Biochemistry and Molecular Biology Education

Using Graph-Based Assessments within Socratic Tutorials to Reveal and Refine Students' Analytical Thinking about Molecular Networks.

Journal:	*Biochemistry and Molecular Biology Education*
Manuscript ID:	BAMBED-11-0102.R2
Wiley - Manuscript type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Trujillo, Caleb; Purdue University, Biological Sciences Cooper, Melanie; Clemson University, Chemistry Klymkowsky, Michael; University of Colorado, Boulder, Molecular, Cellular & Developmental Biology
Keywords:	active learning, assessment of educational activities, cooperative/collaborative education, curriculum assessment, effective in-class problems, gene expression, learning and curriculum design, molecular biology, protein synthesis and turnover, student conceptual and reasoning difficulties

Using Graph-Based Assessments within Socratic Tutorials to Reveal and Refine Students' Analytical Thinking about Molecular Networks.

Caleb Trujillo[1,3] , Melanie M. Cooper[2] & Michael W. Klymkowsky [1,4]

[1] Molecular, Cellular & Developmental Biology
University of Colorado, Boulder, Boulder, CO. 80309.

[3] Chemistry,
Clemson University, Clemson, SC. 29634.

[2] Current address: Biological Sciences,
Purdue University, West Lafayette, IN. 47907. ctrujil@purdue.edu

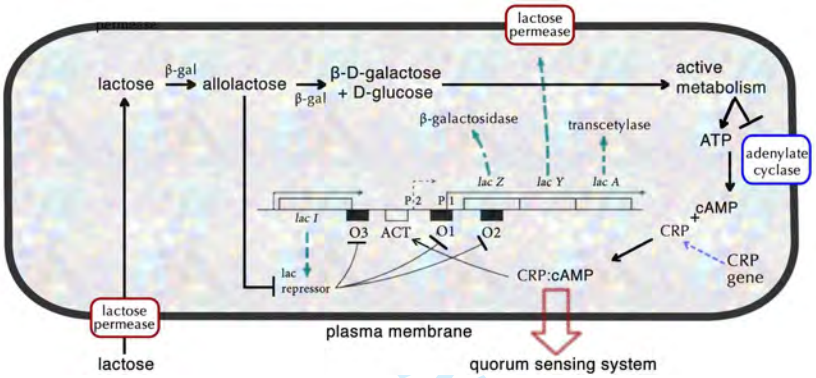
[2] to whom correspondence should be addressed: michael.klymkowsky@colorado.edu

Keywords: Molecular networks, Socratic tutorials, Graphing, Undergraduate Education

Abstract: Biological systems, from the molecular to the ecological, involve dynamic interaction networks. To examine student thinking about networks we used graphical responses, since they are easier to evaluate for implied, but unarticulated assumptions. Senior college level molecular biology students were presented with simple molecular level scenarios; surprisingly, most students failed to articulate the basic assumptions needed to generate reasonable graphical representations; their graphs often contradicted their explicit assumptions. We then developed a tiered Socratic tutorial based on leading questions designed to provoke metacognitive reflection. The activity is characterized by leading questions (prompts) designed to provoke meta-cognitive reflection. When applied in a group or individual setting, there was clear improvement in targeted areas. Our results highlight the promise of using graphical responses and Socratic prompts in a tutorial context as both a formative assessment for students and an informative feedback system for instructors, in part because graphical responses are relatively easy to evaluate for implied, but unarticulated assumptions.

Introduction

Dynamic interaction networks, whether at the molecular, cellular, developmental, physiological, or ecological level, are a universal feature of biological systems. These networks can be homeostatic, adaptive, evolving, or commonly a combination of all three. They typically involve a defined set of entities, i.e. molecules, genes, cells, tissues, organs, organisms, populations, etc., and a defined set of interactions. In molecular biology, these include activation, repression, assembly, modification, stabilization, degradation, and altered localization. Predicting the behavior of such networks and how they respond to perturbations is not a simple skill, yet this type of thinking lies at the heart of a robust understanding of what characterizes living systems and their behaviors. At the cellular and molecular levels, threshold effects, feed-back and feed-forward interactions, and redundant pathways, together with the stochastic nature molecular events, all make significant contributions to system behavior. The recent explosion in the number of direct observations [see 1, 2] and data-driven models of molecular level network behavior [see 3-7], has been followed by a growing recognition of the importance of understanding the behavior of regulatory circuits, particularly their emergent behaviors in the context of biological systems [8-12].

