

Pathway Logic: Helping Biologists to Organize and Understand Pathway Information

M. Knapp, L. Briesemeister,
S. Eker, P. Lincoln, I. Mason,
A. Poggio, C. Talcott, and
K. Laderoute

**Find out more about Pathway Logic at:
<http://www.csl.sri.com/~clt/PLweb/>**

Abstract

Pathway Logic [1-3] 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 [4,5], a symbolic language founded on rewriting logic [6]. 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 the results of both literature curation and experiments.

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 have been derived can be accessed, as can information about the molecules involved. The user can zoom in on regions of interest, or zoom out to see the overall network structure. A model can be queried to find relevant subnetworks [3,7] or pathways leading to interesting biological results such as protein activation or gene expression. Situations to be avoided can also be specified allowing the user to explore the effects of gene knockouts or alternative pathways. Query results are also represented graphically.

Here we illustrate both Pathway Logic and the PLA using a large curated model of intracellular signaling in a human mammary epithelial cell (HMEC).

Constructing a Pathway Logic Model of Biological Signaling Pathways

To construct a Pathway Logic model we use the following overall process:

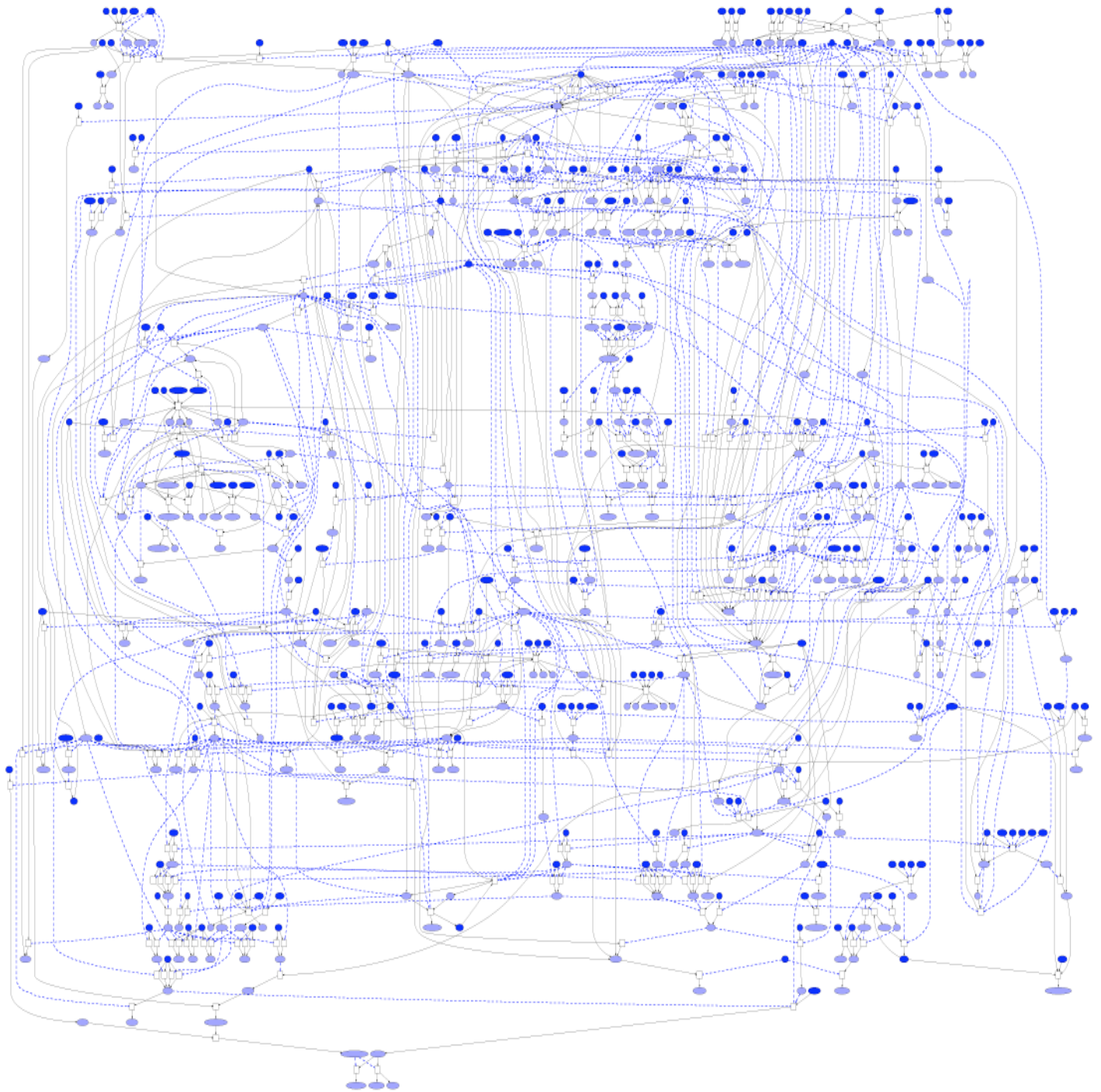
- 1) Obtain a comprehensive “parts list” of proteins and other signaling molecules present within a particular type of mammalian cell.
- 2) Include (add to the list) relevant extracellular ligands and/or other stimuli that could affect the biology of this cell type—the cell culture or experimental conditions.
- 3) Apply the Pathway Logic rules to the list of components.
- 4) Collect the results of this analysis and construct a Petri Net.
- 5) Load the Petri Net into a graphical viewer to display the signaling network (using the dot drawing tool from GraphViz).

