Pathway Logic Modeling of Protein States and Functional Domains

Merrill Knapp, Carolyn Talcott, Steven Eker,
Patrick Lincoln, Keith Laderoute
(firstname.lastname@sri.com)

Computer Science Laboratory
Center for Computational Biology
SRI International, Menlo Park, CA

About Pathway Logic

- Pathway Logic is a computational approach to modeling biochemical and biological processes using formal methods and rewriting logic
- Pathway Logic uses a computer language called Maude (http://maude.cs.uiuc.edu, http://www.csl.sri.com)
- Biological networks are automatically assembled using collections of rewrite rules representing network elements (biochemical processes)
- These formal networks can then be queried or analyzed using formal methods tools
- For example, by choosing an initial condition/state the following tasks could be performed:

- Execution—show a specific pathway through the network
- Search—show all pathways converging on a specific final condition/state
- Model checking—is there a pathway having particular properties?

Pathway Logic and Databases

- Pathway Logic can be used as a novel database technology
- Pathway Logic models make information explicit that is only implicit in traditional biological databases. For example,
  - Models incorporate knowledge about high order structures (e.g., cellular compartments, scaffolds for protein kinases)
  - New pathways or the states of components (e.g., protein phosphorylation states) can be generated on demand
  - Models can be used make inferences and predictions (e.g., the effect of knockouts on a network)
  - Biological subsystems/modules correspond to formal modules and can be combined or transformed using the same computational methods
  - Models can be analyzed for consistency or compared with other models

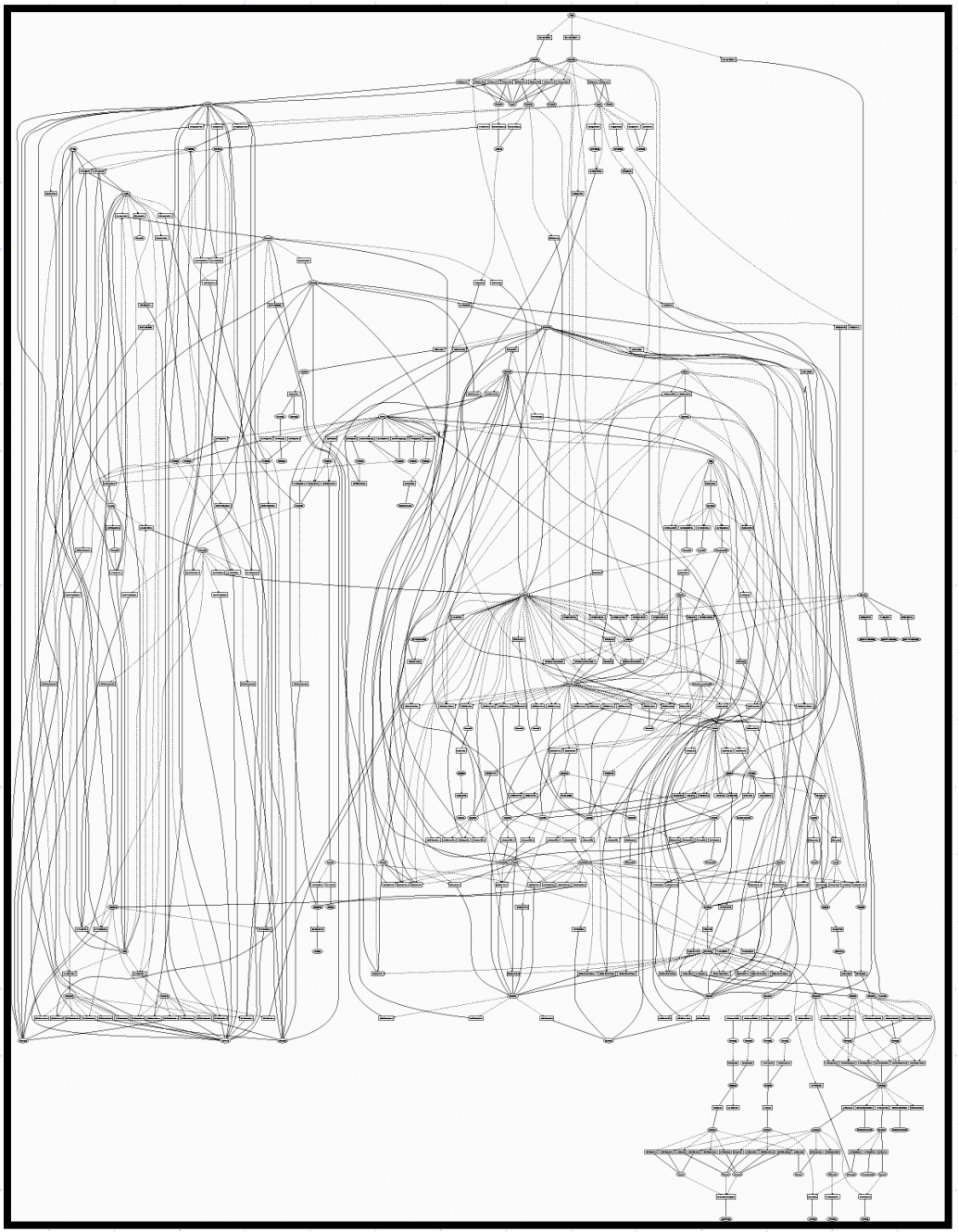
A Key Concept: Levels of Abstraction

- In Pathway Logic signaling processes or metabolic pathways are modeled at different levels of abstraction
- Level I (e.g., an EGF receptor or ErbB1 network)
  - Here proteins in the network are represented as abstract symbols that may be annotated (e.g., activated/inactivated)
  - Biochemical pathways or processes in the network (e.g., activation by phosphorylation) are represented by rules describing transitions
- Level II
  - Here proteins in the network are represented as collections of functional domains that can include biochemical modifications (e.g., phosphorylation sites)
- Pathway Logic is designed to navigate between Levels I and II, as desired

A Level I Pathway Logic Model

This graph represents an a ErbB1 signaling network generated by Pathway Logic, and depicts some potential consequences of the binding of EGF to its receptor. The network is based on rules derived entirely from the biological literature and appropriate databases.

