Kinesthetic Pathways: A Tabletop Visualization to Support Discovery in Systems Biology

Andy Wu, Jung-Bin Yim, Eric Caspary, Ali Mazalek,
Sanjay Chandrasekharan, Nancy J. Nersessian
GVU Center, Georgia Institute of Technology
TSRB, 85 5th St. NW
Atlanta, GA 30318, USA
{ andywu, jyim, ecaspary3, mazalek }@gatech.edu, { sanjayan, nancyn }@cc.gatech.edu

ABSTRACT
We report on an ethnographic study of the work practice and discovery process in a systems biology lab, and outline a tabletop visualization that was developed based on this study, in collaboration with the researchers. The feedback from the researchers on the current prototype is presented, and ongoing revisions are outlined. We conclude with some of the challenges involved in developing such tangible visualizations for discovery.

Author Keywords
Pathways, systems biology, simulation, modeling, tabletop, HCI, tangible user interface, kinesthetic interaction

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces---*input devices and strategies, interaction styles, user-centered design*; I.6.7 [Simulation and Modeling]: Simulation Support Systems---environments.

General Terms
Design, Experimentation

INTRODUCTION

In August 2010, *Nature* published a paper where roughly 200,000 players of a videogame were included as authors. The paper reported how the re-representation of the protein-folding problem as a multi-player videogame, Foldit, allowed collaborative solving of the problem by ordinary people. It also proposed that such harnessing of people's visual and spatial reasoning abilities could be a new method to solve computationally-challenging scientific problems.

Much of the attention on Foldit has focused on the "citizen science" aspect [2], but equally important is its contribution to an understanding of how perception and action systems interact in creativity and cognition, particularly in relation to scientific discovery. A central component of the success of Foldit is its direct manipulation interface, which allows players to grasp and pull and move and twist different protein strands. This type of interface, where the user actively controls and explores objects on screen to develop an understanding, is common in educational applications in science (e.g. PhET [16]). But most scientific visualizations do not seek to support this type of kinesthetic interaction-based discovery [4]. Rather, they seek to represent data in new ways, and the insight is expected to come from the different visual perspectives on the data.

One of the reasons for the success of Foldit is the use of the kinesthetic interaction, which allows the motor system to work in tandem with the visual and imagination systems while solving spatial problems (for a detailed model of this interaction, see [1]. In this paper, we report a prototype system where we seek to extend such kinesthetic interaction for discovery to systems biology, particularly to the modeling and simulation of metabolic systems. This is a more complex and abstract problem than protein-folding, as the entities involved do not have spatial features. In the following section, we provide an outline of the practice and problems in a systems biology lab, based on a two-year ethnographic study. The next section outlines the prototype we have developed based on this study, and the feedback we have received from the lab members on the prototype. We conclude with some of the challenges involved in developing systems to support discovery in this domain and future work. The images in this paper are retrieved from different stages of the prototype. Since we have been trying different types of visualizations techniques, some of them are different from the current working system.

PRACTICE IN A SYSTEMS BIOLOGY LAB

In our current project we are studying problem-solving practices in two integrative systems biology labs. We focus here on one lab that does only computational modeling ("Lab G"). The modelers in Lab G are mainly from engineering fields, but work on building computational models of biochemical pathways, to simulate and understand phenomena as varied as Parkinson's disease, plant systems for bio-fuels, and atherosclerosis. The problems Lab G modelers work on are provided by outside experimental collaborators, who see modeling primarily as a method for identifying key experiments of scientific or

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
C&C'11, November 3–6, 2011, Atlanta, Georgia, USA.
Copyright 2011 ACM 978-1-4503-0820-5/11/11...$10.00.

commercial importance. The collaborators provide experimental data for modeling, and also data needed for developing pathway diagrams, or validating the model.

The modeling process can be classified into three phases – building, fitting/testing and perturbation – even though these phases overlap. In general, the model is built up in a 'ripple' fashion, with small models in the beginning, each going through these three phases. More elements are then added to these models, and these complex models then go through the three phases.

The Modeling Process

Lab G researchers mostly build ordinary differential equation (ODE) models of metabolic systems, which capture how the concentration levels of different metabolites in a given biological pathway change over time. The first step in this building process is the development of a pathway diagram, which shows the main reactions involved. The pathway diagram also captures positive and negative regulation effects, which specify how the presence of different metabolites has a positive or negative influence on different reactions (see Figure 1). The experimental collaborators usually provide a rough diagram of the pathway. But the modelers, who mostly come from engineering backgrounds, have to estimate the details of the pathway by themselves, particularly values of parameters related to metabolites, such as speed of change (kinetic order) and concentration level (rate constant), which are usually not measured by experimenters. This information is available in rough form (with varying degrees of reliability) from online databases, but most often these values need to be estimated, usually through iterative testing of the model, using a range of numbers as parameter values.

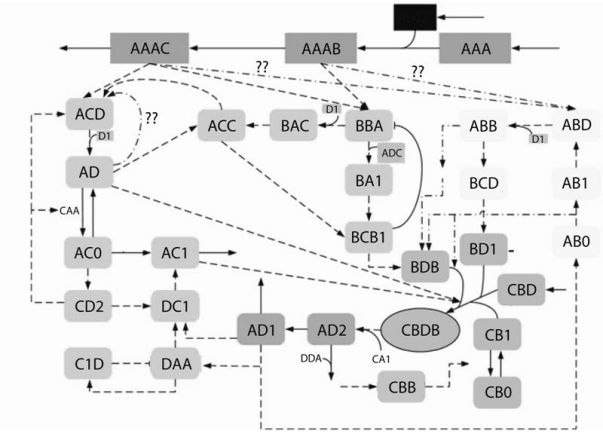

Figure 1. A sample pathway diagram. Metabolite names are replaced with alphabet characters. The dark lines indicate connections where material moves across nodes, the dotted lines indicate regulatory connections. The question marks show connections that are postulated by the modeler.