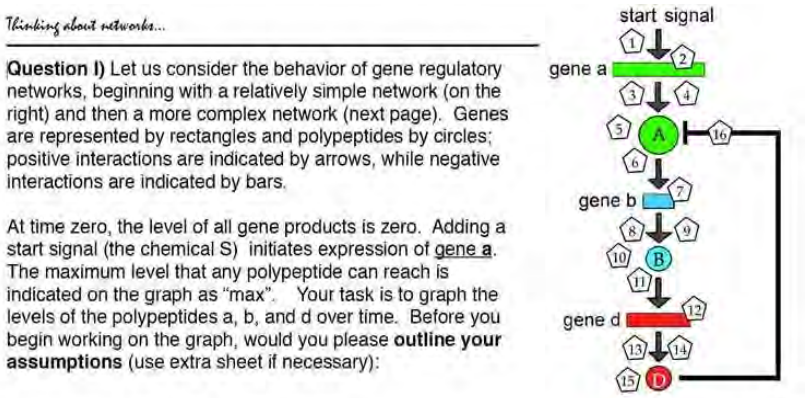


The challenges involved in reaching an understanding of molecular level networks can be illustrated by considering the *lac* operon of *E. coli* [13](**FIG. 1**). While often presented as a simple system, the *lac* operon is part of a complex network. The operon itself contains three genes, *lacZ*, *lacY*, and *lacA*. A distinct gene, *lacI*, encodes the lac repressor polypeptide. The functional lac repressor protein is a tetramer of LacI polypeptides. Lactose enters the cell via lactose permease, encoded by *lacY*, and is transformed in the allosteric regulator of the lac repressor, allolactose, via the catalytic action of β-galactosidase, encoded by *lacZ*. Allolactose binds to the lactose repressor, reducing its affinity for the operon's repressor binding sites, known as operators (O1, O2 and O3). The binding of the repressor blocks the binding and activation of DNA-dependent, RNA polymerase. A second level of regulation involves the activation of the Catabolite Repressor Protein (CRP) by the binding of cyclic AMP. Cyclic AMP is formed through

the action of a membrane bound protein, adenylate cyclase, acting on ATP. At normal metabolic levels, adenylate cyclase activity is inhibited and the intracellular cAMP concentration is low. As metabolic activity drops, cAMP levels rise, leading to increased levels of cAMP-bound CRP. In addition to turning on target genes, cAMP-bound CRP also regulates the quorum sensing system that mediates intercellular communication, swarming, and community behaviors [14]. An important point to consider is that to function the operon must be "leaky", that is, even in the absence of lactose the *lac* operon is periodically transcribed, LacY and LacZ polypeptides are synthesized, and low levels of lactose permease and β-galactosidase activities are present in the cell. Leakiness is insured in part by the low level of *lacI* expression and the stochastic nature of protein-DNA, RNA-protein and protein-protein interactions. Together, these lead to significant variation between cells in terms of RNA and protein levels [15]. As one can imagine, teaching the complexity of this paradigmatic system would be a daunting task. It requires a working understanding of the interconnectedness of this system would be a unique expert-level skill. However, in conjunction with the specific content knowledge of the *lac* operon there are the more general higher-order thinking abilities. Cellular and molecular biologists use principled reasoning and can analyze such systems by focusing on the entities, interactions between and among entities, and the functional activities that emerge to build representations the depict causal mechanisms within or between cells [16].

Given the importance of network thinking in biological systems, we set to answer the question; how do senior-level molecular biology majors perform on tasks related to representing systems of gene networks? Expert problem solvers often spontaneously create alternative representations when addressing a problem [17, 18]. One type of alternative representation involves transforming a problem into various "visualizations." We use the term as Tufte [19] defines it, that is, the systematic and focused visual display of information in the form of tables, graphs and diagrams. Graphing has been demonstrated to be an effective way to capture students' alternative conceptions and to help them learn in physics [20] and other areas [see 21 and references therein]. While there have been studies on how students interpret graphs and other types of visual representations [22-24], we found no published studies in which students were called upon to generate graphical or quantitative visualizations of gene network behavior. In addition to its inherent importance, the ability to analyze biological networks provides a relevant context within which students can apply and explore their understanding of the underlying principles at play at the molecular and cellular level.

To better understand student thinking about gene networks and their behavior we asked college-level molecular biology students nearing graduation to produce graphs that describe the behavior of simple systems, in terms of polypeptide concentrations over time. As a result of our findings we designed a tutorial activity that provided tiered feedback to students through sequential prompts for reflection aided at the progressive refinement of a student's produced visualization. We found that working through this activity improved student understanding of the underlying system, but that this improvement was restricted to the basic topics covered and did not transfer to negative feedback interactions. That said, we believe that a more extensive set of activities could extend student understanding of more complete and complex sets of network behaviors.