The next panel shows an example of the outcome of this Process—a Pathway Logic model of a human mammary epithelial cell (**Panel 4**).

Notice how individual signaling pathways are embedded within a complex network, which is difficult to grasp or analyze at first glance.

However, Pathway Logic models are both scalable and navigable, so that a user can readily explore such a complex network at a more detailed level (**Panel 5**).

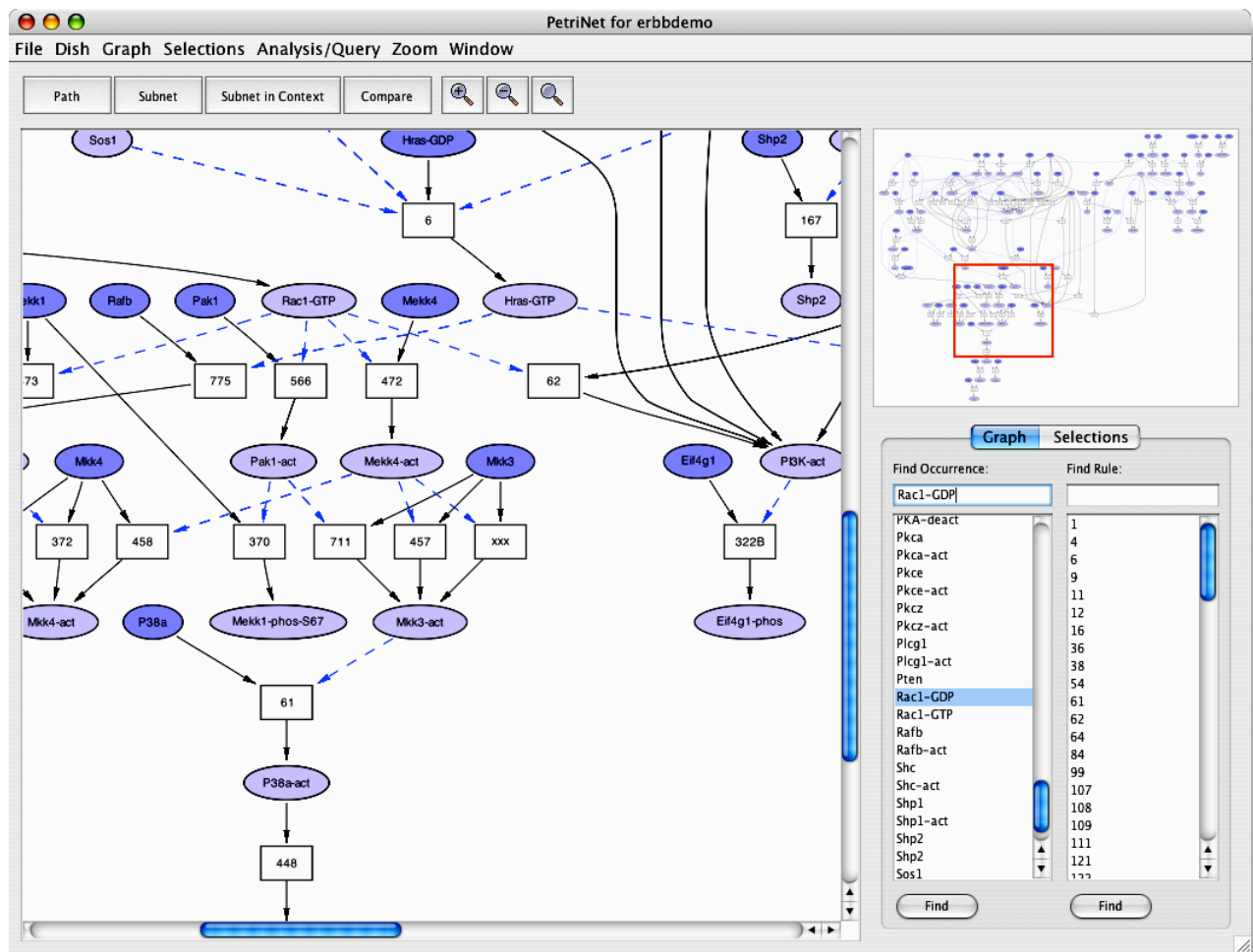
A Pathway Logic Model of a Human Mammary Epithelial Cell



The Pathway Logic Viewer: A Tool for Exploring Complex Biological Networks

Analysis buttons
send user queries to
appropriate programs

The **Navigator window** shows a