Modelers also add some components to the pathway, usually metabolites that are known to interact with the network provided by the experimenters. These components are found by reading and searching biology journal articles and databases related to the problem being modeled, and also based on results from preliminary models. Even though experimentalists provide much of the pathway, these additions based on literature searches are required, because the provided pathway does not identify all the components, and the regulatory influences they have on the reaction.

The pathway developed by the modeler thus brings together pieces of information that are spread over a wide set of papers, databases, and unreported data from the experimentalists. This pathway is usually trimmed, based on some simplifying assumptions, mostly to lower the mathematical and computational complexity involved in numerically solving the differential equations. After the trimming, differential equations are generated to capture the trimmed pathway. A variable is used to represent the metabolite, while the speed of its change (kinetic order) and its concentration level (rate constant) are represented by parameters, which can take many values. The next step involves estimating values for these parameters, and these values are then used to initialize simulations of the models. The simulation results are then compared to actual experimental results, to judge the 'fit' of the model.

Usually, modelers split available experimental data into two, one set is used to develop the model (training data), and the other set is used to test the completed model (test data). When the model data do not fit the test data, the parameters are "tuned" to get model results that fit. Once the model fits the test data, it is run through a series of diagnostic tests, such as stability (e.g. does not crash for a range of values), sensitivity (e.g. input is proportional to output) and consistency (e.g. reactant material is not lost or added). If the diagnostic tests fail, the parameters are tuned again, and in some cases, the pathway changed, until the model meets both the fit and diagnostic tests. Figure 2 provides a broad outline of the modeling process.

Lab G models do not have real-time dynamic visualizations. Parameter values are changed manually or using scripts. Results for different parameter values are compared using a deck of graphs, where each graph plots the concentration value of an element in the pathway across time. The modeler uses these graphs while discussing the model with collaborators and other team members. A significant chunk of the parameter estimation problem is tackled using optimization algorithms (such as simulated annealing and genetic algorithms), which automatically do the 'tuning' of parameters, by comparing the output values (for different parameter inputs) against a desired value. Importantly, the linear workflow suggested by the above description is very deceptive – the modeling process is highly iterative. For instance, to develop the pathway diagram, preliminary models are built using the network provided by the experimenters, and these are run using tentative parameter values, and the generated model data

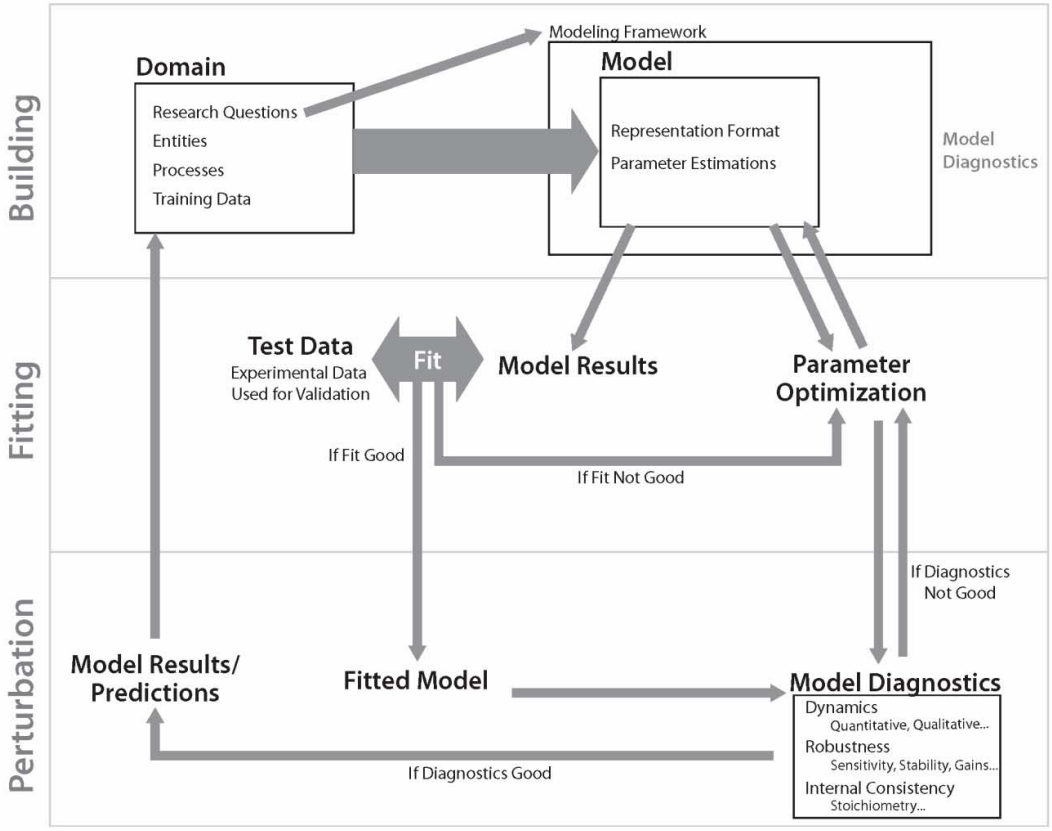

Figure 2. An outline of the modeling process in Lab G. Note that the 'building' phase involves both 'fitting' (of the training data) and perturbation (model diagnostics).