Methods and experimental design:

These studies were carried out with the informed consent of students and judged as exempt by the UC Boulder internal review board (IRB protocol 0304.09). We were able to examine three upper division classes at a large Western US university using either the preliminary network activity (**FIG. 2**)(supplied as supplement-1.pdf) or the Network-I Socratic tutorial (supplied as supplement-2.pdf). These materials were administered to students so as to integrate them with the teaching practices of the cooperating instructors. Student responses were collected and analyzed using rubrics created through an emergent coding scheme in the manner described by Haney *et al*. [25]. Emergent coding was used to create an initial rubric; this coding rubric was later revised by discussion between researchers and in the light of student responses to the Network-1 activity. One researcher grouped student responses into bins based on similar shaped graphs or similar written assumptions; this was performed in an iterative manner with input, evaluation, and agreement from the second researcher. A final common rubric was used to code the responses of the preliminary network activity and the Network-1 Socratic tutorial. The final rubrics contain the codes and descriptions for the assumptions derived from written statements (**Table 1**) and the implicit assumptions displayed in graphs (**Table 2**). An analysis of the system indicates that a rigorous representation of the system requires a consideration of 16 factors (**FIG. 2**). Assuming we are dealing with a prokaryotic system, lacking transcript processing or nuclear export, these can be collapsed into four major considerations: i) the effects of an activator/repressor on

the rate of transcription initiation (a function of concentration); ii) the time required for RNA and polypeptide synthesis (a function of transcript and open-reading frame lengths); iii) the half-lives of the RNA and polypeptides produced; and iv) the interaction affinities between proteins, and between proteins and DNA regulatory regions (again, a function of concentration). The implicit assumptions in the visualization (based on the shapes of the graphs drawn)(**FIG. 3**) were compared to the explicit assumptions in the written section to generate a measure of the internal consistency within students' responses (**Table 3**). One researcher coded all graphs and written assumptions using this rubric, while a second researcher coded 20 randomly selected students' responses for pages 1 and 5 of the Socratic tutorial to establish reliability. The rubric used to evaluate visualizations has an inter-reliability rating of 0.86 proportion in agreement and 0.65 as calculated by Cohen's Kappa, which is considered 'substantial' [26]. To evaluate the effect resulting from the Network-1 Socratic tutorial we identified the frequency of three common errors made by students between the first and last pages of the activity. These errors were selected because they provided the best comparison for change within the tutorial.

Preliminary network tutorial and its administration: The initial group of 102 students enrolled in a major's "capstone" course (group 1) was given the preliminary network activity as homework. The majority of these students were two months away from completing their required course series and within a year or less of graduation. The activity presented a three-component gene network schematic of a negative feedback loop initiated by the introduction of a signal (**FIG. 2**). Students were asked to generate a graph for the concentration of each of the three polypeptides as a function of time, following the addition of the "start signal", which initiated transcription of gene a. Students were also asked to outline their assumptions. Participants were given two days to complete the homework and responses were collected at the start of class.

Network-I tutorial design and administration: In the light of student responses, we designed a scaffolded [27] five-page Socratic "Network-I" tutorial that could be administered either on paper or on-line [1]. When we indicate that the Network 1 tutorial was scaffolded, we mean that it relies on simplifying problems and presenting them in a series, such that each component of the tutorial presents the students with a manageable cognitive load [27]. The tutorial was designed to target the assumptions that are necessary, but generally found to be lacking in students' responses to the initial

[1] The original Network-I activity can be accessed at http://besocratic.colorado.edu/network/network-page1.htm

network activity. Page 1 begins with a single gene, *gene a*, making a single polypeptide (A) and asks students to graph the concentration of A as a function of time after an "activating" signal is introduced, as well as to define the term "gene expression." Page 2 asks students to reflect on the previous graph and on the time delays associated with transcription and translation; it asks students to reconsider their graphs in the light of these considerations. Page 3 asks students to consider the rate of degradation of polypeptide A, and presents a new scenario in which the activating signal (S) is withdrawn sometime after its introduction. On page 4, students are asked to examine a scenario in which polypeptide A activates the expression of a second gene, *gene b*, and to graph the concentrations of both A and B polypeptides. Here we ask explicitly how *gene b's* transcript and polypeptide lengths influence the time it takes to produce polypeptide B. The final page, page 5, adds a negative feedback interaction between the B and A polypeptides. Students are given the option of choosing whether this interaction involves the enhanced degradation of polypeptide A or the inhibition of A's activity, that is, its ability to activate the transcription of *gene b*. They are also asked to comment on how their graph would differ if they had chosen the alternative interaction mechanism. Throughout the tutorial we used prompts (questions about assumptions) and perturbations (changes in the problem) to encourage students to consider the implications of their assumptions and possible alternatives, without explicitly telling them whether they were right or wrong. The goal is to encourage students to make their assumptions explicit, and so help them recognize the assumptions that they are making.

Students in groups 2 and 3 were presented with a paper-based version of the "Network-I" tutorial in two different settings. Group 2 (N=34; students enrolled in a major's molecular biology "core course") were given the Socratic tutorial as an out-of-class extra credit homework assignment. This administration was consistent with the teaching practices of the professor. Group 3 (N=90; students enrolled in a major's "capstone" course and nearing graduation) were given the same Socratic tutorial as an in-class group activity (31 groups of 2-3 students each). Written and graphic responses to the tutorial activity were analyzed using the coding rubric presented in Tables 1 and 2 respectively.

