Pathway Logic
Helping Biologists Understand and Organize Pathway Information

Merrill Knapp Linda Briesemeister Steven Eker Patrick Lincoln Andy Poggio
Carolyn Talcott Keith Laderoute

SRI International, Menlo Park, CA
{firstname.lastname}@sri.com

1. Introduction

Pathway Logic [3, 4, 8] is an application of techniques from formal methods to the modeling and analysis of signal transduction networks in mammalian cells. These signaling network models are developed using Maude [1, 2], a symbolic language founded on rewriting logic [5]. Network elements (reactions) are represented as rewrite rules. Models can be queried (analyzed) using the execution, search and model-checking tools of the Maude system. Collections of rules and initial states of interest form a novel kind of database where a biologist can record results of both curation and experiments.

2. Pathway Logic
A tool for the working biologist

The requirements for a Process Diagram: (from [6], page 4 §2.2)

> A successful graphical notation system must
>
> 1. allow representation of diverse biological objects,
> 2. be semantically and visually unambiguous,
> 3. be able to incorporate notations,
> 4. allow tools to convert a graphically represented model into mathematical formulas for analysis and simulation, and
> 5. have software support to draw the diagrams.

What additional requirements are there from the perspective of working bench biologists?

Imagine a biologist in the lab studying signal transduction or to be more specific: the effect of a stimulus on a cultured cell. He takes the cell in a particular condition (state) and adds a stimulus (input) and, after a known amount of time, does a number of measurements. He then looks at the measurment results (output) and makes a hypothesis about how his system jives with the rest of the world. This assumes that he knows what the rest of the world is like. This is where a process diagram would be helpful. But he has additional requirements, as desceribed below.

(6) All the "pathways" need to be interconnected. Cells do not care how their signals are categorized. In the lab, a biologist cannot ignore the fact that the outcome of hitting an adherent quiescent cell containing EGFR, ErbB2, and ErbB3 with EGF will depend on signals from EGFR homodimers, EGFR-ErbB2 heterodimers, and EGFR-ErbB3 heterodimers, or that an adherent cell will respond differently to a particular stimulus if it is in suspension than if it is stuck to a surface.

(7) It has to be scalable and navigable. One reason that signalling is broken into "pathways" is that there is just too much information to handle at once. Only, there is so much "cross-talk" between pathways that it is not practical for the biologist working with whole cells to divide the information that way. If the experimenter can only measure the phosphorylation state of Erk, Akt, and Shc, he does not particularly care if Raf1 gets phosphorylated first on S338 and then Y341 or the other way around. Hiding details that are not relevant to the results can substantially simplify analysis. A better way to organize the data in a map is to divide it into levels of detail—just like Yahoo maps. One wants to know how to get to the city before trying to navigate the streets within the city. If the city is big enough, divide it into neighborhoods and then look at the street names.

(8) The components have to be unambiguous but also easily recognized. Biologist have been divided into their little pockets of expertise for so long that their vocabularies have almost become different species. We now have ways to standardize language and we ought to use it in a way that still allows the use of the old familiar names but translates them easily into the old familiar name in other fields.

(9) The elements of a map (reactions, modifications, translocations, etc.) need to be linked to the data from which they were derived. Most computer represented maps can and do link you to a database record for a particular component. And many maps give a set of references at the end of the map or a "guru" that one can contact to get additional information (what is one supposed to do when the guru retires?). This is sufficient for establishing credibility. However, it does not help a biologist working in the lab who needs an assay for the activation of Mek, or who wants to know which cell line was used for particular observation, or the particular antibody that was used to distinguish between Raf1 and B-Raf. It makes it hard for the user if he has to wade through all of the references dumped in one spot, especially when he discovers that the reference is a review article the refers to another paper that actually contains the data.

(10) It should be possible to do the sort of analysis that bench biologists need from a map.

- Show me a path from `<here>` to `<there>`.
- Show me all the paths from `<here>` to `<there>`.
- Show me a path from `<here>` to `<there>` that uses `<this>`.
- Show me a path from `<here>` to `<there>` that does not use `<this>`.
- If I knock out `<this>` can I get from `<here>` to `<there>`?
- Show me all the things that, if knocked out, would prevent me from getting from `<here>` to `<there>`.

Note that the latter two items involve a redrawing of the map.

Pathway Logic is designed to take all these requirements into consideration. The Pathway Logic Assistant (PLA) [7] is a tool for browsing and querying Pathway Logic models via an interactive graphical representation. Citations from which rules (map elements) have been derived can be accessed, as can information about the proteins involved. The user can zoom in on regions of interest, or zoom out to see overall network structure. A model can be queried to find relevant subnetworks and pathways leading to situations of interest such as an activated protein or an expressed gene. Situations to be avoided can be specified allowing the user to explore knockouts and alternative pathways. Query results are also represented graphically, either in isolation or in the context of the larger network.

The modeling approach and the PLA will be illustrated using a large curated model of signaling in human mammary epithelial cell. The poster will give examples of how the above requirements are met. The reader can find out more about Pathway Logic at `http://www.csl.sri.com/~clt/PLweb/`.

References

[1] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Marti-Oliet, J. Meseguer, and C. Talcott. Maude 2.0 Manual, 2003. `http://maude.cs.uiuc.edu`.

[2] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer, and C. L. Talcott. The Maude 2.0 system. In R. Nieuwenhuis, editor, *Rewriting Techniques and Applications (RTA 2003)*, volume 2706 of *Lecture Notes in Computer Science*, pages 76–87. Springer-Verlag, 2003.

[3] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and K. Sonmez. Pathway Logic: Symbolic analysis of biological signaling. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 400–412, January 2002.

[4] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, and C. Talcott. Pathway Logic: Executable models of biological networks. In *Fourth International Workshop on Rewriting Logic and Its Applications (WRLA'2002), Pisa, Italy, September 19 — 21, 2002*, volume 71 of *Electronic Notes in Theoretical Computer Science*. Elsevier, 2002. `http://www.elsevier.nl/locate/entcs/volume71.html`.

[5] J. Meseguer. Conditional Rewriting Logic as a unified model of concurrency. *Theoretical Computer Science*, 96(1):73–155, 1992.

[6] K. Oda, T. Kimura, Y. Matsuoka, A. Funahashi, M. Muramatsu, and H. Kitano. Molecular interaction map of a macrophage. *AfCS Research Reports*, 2(14 DA), Aug. 2004.

[7] C. Talcott and D. L. Dill. The pathway logic assistant. In G. Plotkin, editor, *Third International Workshop on Computational Methods in Systems Biology*, pages 228–239, 2005.

[8] C. Talcott, S. Eker, M. Knapp, P. Lincoln, and K. Laderoute. Pathway logic modeling of protein functional domains in signal transduction. In *Proceedings of the Pacific Symposium on Biocomputing*, January 2004.