are fit to the training data. The parameter values are then revised based on this fit. If the model data do not fit after a large number of these parameter revisions – particularly if the data trends are the exact opposite of experimental data – the modeler will add some components to the network, based on elements that are known (in the literature) to be related to the pathway. There is also an instance where the modeling led to the discovery that an element (named X) in another pathway was influencing a biofuel pathway. Later experimental work by collaborators identified a candidate element for X. Thus, *some of the model's components – elements and values – are set by building and running the model itself.* These pathway revisions, and their justifications, are discussed with the collaborators, and if a revision is considered "reasonable" by the experimenter, it becomes a stable component of the pathway. This pathway identification process is usually "bottom-up," and creates a "composite" network, made up of parameter values and metabolites extracted from experiments in different species, different cell lines etc. This composite is usually unique, and does not exist anywhere else in the literature.

One of central problems the lab members face is the unavailability of rich, and dependable, data. In modeling, data are used for many purposes. One central use of data is to establish that the model captures a possible biological mechanism, and this is done by showing that the model's output matches the output from experiments (fitting data). A second use of data is to tune parameter values during the training phase of building the model. The fit with the experimental data from each training simulation can indicate how the model parameters need to be changed, to generate model data that fit the training data. This use is highly dependent on the type of data available. Most of the time, the available data are 'qualitative' in nature – usually how an experimental manipulation led to a change in a metabolite level from a baseline. Mostly, this is reported as a single data point, indicating the level going up or down, and then holding steady. However, when this type of "steady-state" data fits the results of the model, this fit does not indicate that the model has captured the biological mechanism. A range of parameter values can generate model results that fit such sparse data – the fit is not unique. Further, since the pathway is an approximation, the modeler is uncertain as to whether the lack of a unique and accurate solution is due to poor estimation of parameters, or because some elements are missing from her pathway.

Example Cases

As an example instance of modeling in this lab, consider G12, an electrical engineer by training, who is modeling atherosclerosis. When she started modeling, she had no background on atherosclerosis. Her experimental collaborators provided her a rough outline of the pathway, and she learned more about the pathway by reading papers. The initial papers were from the collaborating lab, but then she spread out using the reference lists of those papers. The data available were mostly steady-state data. Once she had read a number of papers, she started building rudimentary computer models and testing these using available data. She then added some components to the model based on connections in the literature. It is worth noting here that while her problem mostly concerned endothelial cells, some of her parameters were taken from experiments with neurons, a very different cell class, and a domain of research (neuroscience) that is not usually connected to research in endothelial cells. Some of her additions were considered "reasonable" by her experimental collaborators.

Estimating parameter values for her model was a tough problem, since the data were sparse. To get parameter values that generated model data that fit the training data, she ran a large number of simulations and compared the results with the training data. Finally, she arrived at a set of values that generated data that roughly matched the training data. Once this was done, she tested her model against the test data, and got a rough fit there as well. Based on this fit, she generated a number of predictions from the model, by changing the parameter values. Her experimental collaborators would test some of these predictions.

A second example is G11, a physicist with a Ph.D. in bioinformatics, who is modeling Parkinson's Disease (PD), particularly the dopamine pathway. He works closely with a neuroscience lab in a neighboring university, and he has access to all their data (excel files), which he uses to build his models. He has also suggested experiments based on his models, and the collaborators are currently testing some of these predictions. He, however, thinks the experiments take a long time, and as he is waiting for the results, he is trying purely mathematical approaches (such as Monte Carlo simulations) to develop a good model of PD. Apart from this 'phase lag' between modeling and experimentation, he also finds the experimental methods limiting. For instance, he can manipulate more than three variables in his model, and give very specific values to his variables (such as concentration of a metabolite in one stage of a reaction). Whereas in experiments, manipulating more than three variables is impossible, and there is no good way to ensure that concentration of the material in the cell is the same as what was specified in the model.

G10 has an electrical engineering background, and he is developing a model of the biosynthesis of lignin (a critical component of plant cell walls). The objective of the model is to provide insight on how the composition of lignin could be manipulated in such a way that the lignin breaks down easily to produce biofuel. A related objective is supporting the development of transgenic species where the lignin composition is optimal for biofuel production. This work is different from other work in the lab in two respects. One, G10 works with non-academic collaborators, particularly a company and a non-governmental organization. Secondly, G10's work deals with plant cells, while all others work with animal cells. In recent work, his modeling work showed that a product of another pathway (identified as X) has significant regulatory influence in the production of lignin. His collaborator has identified a candidate molecule that could play this regulatory role. Also, his modeling work suggests that some of the reactions are reversible, even though they have traditionally been considered as just forward reactions. These discoveries suggest that modeling has a self-illuminating effect, where the model can provide insight into the very pathways on which the model is based.

PATHWAYS

Based on this detailed understanding of the scientific practice in Lab G, we are working to develop a tangible tabletop visualization system to support discovery in systems biology, with a focus on developing an interface that would allow the researchers to explore parameters in a kinesthetic fashion, and thus estimate the parameters that would best fit model data with experimental data. The system also seeks to support modeling by experimentalists, who may not have a detailed understanding of the mathematical techniques, and also efficient collaboration between modelers and biologists.