Results:

Preliminary Network activity results: Analysis of written and graphical responses to the preliminary network activity revealed a number of common features. First, when asked to explicitly state their assumptions, most students listed relatively few of those needed to accurately describe the system (**Table 4**). Of the four classes of assumptions needed (see above), ~50% of students explicitly stated only one, ~ 30% stated two, and ~12% stated more than two. These articulated assumptions were distributed unevenly. With regards to signal duration, polypeptide (or RNA) half-life, or the time required for transcription and translation, each assumption was explicitly made by less than 25% of the students. Explicit assumptions centered primarily on the nature of the negative feedback interaction between polypeptides D and A (~80%). This suggests that students recognized the inherent ambiguity of the schematic in regards to the nature of this interaction, but that other equally important aspects of the interaction network were overlooked.

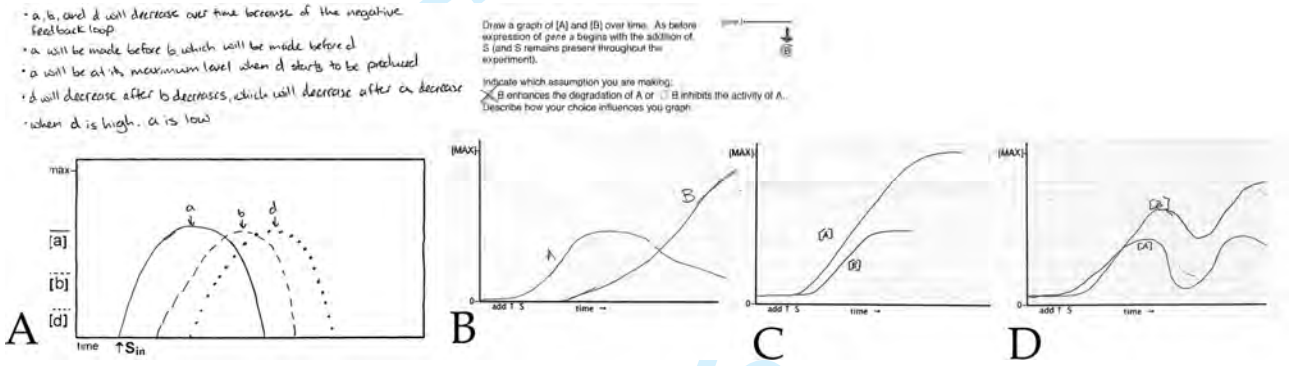
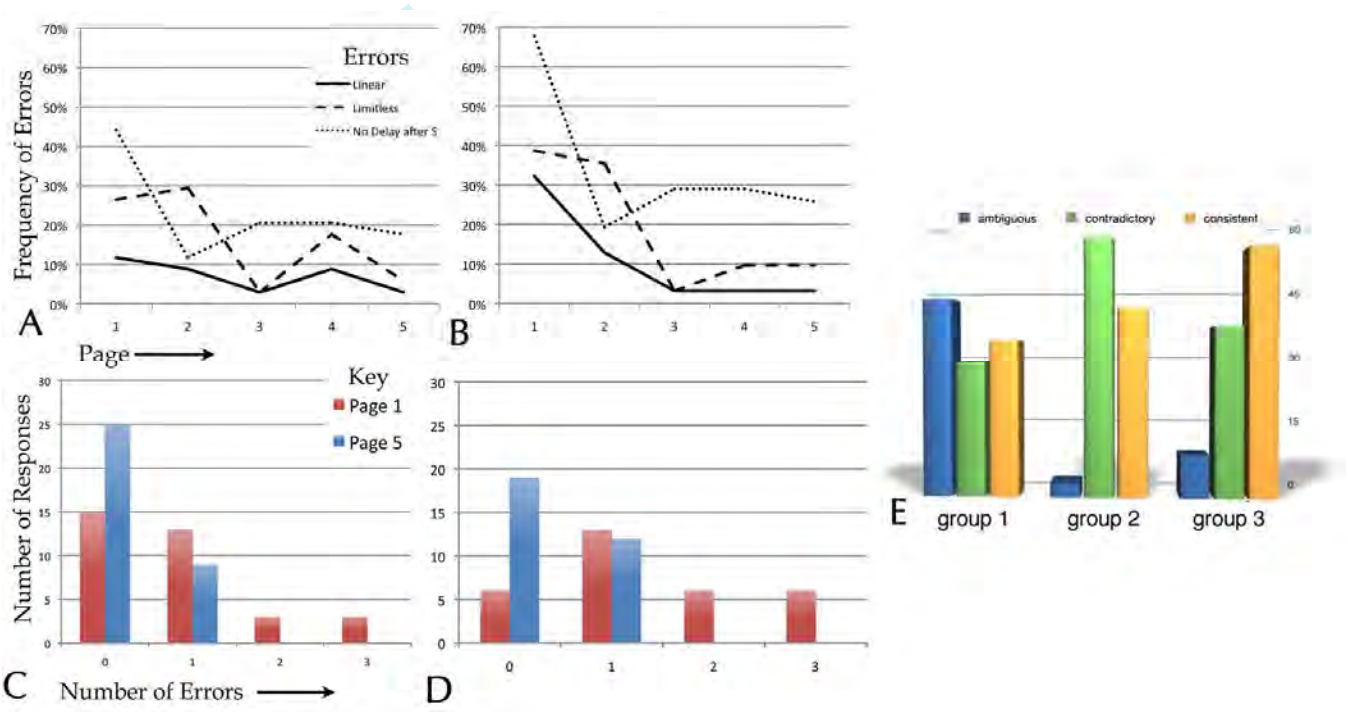