thumbnail of the whole Petri Net,
which enables a user to move
quickly to different areas of interest

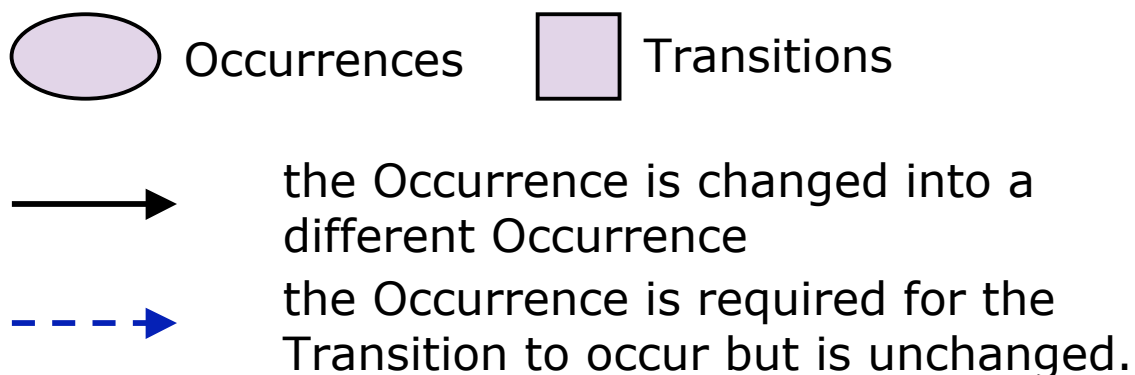


The main **Viewer window**
shows an enlargement of an
area of the whole Petri Net,
chosen by using the **Navigator**

The **Info window** contains
user query menus, shows
analysis results, and displays
information requested by a user

Graphical Notations in Pathway Logic Models

Pathway Logic Petri Nets can be very large—therefore, we use minimal notation for clarity. In fact, as illustrated below, there are only four notations (symbols) that a user needs to learn: occurrences (**Panel 9**), transitions (**Panel 11**), and two types of transitional arrows.