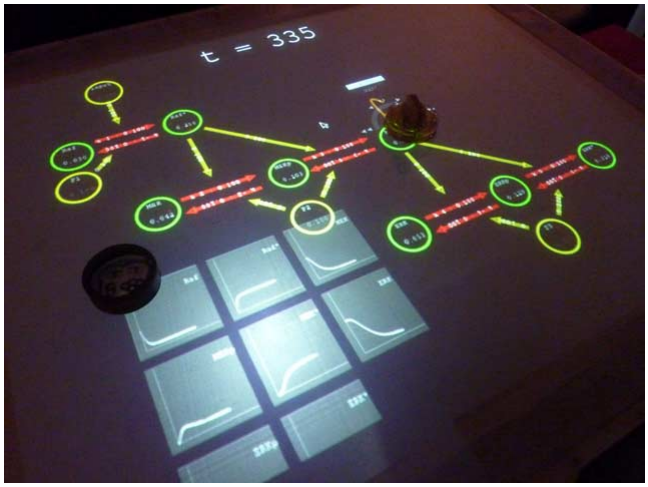

Figure 3. Pathways runs on an interactive tabletop display.

Pathways is a tangible visualization for systems biology. It works with an interactive tabletop display (see Figure 1). The interface shows a dynamic display of the pathway as the simulation is running. Manipulating objects on the table changes the concentration and kinetic order values in the simulation. Additionally, the interface provides a display of the graphs associated with the pathway, and also the

possibility of creating the pathway and its associated equations by just drawing pictures on the table, or using objects to create nodes and arrows. The interface allows even a novice researcher to create a network and simulation from scratch, and see both the global changes (entire network) and the local changes (graphs) when she manipulates parameter values kinesthetically. In the following sections, we outline some related work, and describe the current prototype, then present the feedback from users, and how we are planning to revise the prototype based on this feedback.

RELATED WORK

Interactive Tabletop displays provide a large screen area and allow multiple people to work collaboratively. They are more friendly interfaces than traditional Graphical User Interfaces (GUIs), since people are familiar with working at tables. Also, people working at a table have virtually the same access to any spot on the table. In our research lab, we have built several interactive tabletop displays [11] capable of tracking multiple finger touches and tagged objects, as well as recognizing several gestures. The display surfaces of these tabletops measure nearly 55 inches diagonally. The tables use reacTIVision [7] as their core engine to read the inputs, and Multitouch for Java (MT4J) [12] as the application level Application Programming Interface (API) to integrate with other programs. Additionally, these interactive tabletops have Bluetooth and WiFi capabilities. It is also feasible to integrate smartphones with the tables. The display of the smartphone can be a separate window in which to show dynamic information.

On the visualization side, the Sensetable [15] is a sensing technology that tracks the positions and orientations of multiple objects on a tabletop display surface electromagnetically. It introduced several tangible visualization techniques. For example, using a "scoring process" to determine the importance of the content, Using a puck as an interactive medium, indicating the center of attention by darkening the less important information, using fisheye effects to provide more detailed information and semantic zooming. This interface provides more direct interaction between the human hand and the actual digital media than does a mouse manipulating GUI widgets. Sensetable also affords better collaboration between users than GUI interfaces. The capability of manipulating multiple objects allows different people to change different interface elements concurrently.

In recent years, many groups have developed tangible tabletop interfaces for a range of applications. An application close to our project is the use of tabletop visualization to control network models. IP Network Bench [10] is based on Sensetable and provides an interface for users to determine the balance between cost and performance in IP network design. Users can have real-time results by controlling the network model collaboratively with multiple controllers. Pico [14] is a Tangible User Interface (TUI) based tabletop surface that can track and move small objects on top of it. The position of these physical props represents and controls application variables for optimizing the configuration of cellular telephone network radio towers. The computer optimizes the network, while the user moves the physical props under the constraints of other physical objects.

LifeLines [17] visualizes the personal histories of various types of biographical data, such as medical and court records. These data are collected and converted into appropriate visualization to explore the personal histories. The linear time scale helps to provide an intuitive approach to visualize histories in different time scales, in months, weeks, days and even minutes. It always begins with a screen showing the overview of the record, and the user can use options, recalling tools or filters, to analyze complex records. Visualizing with different colors and thickness of lines, rescaling, and filtering allow users to focus on parts of the record and find out more details.

ThemeRiver [5] is another visualization representing thematic variation over time within a large collection of documents. The river flows from left to right within the context of a timeline, and the variation of data is depicted with different colors and width. This visual metaphor facilitates discovery by presenting data in an intuitive, easily ingested format and builds a relationship with user's perceptual and cognitive abilities. This tool is useful when the user needs to figure out trends, relationships, unexpected data, and structure in the data.

There are also other systems designed for visualizing biological pathways. VisANT [6] is a web-based software framework for visualizing the network models of biological interactions and relations. Users can import data from either their own data source, or from standard exchange format and they can be represented with millions of edges and nodes. The system not only provides analytical tools for extracting topological properties, but also supports customizing and modifying the network with user-defined sets. Another visualizing and modeling tool, iFBA (Integrated Flux-Balance Analysis) [3] has the advantage of flux balance models over traditional sets of ODEs that allows for analysis of the entire metabolic and regulatory networks. It also has the advantage of the ODE models to capture intracellular concentrations and short time-scale dynamics. iFBA approach has the potential to incorporate the advantages of both perspectives.

Cytoscape [18] is a software tool for integrating bio-molecular components and their interactions with expression profiles, phenotypes and other molecular states. It supports various automated network layout algorithms, and visualizes various sizes and colors of nodes and edges. The system is extensible by adding new plug-ins, which allows rapid development of additional visualization and computational analyses.