Figure 3: Representative student response to the (A) preliminary network assessment and (B, C, & D) the final page of the Network-1 Socratic tutorial.

In contrast to student's textual responses, which often contain multiple, often internally inconsistent, incorrect, or irrelevant ideas [28], graphical responses were, superficially at least, unambiguous (**Table 5**). They suggested a failure to consider the delay between the onset of transcription and the appearance of polypeptide, inferred by an instantaneous rise in polypeptide concentration following the addition of the activator S (~85% of all responses). Together with linear representations of changing polypeptide concentrations, there appeared to be a common failure to consider the effects of RNA and polypeptide turnover on polypeptide accumulation. Of the ~80% of students who made some explicit assumptions about the nature of the negative feedback interaction (D ⊣ A), ~78% of the graphic responses implied that D's interaction with A enhanced A's degradation

(indicated by a fall in [A] in the presence of D), but only ~13% stated this assumption explicitly. The extent of discordance between graphical and written responses is often dramatic (**FIG. 3A**). Note that this student's written assumptions address the negative feedback of D on A, but fail to indicate its molecular level effect. In contrast, their graph suggests that D induces the enhanced degradation of A. The graph also indicates no time delay between the introduction of S and the appearance of A, and suggests that both B and D are rapidly degraded (since they disappear "quickly", as quickly as A) following D's accumulation and negative interaction with A. Also, there is the implication that S has disappeared from the system, since A does not reappear after D disappears. Such, "three-hump" graphs were common.



Across all five pages of the tutorial, three common errors were tracked: limitless growth of the curve, linear graphs, and absence of a delay between the signal input and rise in the concentration of the encoded polypeptide (**FIG.4**). The relative frequency of errors for each category decreased dramatically immediately after targeting, but then rose slightly on subsequent graphs (**FIG. 4 A,B**); for example, considering the time-delay between the signal and product was targeted on page two and this impacted the graphs on page 2 for both groups, but a small percentage returned to errors in time delay on future pages. Note that limitless growth was not a common feature of the original activity (**Table 5**), but emerged when the system was simplified. This was apparently due to student's overlooking protein degradation, a conclusion supported by analysis of reflective question responses (data not shown). It is worth noting that an informal survey of instructors suggests that degradative processes

are rarely presented within the curriculum. Group 2 displayed an average of 0.82 errors (out of the possible 3) on the first page and displayed an average of 0.26 errors (out of the same 3) on the final page. Group 3 began with 1.39 average errors and ended with 0.39. Histograms show the change in the distribution in errors (**FIG. 4 C,D**). A Wilcoxon signed-rank test was used to evaluate the change in number of errors for the non-parametric data and suggests that the decreased errors between the beginning and end of the tutorial were significant for both groups (Group 2: $W_+ = 32$, $W_- = 158$, $N_{pairs}=19$, $p<0.01$ and Group 3: $W_+ =12$, $W_- = 241$, $N_{pairs} = 22$, $p<0.005$). On the final page of the activity, students were asked to interpret a negative feedback loop somewhat similar to that in the preliminary network activity. In contrast to the original "survey" activity where participants listed assumptions openly, in the network tutorial, the students were asked to make an explicit choice of the type of negative interaction, that is whether B acted to inhibit the activity of A or increased the rate of its degradation. Response analysis revealed serious inconsistencies between their explicit choice and the graphs students generated (**FIG. 4 E**). For example, if a student selected inhibition of activity as the negative interaction but graphed degradation (indicated by a falling concentration of the first product upon the presence of the second), then the response was considered inconsistent. Likewise, if the explicit assumption made by the student was implicitly represented in the graph, then the graph was considered consistent. For graphs that were ambiguous or when interactions were not indicated, the response was considered ambiguous. These observations suggest that students have a poorly defined idea of how degrading a polypeptide and inhibiting its activity differ. We did not measure the long term persistence of the apparent improvements students display within the tutorial, in part because the intervention was restricted to a single class period, but clearly this is an important point that needs to be addressed in future studies.