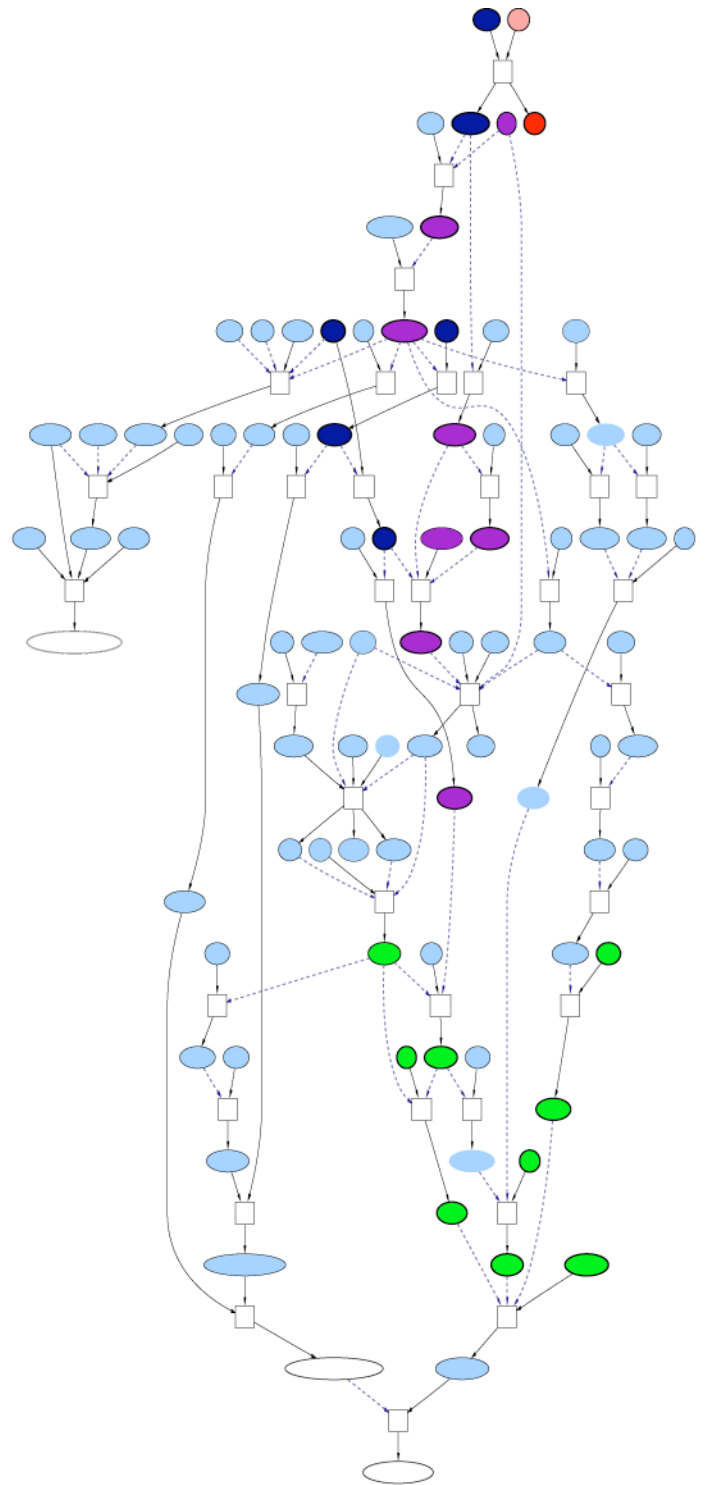
