An Interactive Environment for Visualizing, Interpreting, and Revising Biological Process Models

Durga Bidaye[1], Juraj Dzifcak[1], David Stracuzzi[1], Richard Chimera[1], Michael Verdicchio[1], John Furber[3], Seungchan Kim[1,2], Pat Langley[1]

1 Background and Motivation

Systems biology aims to understand living organisms in terms of interacting processes and their effects. Informatics researchers have developed a variety of methods that enable biologists to study these processes more effectively, but existing software suffers from three limitations: knowledge about the biological system is distributed among many sources, users cannot compare and evaluate competing hypotheses efficiently, and there is little support for automated reasoning over knowledge and hypotheses. Taken together, these pose significant hurdles to progress in developing integrated accounts of complex biological systems.

Current software environments provide support for some of these activities, but not all of them. For instance, EcoCyc [5] lets users work with their own knowledge base, but they cannot modify the core curated knowledge. A community-based knowledge repository would provide a wider scope for scientists to explore new ideas and to collaborate. KEGG [4] lets users explore relationships among biological entities, but it does not reason over its knowledge and cannot determine the effects of adding new hypotheses or modifying existing ones. Ingenuity [3] lets users visually explore biological knowledge that curators have extracted from the literature, but it does not let them modify the core knowledge or provide the ability to reason over this content. Systems biology would benefit from software that supports all of these abilities.

2 Technical Approach

Our research addresses the above concerns by developing an interactive software environment that lets scientists visualize, interpret, and revise explanatory models. The framework encodes assumptions and hypotheses about biological processes in relational logic, which lets it provide users with three important capabilities. First, an interface graphically displays the entities, events, and processes in a model, which helps the user visualize complex relationships and how they interact. Second, the environment supports reasoning over a model, which helps researchers understand how its assumptions and hypotheses lead to various predictions. Finally, the environment lets users revise elements of a model and examine the implications on predictions, which in turn helps them converge on an account that is consistent with results from the literature.

Our aim is not to supplant systems biologists but rather to support them in their work. The first step in applying our environment to a particular area is to create and store a model that incorporates processes borrowed from the relevant literature. Users enter this content in constrained English syntax, which the environment then translates into logical structures that encode processes, events, and entities. Each process is stated as a relational rule that involves one or more entities and events, and that specifies its effects on other entities and events. For example, an aging model may contain the rule "disassembling of junk exposes free iron in the lysosome", which states how the junk disassembly process affects lysosome entities, namely by introducing free iron entities into the system.

This formal representation directly supports the environment's three main abilities. The module for model visualization displays entities and events that are connected by processes in a bipartite graph. The user can ask specific questions about the model, such as how it explains a given result from the literature

[1] School of Computing and Informatics, Arizona State University, USA. E-mail: {durga.bidaye, juraj.dzifcak, david.stracuzzi, richard.chimera, michael.verdicchio, langley}@asu.edu

[2] Translational Genomics Research Institute (TGen), USA. E-mail: dolchan@asu.edu

[3] Legendary Pharmaceuticals, USA. Email: johnfurber@legendarypharma.com

(e.g., why lipofuscin increases in lysosomes over time). The environment invokes Answer Set Prolog [1] to reason over the model's rules and assumptions to generate this explanation, which the visualization module depicts by highlighting participating rules and events on the screen. The user can also ask how the model's predictions would change if one removed or added specific assumptions, say about the presence of free iron in the lysosome. In response, the reasoning engine finds one or more consistent sets of predictions, which the environment displays for the user's inspection. Finally, the representation supports the modular addition, removal, and modification of processes that comprise the model, followed additional reasoning that determines the implications of these changes and graphics that highlight the alterations. Taken together, these features support an iterative process of constructing, reasoning over, evaluating, and revising biological process models, as summarized in Figure 1.

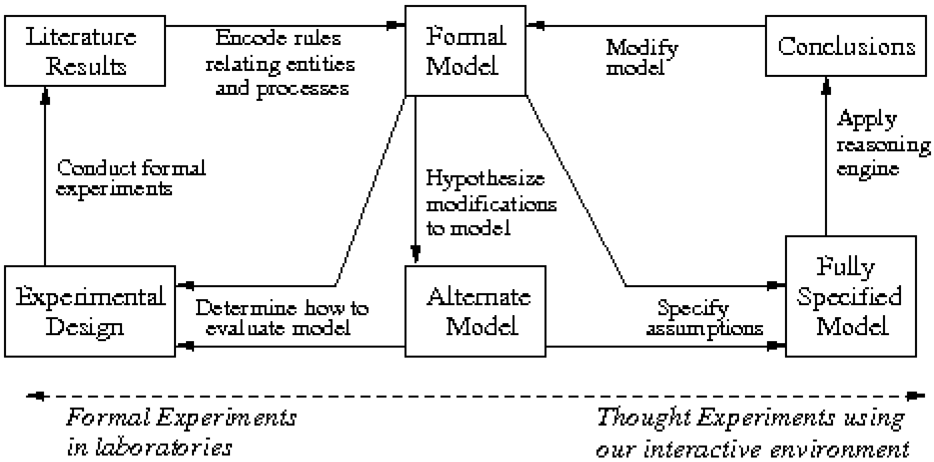

Figure 1. Using the interactive modeling environment.

3 Preliminary Results and Discussion

We have implemented this interactive software environment for systems biology and started to evaluate our claims about its efficacy. Our initial tests have revolved around a systems-level model of human aging developed by Furber [2]. We have encoded the processes, events, and entities in his model into our formalism and used the reasoning engine to detect errors in our encoding. The environment has also uncovered inconsistencies in Furber's original model, which he has since revised. These preliminary results are encouraging, but we need to evaluate the environment further by incorporating additional portions of Furber's model, obtaining feedback from other biologists on its usability, and testing it on other biological domains. Only then can we make strong claims about its usefulness to systems biology.

In summary, our research addresses the limitations of informatics tools by integrating the visualization of complex biological process models, reasoning over these models to identify explanatory chains and to determine the implications of model assumptions, and letting users modify the model's processes in an iterative effort to better account for results from the literature. We also believe that the environment, when made available on the Web, will provide a means for biologists at different sites to share hypotheses and jointly develop the complex models on which systems biology depends.

References

[1] Baral, C. 2003. *Knowledge representation, reasoning and declarative problem solving*. Cambridge, UK: Cambridge Univ. Press.

[2] Furber, J. 2009. Systems biology of human aging: Network model of biochemical and physiological interactions in human senescence. Retrieved February 16, 2009, from Legendary Pharmaceuticals: http://www.legendarypharma.com/chartbg.html.

[3] Ingenuity Systems. 1998. Products and Services. Retrieved February 16, 2009, from Ingenuity Systems: http://www.ingenuity.com/products/prod_overview.html.

[4] Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res. 36*, D480-D484.

[5] Karp, D., Riley, M., Paley, S.M., Pellegrini-Toole, A. and Krummenacker, M. 2009. EcoCyc: Encyclopedia of Escherichia coli genes and metabolism. Retrieved February 16, 2009, from EcoCyc.org: http://ecocyc.org/.