Discussion:

Molecular biology students are commonly taught the basics of gene expression, including some exposure to the regulatory interactions that underlie life in general and specific biological systems in particular. However, our preliminary network activity revealed that molecular biology students nearing graduation appear unprepared to critically analyze and visualize the task with which they were presented; even though it was quite simple compared to typical molecular networks found in nature [29, 30], including the oft taught *lac* operon (see above). Few students were able to articulate the majority of assumptions needed to address the problem. We propose two non-exclusionary models to explain student's responses. **Model 1**: Students do not have either the skills, or sufficient confidence in their skills, to make appropriate assumptions. This implies that their ability to transfer appropriate knowledge is poorly developed. By transfer we mean extending what is learned in an initial context to other contexts through recognition of larger principles [31]. **Model 2**: The novelty of the problem (and the task involved) may have deterred students from responding fully. It is possible that even simple problems lead to cognitive overload. It is certainly the case that few students had "practiced" on this type of question during the course of their education. Based on preliminary discussions with students, we suspect that both models are in play, but in-depth interviews in the context of longer interactions with network activities are needed to fully map out student thinking in this area. That said, the diverse and often ambiguous and self-contradictory graphs produced suggest that students have enormous trouble not only with the complexities of the ideas, but also representing them in graphical or pictorial formats.

Based on students' responses to the initial survey activity, we examined their responses to a simpler, tiered scenario. In the course of this activity, students admitted to overlooking specific qualities when prompted as part of the tutorial *(data not shown)*, but could properly correct graphs after reflection. This suggests that they possess some of the necessary "raw" content knowledge and that targeting prompts could initiate the transfer of their knowledge to the task of generating accurate graphical representations. We have seen similar "correctable responses" in the course of graphical activities carried out on the generation of Lewis structure [24], coupled chemical reactions and the effects of introduced pathogens on population dynamics in the context of the introductory major's molecular biology course, MCDB 1150 Biofundamentals (data not shown). The improvements displayed by students in the course of the network tutorial were clearly not readily transferable to new

tasks, as witnessed by their apparent difficulties when they were confronted with a variant of the task, such as a negative feedback interaction. This begs the question of whether the relevant content from molecular biology/biochemistry is developed enough for most students to use for thinking about gene networks. Thinking about molecular and cellular biology in a systematic way is a somewhat recent trend, which continues to rise, but the degree of knowledge (i.e. where along the curriculum) a student needs before being able to analyze networks in a useful way is uncertain. Additionally, it is unclear who is qualified to teach these topics and what are the universally appreciated concepts for this field in flux. The Network-I tutorial used here is conceptually quite limited, and it is easy to imagine how more complex topics, such as the influence of concentration changes, interaction affinities, cooperative thresholds and stochastic interactions, could be incorporated into more physiologically realistic follow-on modules.

It could be argued that students do understand the entities, interactions and activities involved in gene networks but that the tutorials failed to reveal their understanding although this begs the question as to what students are supposed to be able to do with their knowledge [33]. Is not having been explicitly taught or trained to solve a novel problem a valid explanation for students' observed struggles? A New Biology for the 21st Century [34] focuses on the challenges that biologists will face in the near future; most of these problems do not have simple linear solutions. Inventive, interdisciplinary approaches must be created for the students we are instructing. Yet, few university level programs appear to have integrated mathematical and modeling skills into their curricular design or pedagogical practice [35, 36]. It is our working hypothesis that a more extensive and integrated use of Socratic tutorials will be key in producing the necessary improvements in students transfer skills by provoking and encouraging meta-cognitive reflection, and in a group context, meta-cognitive discussions.

Acknowledgements*: We thank Alex Wyman and Denise Kilates for contributing to data analysis, and colleagues for access to their classes. The work was supported primarily by a grant from ASSETT (UC Boulder) to M.K. and stimulated by work funded on gene network interactions in Xenopus (NIH* GM84133*.) C.T. was supported in part by UC Boulder's Science Education Initiative, MMC and MWK acknowledge support from the NSF (*DUE -816692 and DUE-1043707).