Karp [8] describes Pathways Tools, a software environment for creating a model-organism database that can integrate the evolving understanding of the genes, proteins, metabolic network and regulatory network of an organism. This paper supports our point that most visualization tools just re-represent data, they don't seek to support the process of fitting data. More broadly, our objective is to help people fit data, and estimate parameters. Visualizing the network is a step in that direction.

DESIGN

We interviewed our potential users to understand how they sketched a graphical pathway, and to see what other pathways look like. It was found that there is no standard method for drawing a pathway. One common characteristic of most pathways is the use of directed graphs. The nodes of these types of graphs are usually enzymes or molecules, and the edges are the reactions.

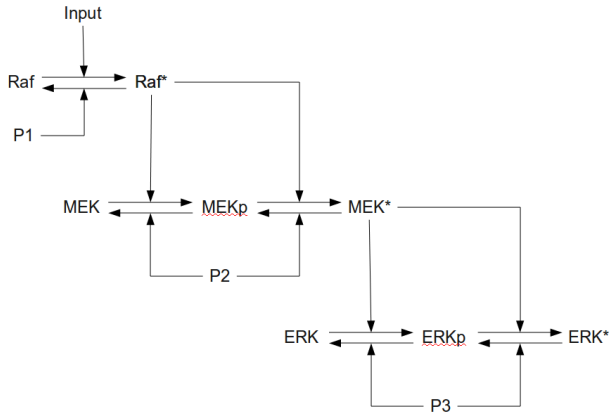

Figure 4. A draft of the MAP Kinase cascade. *Raf*, *Raf, *MEK*, *MEKp*, *MEK**, *ERK*, *ERKp*, and *ERK** are molecules. *Input*, *P1*, *P2*, and *P3* are enzymes.**

The default data running on our current implementation of Pathways is the MAP Kinase cascade in the MAPK/ERK pathway. There are 8 molecules and 4 enzymes in this pathway. A draft of the pathway from one systems biologist is shown in Figure 4.

Simulation

The underlying math of a pathway is a series of Ordinary Differential Equations (ODEs) that represent the producers and consumers in the chain reactions. Currently, systems biologists have to write programs to solve these ODEs. We expect them to interact with the graphical representation of the data once Pathways is completed. There are several mathematical libraries available for solving the ODE set. We tried two different approaches to create a sample pathway.

Graphical Pathways with Supporting Math

The idea is to allow the systems biologist to create a pathway on an interactive tabletop display. This can either be done using a graph description language or using the tools provided by our tabletop application. DOT [19], a language designed to describe graphs was used in the first prototype. DOT is good for describing complex graphs. To describe a graph, one has to use this syntax:

```
A -> B
```

This statement describes the start-node A, a directional edge from A to B and the end-node B. DOT is written in plain text and can be easily parsed by computer programs. However, one cannot describe a graph using this DOT without actually having a draft beforehand.

In the current system, a user drafts molecules by using the visual tools provided by the tabletop. After that, she creates reactions between molecules by dragging her finger from one molecule to another. The ODEs are generated when she creates the pathways. This approach is intuitive, but the user needs to have a rough pathway image in her mind in advance to create the graph. Another drawback of this approach is the freedom of creation can create unreasonable chemical chain reactions and unsolvable pathways. This problem is partially offset by the knowledge of our users, who are very familiar with their domain.

Math Equations with Graphics

Another approach is to allow the user to enter the ODEs. After that, the system generates the pathways based on the equations. Eventually, they have to approach the tabletop to interact with the visualizations on the tabletop surface. In this approach the user can concentrate on the numerical data and mathematical equations rather than on the visual representation of data. However, few people can think with ODEs only, especially when the pathway grows larger. Even if some claim they do, they most likely have graphical representations of data in their minds.

Visualization

There are benefits and drawbacks of using different types of visualization techniques. One design decision we needed to make was using abstract versus more organic looking pathways. An abstract visualization [8, 9] is very similar to a subway map. It gives users a clear overview of the pathways. However, we wanted to make the pathways look and feel more organic. For example, molecules should move and morph slowly like amoebas. A very structured abstract pathway similar to a subway map looks more robotic than organic.

Reactions

To show the reaction speed, we had several options. Changing the stroke weight of a reaction arrow based on its flux, namely the reaction speed, is a simple solution. Another similar visualization is to change the arrow color based on its data. However, these two methods do not always generate steady and clear visualizations. When the change is subtle, users can hardly tell the difference.

Another technique that is more perceivable to users is to use animations (see Figure 5). The animation moves small arrows one by one from one node to another. This would give users clearer feedback about the process. Other visualization options include using numerical presentations of flux, using pie charts to show normalized speeds, or using a progress bar to show the values.

However, except for the "numerical presentation", all alternatives have the problem of reflecting the actual data. In other words, users can hardly distinguish the subtle changes of values. Consequently, our final decision was to use animations to show the speed, since it provides the most obvious feedback.

Figure 5. An image of three molecules "breathing". From their shapes, we can tell *Raf* was exhaling while *Raf and *P1* were most likely inhaling when they were captured.**

Molecules

We changed the size, the filled color and the stroke color of the molecules based on the concentration value in the beginning, but encountered similar problems – users can hardly tell the difference between different molecules. Therefore, we used similar animation techniques as above to visualize the molecules. The molecules move slowly and randomly in a confined area. They also "breathe" slowly by changing the width of the shape (see Figure 5). The stroke color of a molecule shows its type. Green is for molecules and yellow is for enzymes, which are also a different kind of molecule. The higher the concentration value, the deeper the "breath" a molecule makes. A molecule also has its name and concentration value on it. We added some red dots in one prototype to evaluate the visualization. The dots are generated randomly inside a molecule. The number of the dots reflects the concentration of a molecule.