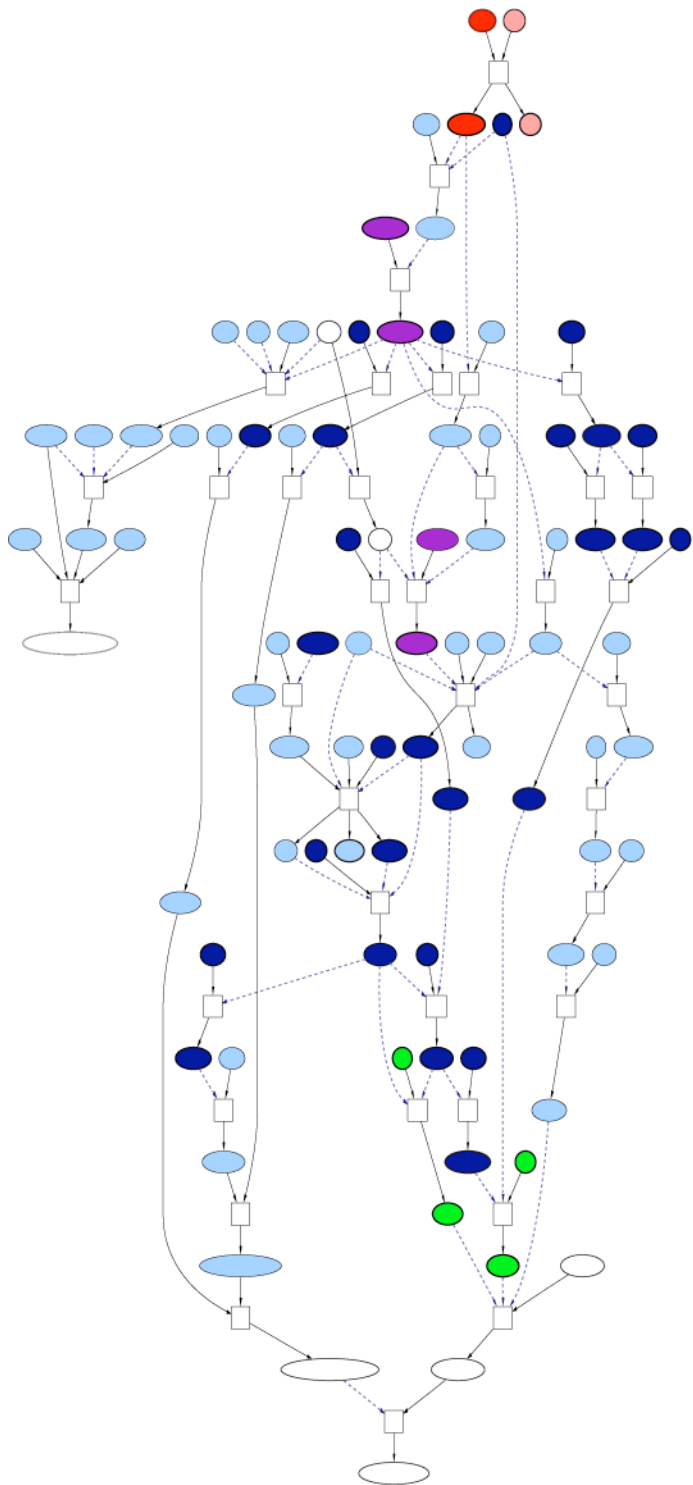


