (2018). 3

Leslie, S.-J., A. Cimpian, M. Meyer, and E. Freeland (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science 347*, 262–265. 3

Lodish, H. F., A. Berk, C. A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, A. Amon, and K. C. Martin (2016). *Molecular cell biology*. 4

Mast, F. D., A. V. Ratushny, and J. D. Aitchison (2014). Systems cell biology. *The Journal of Cell Biology 206*, 695–706. 3

Ravenel, J., D. T. Lambeth, and B. Spires (2014). Effects of computer-based programs on mathematical achievement scores for fourth-grade students. *Journal on School Educational Technology 10*(1), 8–21. 3

Shi, J., W. B. Wood, J. M. Martin, N. A. Guild, Q. Vicens, and J. K. Knight (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE-Life Sciences Education 9*, 453–461. 9

Sullivan, G. M. and Ar (2013). Analyzing and interpreting data from likert-type scales. *J Grad Med Educ 5*(4), 541–542. 10

Uitto, A., K. Juuti, J. Lavonen, and V. Meisalo (2006). Students' interest in biology and their out-of-school experiences. *Journal of Biological Education 40*, 124–129. 3

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing 25*, 10. 8

Willis, G. B. and A. R. Artino (2013). What do our respondents think we're asking? using cognitive interviewing to improve medical education surveys. *Journal of Graduate Medical Education 5*, 353–356. 8, 11

Woodin, T., V. C. Carter, and L. Fletcher (2010). Vision and change in biology undergraduate education, a call for action–initial responses. *Cell Biology Education 9*, 71–73. 3