Charts

According to Lab G researchers, the chart is the most important channel for them to understand a series of reactions under certain initial conditions. One of our goals was to allow them to work more effectively without frequently switching between charts, ODEs and the sketched pathways. Therefore, a chart is a supplementary tool in our application. Ideally, the researchers would read all information from the graphical representation of Pathways. Currently, a tangible object is used to control the charts' position.

The charts in Pathways show the concentration changes across time. In each single chart, the corresponding molecule's name and its concentration value at that time are shown. A vertical baseline moves when time passes. A user can compare the concentration values of different molecules at the same time (see Figure 6).

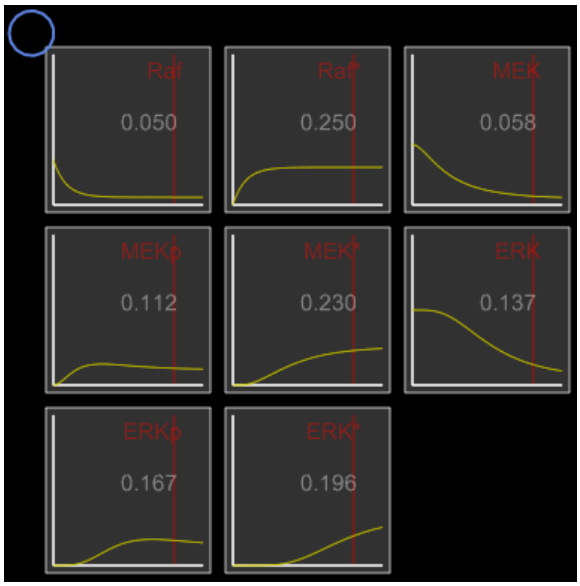

Figure 6. The latest output charts of the MAP Kinase cascade pathway. The red vertical line is moving with time to show the concentration at a certain time. Also, the curves are plotted in advance.

Tangible Input

It is hard to make a decision of when to use tangibles, and when to use more direct manipulation with finger touches. As the interaction surface supports both the tracking of tangible objects and finger touches, we frequently encountered decisions regarding when to prefer one method of interaction to the other. With tangibles, only the person holding the tangible has the control of the tangible's corresponding digital component, whereas finger touches allow all users an equal chance to manipulate the media on the tabletop. Tangibles thereby lead to scenarios with more centralized control of the simulation, and (perhaps) fewer accidental modifications.

A limitation to the current set of tangibles is that they occasionally block the view of the digital content on the tabletop. Since the orientation of a tangible can be determined, the orientation of digital content can be adjusted to face the user. The current build of the application uses tangibles for all major interactions, which include: modifying molecule concentrations, adding molecules to the network, positioning the graphs of the ODE, and starting and stopping the reaction simulation.

One tangible we use is modeled after the dials on old VCR remote controllers (see Figure 7). When a user rotates the VCR dial clockwise, video playback speeds up. The more she rotates, the greater the speed. When the user releases the dial, it returns to the initial state and plays the video in a normal speed. Likewise, a counter-clockwise rotation rewinds the video. The dial in Pathways operates similarly. Only instead of manipulating time, it adjusts the initial concentration of a molecule (see Figure 8). When a user releases the dial in Pathways, the dial returns to its balance state and the corresponding number stops changing.

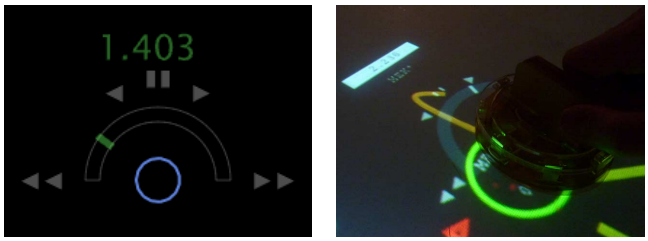

Figure 7. The left image is the digital content when a physical dial is pointing at about 10 o'clock. Turning in this direction decreases the number. In the right photo, a user increases the number by spinning the dial toward "►", which is the symbol of "*forward*".

User Scenario

When a user comes to Pathways, she can either load a pre-defined metabolic network into the simulation environment, or start from a blank slate. The user creates the elements of the metabolic pathway using both physical and graphical widgets. By stamping the molecule widget onto the tabletop, she creates a molecule. Following the stamping, a virtual keyboard appears and prompts the user for the name and initial concentration of the molecule. After creating several molecules, she uses either her finger or a stylus to draw lines between molecules, thus creating a reaction network. Now that the network is constructed, she can run the simulation. To start the simulation, she simply places the simulation widget onto the surface.

FEEDBACK

As this is a system designed to support scientific discovery in a complex area, the available user base is very low (around 10 researchers) and their needs are diverse, so it is difficult to do formal evaluations to find out the efficiency of different visualizations. We have conducted a series of user feedback sessions with individual researchers, and we are using this feedback to develop a new prototype. We expect many cycles of this iteration process before a final prototype is reached. Here we present the initial feedback on the current prototype.