In the default format of a Petri Net (**Panel 7**), the occurrence ovals are dark blue, which represents the starting state. Light blue ovals represent new occurrences resulting from applying the starting state to the rules. Green ovals represent goals chosen by a user—selected endpoints of signaling pathways.


The colors of the occurrences can be changed from the default viewing format by using a series of toggle switches. In one type of view, component types (e.g., ligands, receptors, transcription factors) can be displayed (**Panel 8**). In another view, Component locations (e.g., cell membrane, cytoplasm, nucleus) can be highlighted (**Panel 8**). Other viewing options include indicating components by their Kinds (e.g., Protein, Chemical, DNA), and displaying a Pathway in the context of a whole Petri Net.

Color by Component

Color by Location



	Ligands		GTPases
	Receptors		Transcription Factors
	Kinases		Other Proteins

	Supernatant		Inside Cell Membrane
	Outside Cell Membrane		Nucleus
	Cell membrane		Cytoplasm

The Occurrences

An **occurrence** is defined as a component, its modifications, and its location.

Currently, a **component** can be a protein, a protein family, a protein composite (physical complex), a chemical (e.g., a small molecule, inorganic ion), a nucleic acid, or a stimulus.

Proteins: The polypeptide product of one gene. Small peptides (e.g. glutathione, a tripeptide) are also defined as proteins.

Protein names must be unambiguous in order to translate a Pathway Logic Model into other platforms such as SBML. In Pathway Logic, each protein has a Swiss Protein accession number and a HUGO symbol for its corresponding gene.

Protein Families: Groups of structurally and functionally related proteins that are treated by the source data as indistinguishable. Example: the "Erk" family includes Erk1 and Erk2, which are considered functionally equivalent.

Protein Composites: Complexes of different proteins that are treated by source data as a functional unit. Example: "Ampk" consists of a catalytic α -subunit (Prkaa1 or Prkaa2), a regulatory β -subunit (Prkab1 or Prkab2), and a regulatory γ -subunit (Prkag1, Prkag2, or Prkag3).

Modifications can be essentially any change to the structure of a component that is consistent both with the biology described by a particular model and its level of abstraction.

The Petri Nets shown in **Panels 4 and 7** use a high level of abstraction (less detail). Here, some of the modifications used are "act" for activation, "on" for induced gene expression, and "reloc" for adapter proteins that are recruited to the inner side of the cell membrane.

Locations refer to specific cellular compartments in the cell type to be modeled—components may be constitutively located at these sites or may translocate to them in response to a stimulus.

Locations are abbreviated using upper case letters to represent the compartment and the lower case letters o, m, i, and c to represent the outer surface, the membrane, the inside surface, and the interior of the compartment, respectively. Examples of cellular compartments are the following:

CL - Cell

NU - Nucleus

MO - Mitochondria Outer Compartment

MI - Mitochondria Inner Compartment

ER - Endoplasmic Reticulum

GA - Golgi Apparatus

LE - Late Endosome

EE - Early Endosome

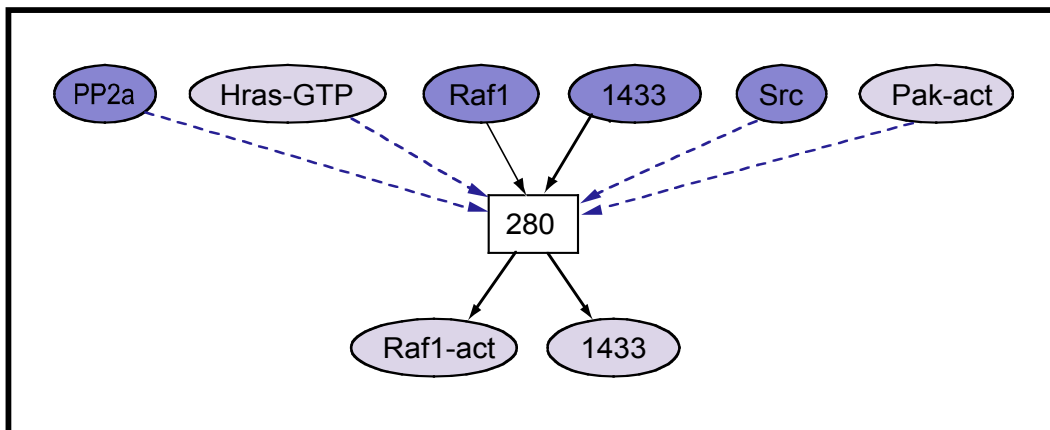
LY - Lysosome

CP - Clathrin Coated Pits

Transitions

A **transition** represents a Maude rule in Petri Net notation.