Bibliography

1. Suter, D.M., et al. (2011) Mammalian genes are transcribed with widely different bursting kinetics. Science. **332**: pp. 472-474.
2. Larson, D.R., et al. (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science. **332**: pp. 475-478.
3. Davidson, E.H., D.R. McClay, and L. Hood (2003) Regulatory gene networks and the properties of the developmental process. Proc Natl Acad Sci U S A. **100**: pp. 1475-1480.
4. Swain, P.S., M.B. Elowitz, and E.D. Siggia (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A. **99**: pp. 12795-12800.
5. Elowitz, M.B. and S. Leibler (2000) A synthetic oscillatory network of transcriptional regulators. Nature. **403**: pp. 335-338.
6. Bolouri, H. and E.H. Davidson (2003) Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc Natl Acad Sci U S A. **100**: pp. 9371-9376.
7. Bieler, J., C. Pozzorini, and F. Naef (2011) Whole-embryo modeling of early segmentation in Drosophila identifies robust and fragile expression domains. Biophys J. **101**: pp. 287-296.
8. Maheshri, N. and E.K. O'Shea (2007) Living with noisy genes: how cells function reliably with inherent variability in gene expression. Annu Rev Biophys Biomol Struct, . **36**: pp. 413-434.
9. Pedraza, J.M. and J. Paulsson (2008) Effects of molecular memory and bursting on fluctuations in gene expression. Science. **319**: pp. 339-43.
10. Raj, A. and A. van Oudenaarden (2009) Single-molecule approaches to stochastic gene expression. Annu Rev Biophys. **38**: pp. 255-270.
11. Mettetal, J.T., et al. (2006) Predicting stochastic gene expression dynamics in single cells. Proc Natl Acad Sci U S A. **103**: pp. 7304-7309.
12. Rinott, R., A. Jaimovich, and N. Friedman (2011) Exploring transcription regulation through cell-to-cell variability. Proc Natl Acad Sci U S A, **108**: pp. 6329-34.
13. Savageau, M.A. (2011) Design of the lac gene circuit revisited. Math Biosci, . **231**: pp. 19-38.
14. Bochkareva, E., et al. (1996) Chaperonin-promoted post-translational membrane insertion of a multispanning membrane protein lactose permease. J Biol Chem, **271**: pp. 22256-22261.
15. Zhou, X., X. Meng, and B. Sun (2008) An EAL domain protein and cyclic AMP contribute to the interaction between the two quorum sensing systems in Escherichia coli. Cell Res. **18**: pp. 937-948.

16. Taniguchi, Y., et al. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. **329**: pp. 533-538.
17. Martin, L. and D.L. Schwartz (2009) Prospective Adaptation in the Use of External Representations. Cognition and Instruction. **24**: pp. 370-400.
18. Bodner, G.M. and S. Domin (2002) Mental Models: The Role of Representations in Problem Solving in Chemistry. Univ. Chem. Educ. **4**: pp. 24-30.
19. Tufte, E.R., (2001) Visualization in Science Education. 2nd ed., Chesire, CT: Graphics Press.
20. Merhar, V.K., G. Planinsic, and M. Cepic, (2009) Sketching graphs—an efficient way of probing students' conceptions. Eur. J. Phys. **30**: pp. 163–175
21. Gomez, K., L. Lyons, J. Radinsky Eds. (2010) *Proceedings of the 9th International Conference of the Learning Sciences - Volume 2*. (International Society of the Learning Sciences Chicago, Il, June 29 to July 2, 2010) pp. 165-171.
22. Beichner, R.J. (1994) Testing student interpretation of kinematics graphs. American Journal of Physics. **62**: pp. 750-762.
23. Picone, C., et al. (2007) Assessing gains in undergraduate students' abilities to analyze graphical data. Teaching Issues and Experiments in Ecology. **5**.
24. Cooper, M.M., et al. (2010) Lost in Lewis Structures: an investigation of student difficulties in developing representational competence. J. Chem. Educ. **87**: pp. 869–874.
25. Haney, W., M. Russel, and D. Bebell (2004) Drawing on Education: Using Drawings to Document Schooling and Support Change. Harvard Educational Review. **74**: pp. 241-272
26. Stemler, S. (2001). An overview of content analysis. Practical Assessment, Research & Evaluation, *7*:Retrieved October 17, 2011 from http://PAREonline.net/getvn.asp?v=7&n=17
27. Reiser, B.J. (2004) Scaffolding complex learning: the mechanisms of structuring and problematizing student work. J. Learning Sciences **13**: pp. 273-304.
28. Wright, J.C., et al. (1998) A novel strategy for assessing the effects of curricular reform on student competence. J. Chem. Educ. **75**: pp. 986-992.
29. Negre, N., et al. (2011) A cis-regulatory map of the Drosophila genome. Nature **471**(7339): pp. 527-531.
30. Praneenararat, T., T. Takagi, and W. Iwasaki (2011) Interactive, multiscale navigation of large and complicated biological networks. Bioinformatics **27**(8): pp. 1121-1127.
31. Bransford, J.D., A.L. Brown, and R.R. Cocking, eds. (2001) How people learn: Brain, Mind, Experience and School. National Academy Press: Washington, D.C.

32. Speth, E.B., et al. (2010) 1, 2, 3, 4: infusing quantitative literacy into introductory biology. CBE Life Sci Educ. **9**: pp. 323-332.
33. McClymer, J.F. and L.Z. Knowles (1992) Ersatz learning, inauthentic testing. J. Excell. Coll. Teaching. 3: pp. 33-50.
34. National Research Council (2009) A New Biology for the 21st Century. National Academy Press: Washington, D.C.
35. Aegerter-Wilmsen, T. and T. Bisseling (2005) Biology by numbers--introducing quantitation into life science education. PLoS Biol **3**(1): pp. e1.
36. Klymkowsky, M.W. (2009) Make room for computing. Science . **326**(5950): pp. 227.