A PhD student from Lab G (G10) attended the early demonstration of the project and gave us feedback based on the first prototype. One of the most important points he raised was the need for graphs. Since the researchers' current practice largely depends on the output charts, they cannot understand the reaction without the graph. Even though one of our goals was to replace the equations and output charts with visualization of the reactions and molecules, at this point, the researchers' thinking relies heavily on the output charts. This user really liked the animation of iFBA mentioned earlier and suggested using the same visualization in Pathways. This involves using the reaction stroke width to indicate the reaction speed, and showing the simulation in different playback speeds. Also, system biologists work with output charts with different initial condition sets. Comparison of charts to show different setups and outputs is necessary for them to work effectively. He also mentioned that there are several types of modeling methods, each generating different ODEs. Thus, it will be helpful to show the ODEs.

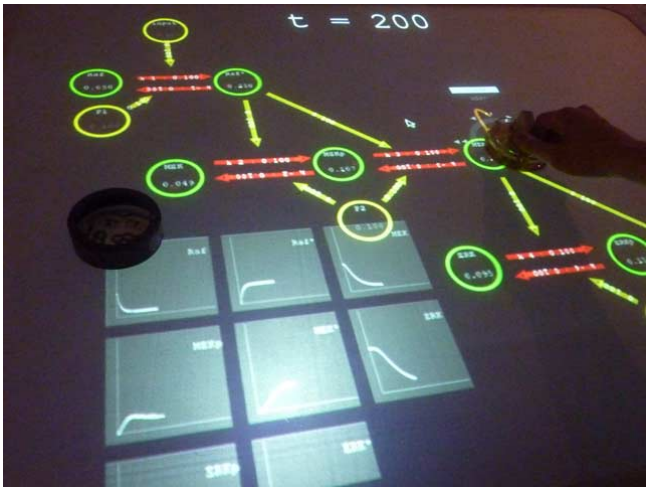

Figure 8. An overview of the running Pathways. A counter is shown above the pathway. A user is manipulating the concentration of a molecule.

G4, a system biologist whose concentration has been in modeling biological systems for many years felt that the glimmering dots inside the molecules distracted him when he tried to read the numbers. He recognized the biochemical reaction immediately and suggested showing more complicated structure. Since the current one is too simple, he had no interest in discovering anything from it. He mentioned that the lack of equations or chemical formulas in the visualization made him feel insecure. He'd like to enter the equations by himself and let the computer generate the graph. G4 mentioned that he has other visualization tools available in his lab, but most researchers do not use them. He doubted whether they would use this novel system. Even though he did not think Pathways would be good for exploration, he thought it would be an ideal platform for demonstration and educational purposes.

C7, a member from another Lab G, attended the latest demonstration of Pathways. He was especially interested in connecting MAP Kinase cascade, a subset to MAPK/ERK, to the pathway. He could read the information from the

charts, but could not relate this information to the entire pathway. He wanted to understand how changing the concentration of one molecule affected the entire system. On the other hand, he also wanted to see how changing a larger pathway impacted a smaller region of it. He said a lot of the time his work is to find out the model based on the experiment results, namely the output charts. However, instead of continuous curves, the charts are often composed of discrete points, which make modeling difficult. He particularly hopes to see a tool that helps him simplify the task. C7 also mentioned the importance of being able to see and compare different points of time in a simulation. Graphs help with this, however we believe this information could also be incorporated into the representation of the pathway itself. G6, another member from Lab G, wanted to be able to change the topology of the pathways. Moving the nodes and edges of a graph around was important for her to connect the model in her mind.

CHALLENGES

Within our lab, the system has been iterated several times to find out the most appropriate representation of the simulation and the most straightforward interaction techniques. The feedback from the respondents indicated several challenges in developing such a visualization.

A Comprehensible Representation

Since there is no standard for drawing a pathway, every researcher has a different method. Sometimes a researcher chooses the representation based on the modeling method. For example, sometimes intermediates are critical under one modeling method, but are ignored in another modeling method. Also, researchers like to work with a graph that matches their mental model. Two researchers may use the same modeling method and visual representation, but the way they organize the pathway can be very different.

The Appropriate Visualization

To make the molecule look organic and life-like, we embedded animations into the simulation. The molecules move randomly in a small confined area. Also, the shape of the molecule morphs with time. The speed of the movement and magnitude of the morph are affected by the concentration of the molecule. We particularly exaggerate this effect for users to see the maximum and minimum values of the visualization. We discussed alternative methods to visualize concentration changes, such as varying the color of the molecule or the thickness of the molecule border. But these techniques make subtle changes hard to notice.

One important topic in information visualization is to maintain graphical integrity.[20] For example, in a bar chart, if the height ratio of three bars are 1:2:3, the ratio of the underlying data values should be 1:2:3. Using animations in visualization can help users understand the overall distribution of the data, but it does not show the details, unless additional features are added. The animations may need to be exaggerated to emphasize the effect; this can cause the user to misinterpret the data. Several of the respondents suggested changing the circle sizes of molecules to signify the concentration changes. However, this changing of a two-dimensional area is not a good mapping for the one-dimensional concentration data.

Tabletop Interaction

It is difficult to design tangible interactions controlling physical objects that map to digital contents. Systems biologists largely rely on pen and paper. The goal is to create an interface that allows them to continue their regular productive tasks more effectively. Even though there is research showing that tabletop interfaces are effective collaborative tools, some users feel awkward when they first approach the table. Moving an object on a digital surface to reveal further information is uncommon in real life. So is moving an object with one finger. We expect the popularity of multi-touch mobile devices will cause users to grow accustomed to this type of interface in the coming years.

ONGOING CHANGES

The objective of Pathways is to help systems biology researchers find the best parameter set for each molecule, so that the data for the whole pathway fits the experimental data. This optimization is currently supported somewhat, as the user can explore the parameter space by manipulating multiple values simultaneously using the tangible user interface. A separate visualization needs to be added, to indicate when an optimized network model is found, and this will help to improve the accuracy, reliability and quality of the model.