Here we show an example for "rule 280" from our Pathway Logic model of an HMEC...



Here is the same transition written as a Maude rule...

```
rl[280.Raf1.on]:
  {CLi | cli [?Ras:Ras - GTP] [Pak - act] Src }
  {CLc | clc Raf1 1433x1 PP2a }
  =>
  {CLi | cli [?Ras:Ras - GTP] [Pak - act] Src [Raf1 - act] 1433x1 }
  {CLc | clc PP2a } .
```

Here is the transition paraphrased in English...

Raf1 activation requires a GTP bound member of the Ras family, activated Pak, and Src located at the inner side of the cell membrane as well as a member of the 1433 family and the composite PP2a located in the cytoplasm.

Annotation

One of the most important and novel features of Pathway Logic is its ability to operate at different levels of abstraction. The Petri Net of the HMEC model that we have used in this demonstration is displayed with most of the information about the nodes removed. This information is not "lost"; it is carried along with the appropriate nodes. The nodes can be queried in the viewer to display the following information:

Occurrences:

The Name of the component

Including the unambiguous identifiers described in **Panel 9**, links to the identifier source, and synonyms.

The Modifications of the component

In as much detail as can be found. For example: "Raf1-act" is phosphorylated on S338, S339, S621, and Y341. It is bound to a member of the 1433 family on phosphorylated S621 and to a member of the Ras family on the Ras Binding Domain.

The Location of the component

Transitions:

The source of the information

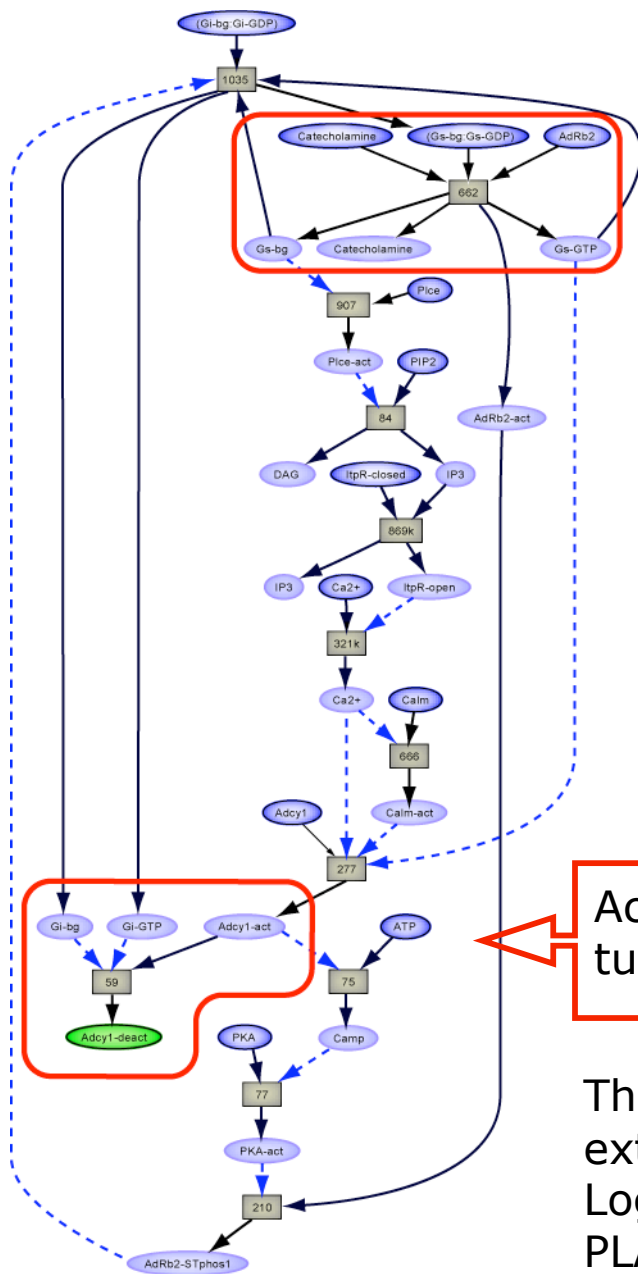
Includes a PubMed ID with a link to its abstract, a designation as to whether the source is data or a review, and a short comment describing which part of the rule was found in the reference.