TABLE 1: Assumption coding rubric used to code written assumptions in the preliminary tutorial

Code	Description of code
	Negative feedback
Regulates Activity	The response indicates that the final product is inhibiting the activity of the first product.
Regulates Stability	The response indicates that the final product is enhancing the degradation of the first product.
Regulates Production	The response indicates that the final product is preventing production of the first product.
Arrows up/down	The response contains arrows to symbolize rising and falling.
Ambiguous	The response acknowledges negative interaction but does indicate nature.
Absent	The response has no mention of the above five.
	Half-life of the gene products
Mentioned	The response considers half-lives, proteins degradation, or equivalent.
Stable	The response states that degradation of the protein does not occur.
Absent	The response has no mention of the above two.
	Signal persistence
Continual	The response states that the signal is present constantly or the first gene is continually activated.
Pulse	The response states the signal will be present for a limited time.
Ambiguous	The response acknowledges that signal has an affect but does not state whether continual or pulsed.
Absent	The response has no mention of the above three.
	Time for Transcription and Translation
Mentioned	The response not only acknowledges transcription and/or translation, but also the time for it to occur (a rate of production is also acceptable).
Absent	The response has no mention of the above

TABLE 2: Graphic coding rubric used to for all five response aspects when applicable.

Code	Description of code
	Signal Persistence Graph
Continuous	The graph shows sustained growth, plateauing or oscillating
Pulsed	The graph displays a single growth and falling period
Ambiguous	Not distinguishable with the above two
	Negative Feedback
Degradation	The slope of first product's concentration declines upon the accumulation of negatively acting product (may represent preventing production or degradation.)
Inhibition	The slope of first product's concentration remains unchanged but downstream targets are affected upon accumulation of negatively acting product.
Ambiguous	Not distinguishable with the above two
	Delay between Signal to Transcription and Translation of Product
Before S	A raise in first product (above basal level) occurs before signal introduced
After S	A raise in first product (above basal level) occurs some time after signal introduction
At S	A raise in first product (above basal level) occurs at time the signal is introduced
Ambiguous	The initial raise is not distinguishable with the above three
	Limitless Growth
Yes	The graph indicates protein concentration grows continually.
No	The graph shows a plateau or negative slope is displayed.
Ambiguous	The slope is not distinguishable with the above two
	Linear Behavior
Yes	The slopes are rigid and composed of straight lines.
No	The slopes are smooth and from curves.
Ambiguous	The slopes are mixed or not distinguished with the above two

Table 3. Coding of the sample student graphs from figure 3 to demonstrate the graphic coding rubric of Table 2.

Student Sample	Signal	Negative interaction	Initial rise in polypeptide A	Linear Behavior	Limitless Growth
	Continuous Pulse Ambiguous	Inhibit Degrade Ambiguous	Before Signal At Signal After Signal Ambiguous	Yes No Ambiguous	Yes No Ambiguous
A	Pulse	Degrade	At Signal	No	No
B	Continuous	Degrade	At Signal	No	No
C	Continuous	Inactivate	After Signal	No	No
D	Continuous	Degrade	After Signal	No	No

Table 4: Absolute and relative frequency of students' written assumptions for the preliminary assessment with Group 1 (N=102)

Four needed assumptions with codes	N (%)
Negative feedback	
Regulates activity	26 (25%)
Regulates stability	4 (4%)
Ambiguous	28 (27%)
Regulates synthesis	9 (9%)
Arrow up/down relation	17 (17%)
Absent	18 (18%)
Half-life of Biomolecules	
Mentioned	11 (11%)
Assumed stable	10 (10%)
Absent	81 (79%)
Signal Persistence	
Pulse	10 (10%)
Continuous	13 (13%)
Absent	79 (77%)
Time for Transcription and Translation	
Mentioned	25 (25%)
Absent	77 (75%)

Table 5: Absolute and relative frequencies of coded graph attributes from preliminary assessment for Group 1 (N=102)

Five qualities of graphs with codes	N (%)
Signal Persistence	
Pulse	42 (41%)
Continual	27 (26%)
Ambiguous	33 (32%)
Delay between Signal to Transcription and Translation of Product	
Before S	27 (26%)
At S	60 (59%)
After S	15 (15%)
Linear Behavior	
Yes	19 (19%)
No	82 (80%)
Ambiguous	1 (1%)
Negative Feedback	
Degradation	80 (78%)
Inhibition	7 (7%)
Ambiguous	15 (15%)
Limitless Growth	
Yes	4 (4%)
No	96 (96%)
Ambiguous	0 (0%)