Our original plan was to generate a dynamic graph that followed a set of physical rules. In this graph, when a node is pulled away, the elastic force will bring it back. A node is connected to another by an invisible spring. However, one user wants the functionality to change the topology of pathways. The ongoing improvement is to build a graph that adheres to specific physical laws, while allowing topology to be adjustable by users.

Charts are very important to systems biologists. The charts have evolved from a single chart with multiple curves to multiple charts and then to multiple charts with baselines that highlight the value. A newer version of charts is now being implemented for each molecule.

Since every biologist has a different way of modeling, the new Pathways interface should be customizable. The system will offer an interface for users to customize the network model by providing a variety of options, such as changing colors of nodes, grouping certain molecules and visualizing with different layouts. Also, users' personal settings and modeling data should be stored separately and be accessible from multiple locations. Controlling

parameters will reflect in the simulation in real-time, and the updated results will be visualized with multiple graphs showing the history of parameter changes. Users will be able to compare and analyze based on time differences, which will also be projected as an animation within the representation of the pathway itself.

For the purpose of discovery, Pathways needs to provide more statistical and computational tools for researchers to analyze the network model. On the other hand, for educational purposes, the system needs to focus on providing a simple intuitive interface and graphical expressions. This will help students get interested in the system, and provide a better understanding of network models.

Other than using tangible objects, finger gestures are another option for controlling the network model, and manipulating it in an intuitive and efficient way. The system currently provides different interfaces and options when there is more than one person using Pathways. For educational purposes, the system can track which person controlled a parameter, and its log files provide a history of data manipulation.

DISCUSSION

Two of the design challenges we face are: what are the appropriate visualization and kinesthetic manipulations of the pathway for a researcher trying to fit a model? And what is the right data, and form of data, to present so that a fit can be achieved using the visualization? In our current prototype, based on the users' preference, we chose to show the concentration values in the visualization, using different representations. However, it is possible that the change of reaction speeds might be a more effective means for them to comprehend the data, such that a fit can be achieved.

More generally, what is the relationship between the fitting process and the representations, and also the visual and kinesthetic manipulations? There is very little known about the cognitive principles and processes underlying modeling, and even less on how visual and motor manipulations contribute to discovery using simulations [1, 13]. We hope that the development and use of the Pathways interface will help shed light on these cognitive principles and processes, and thus pave the way for more sophisticated kinesthetic interfaces that support discovery.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of National Science Foundation REESE grant DRL097394084 and a research collaboration grant from Steelcase Inc. in conducting this research.

REFERENCES

1. Chandrasekharan, S. 2009. Building to discover: a common coding model. *Cognitive Science*. 33, 6 (2009), 1059-1086.
2. Cooper, S. et al. 2010. Predicting protein structures with a multiplayer online game. *Nature*. 466, 7307 (2010), 756-760.
3. Covert, M.W. et al. 2008. Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. *Bioinformatics*. 24, 18 (2008), 2044.
4. Davis, R. 2002. Sketch understanding in design: Overview of work at the MIT AI lab. *Sketch Understanding, Papers from the 2002 AAAI Spring Symposium* (2002), 24–31.
5. Havre, S. et al. 2002. Themeriver: Visualizing thematic changes in large document collections. *IEEE transactions on visualization and computer graphics*. (2002), 9-20.
6. Hu, Z. et al. 2005. VisANT: data-integrating visual framework for biological networks and modules. *Nucleic acids research*. 33, suppl 2 (2005), W352.
7. Kaltenbrunner, M. and Bencina, R. 2007. reacTIVision: a computer-vision framework for table-based tangible interaction. *Proc. TEI '07, (*2007), 69-74.
8. Karp, P.D. et al. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in bioinformatics*. 11, 1 (2010), 40.
9. KEGG PATHWAY: Metabolic pathways - Reference pathway: *http://www.genome.jp/kegg-bin/show_pathway?query=&map=map01100&scale=0.71&show_description=show&multi_query=*. Accessed: 2011-04-29.
10. Kobayashi, K. et al. 2003. A tangible interface for IP network simulation. *CHI '03 extended abstracts on Human factors in computing systems* (2003), 800-801.
11. Mazalek, A. et al. 2009. Architales: physical/digital co-design of an interactive story table. *Proc. of TEI '09*, (2009), 241-248.
12. MT4j - Multitouch For Java: *http://www.mt4j.org/mediawiki/index.php/Main_Page*. Accessed: 2011-04-29.
13. Nersessian, N.J. 2008. *Creating scientific concepts*. The MIT Press.
14. Patten, J. and Ishii, H. 2007. Mechanical constraints as computational constraints in tabletop tangible interfaces. *Proc. of CHI (*2007), 809-818.
15. Patten, J. et al. 2001. Sensetable: a wireless object tracking platform for tangible user interfaces. *Proc. of CHI* '01, (2001), 253-260.
16. PhET: Free online physics, chemistry, biology, earth science and math simulations: *http://phet.colorado.edu/*. Accessed: 2011-04-29.
17. Plaisant, C. et al. 1996. LifeLines: visualizing personal histories. *Proc. of CHI '96* ,(1996), 221-227.
18. Shannon, P. et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 13, 11 (2003), 2498.
19. The DOT Language: *http://www.graphviz.org/doc/info/lang.html*. Accessed: 2011-04-30.
20. Tufte, E.R. and Howard, G. 1983. *The visual display of quantitative information*. Graphics press Cheshire, CT.