Mathematical information

Includes rate constants, concentrations, probabilities, and stoichiometry.

Analysis

One of the most frequently asked questions about Pathway Logic is how it handles **negative feedback**. A classic biological example of negative feedback is the switching on and off of Adenylate Cyclase (Adcy1) by the β -2 Adrenergic Receptor (AdRb2).



When AdRb2 is bound by its catecholamine ligand it couples to and activates the heterotrimeric G-protein complex Gs (Gs-GDP:Gs-bg).

This event leads to the activation of Adcy1, which then activates PKA. PKA phosphorylates AdRb2 on S262, which causes the dissociation of Gs subunits from the receptor and association of the Gi complex (Gi-GDP:Gi-bg) instead.

Activated Gi-GTP deactivates Adcy1, turning off the signal to PKA.

This tiny switching mechanism was extracted from our big Pathway Logic model (**Panel 4**) using the PLA analysis tools.

The complete analysis will be made available on our website at <http://www.csl.sri.com/~clt/PLweb/>.

Summary

In their report to the Alliance for Cell Signaling, Oda et al. discussed some desirable requirements for a Biological Signalling Process Diagram [8]. They felt that an effective graphical notation system should have the following features:

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
5. Have software support to draw diagrams

What are important additional requirements from the perspective of working bench biologists? We suggest the following:

6. All the "pathways" should be interconnected into one network.
7. The network should be scalable and navigable.
8. The components must be unambiguous but also easily recognized by the user.
9. The elements of the network (reactions/transitions, modifications, translocations, etc.) need to be linked to the data from which they were derived.
10. The system should be capable of doing the kind of analysis of a network diagram that would be of practical value to a bench biologist.

This Poster has demonstrated that Pathway Logic is capable of meeting the requirements of both Systems Biologists, who require diagrams for analyzing existing biological information, and bench biologists, who need a place to store the new experimental information in a readily accessible form.

References

1. 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.
2. 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>.
3. CL 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.
4. M Clavel, F Durán, S Eker, P Lincoln, N Martí-Oliet, J Meseguer, and CL Talcott. Maude 2.0 Manual, 2003. <http://maude.cs.uiuc.edu>.
5. M Clavel, F Durán, S Eker, P Lincoln, N Martí-Oliet, J Meseguer, and CL Talcott. The Maude 2.0 system. In R Nieuwenhuis, editor, *Rewriting Techniques and Applications (RTA2003)*, volume 2706 of *Lecture Notes in Computer Science*, pages 76-87. Springer-Verlag, 2003.
6. J Meseguer. Conditional Rewriting Logic as a unified model of concurrency. *Theoretical Computer Science*, 96(1):73-155, 1992.
7. CL Talcott and DL Dill. The pathway logic assistant. In G. Plotkin, editor, *Third International Workshop on Computational Methods in Systems Biology*, pages 228-239, 2005.
8. K Oda, T Kimura, Y Matsuoka, A Funahashi, M Muramatsu, and H Kitano. Molecular interaction map of a macrophage. *AfCS Research Reports* [online]. 2004, Vol 2, no. 14DA [cited 25 August 2004].
<http://www.signaling-gateway.org/reports/v2/DA0014>.

This recent work was supported by grants GM068146 and CA112970-01 from the National Institutes of Health.