Clearly, such a complex network requires tools to enable analysis and reasoning.



Examples of Level I Analysis Tools

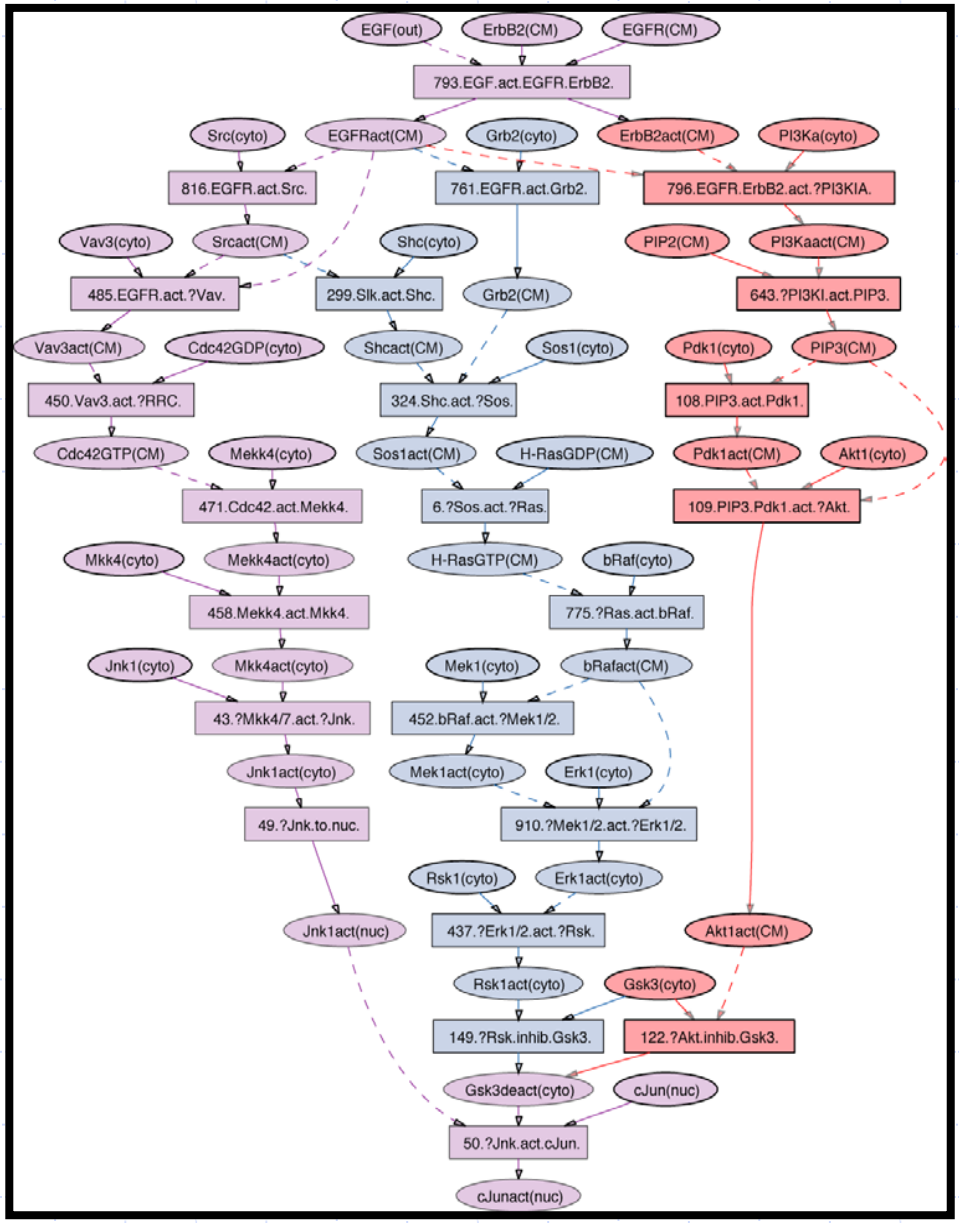
Query: Show the shortest pathway leading from EGF binding to its receptor ErbB1 and the activation of c-Jun (phosphorylation).

Result: The pathway is shown in blue and violet*.

Query: If we knock-out Grb2, can c-Jun still be activated? If so, what is the pathway?

Result: The pathway is shown in orange and violet.

*The violet pathway is common to both queries.



Why Two Levels of Abstraction?

- To generate a meaningful biological network, all the rules in the network must consist of the same data types
- In Level I the rules describe

  - components (e.g., proteins, small molecules)
  - locations (e.g., cell membrane, cytoplasm)
  - attributes (e.g., activated, GTP-bound)

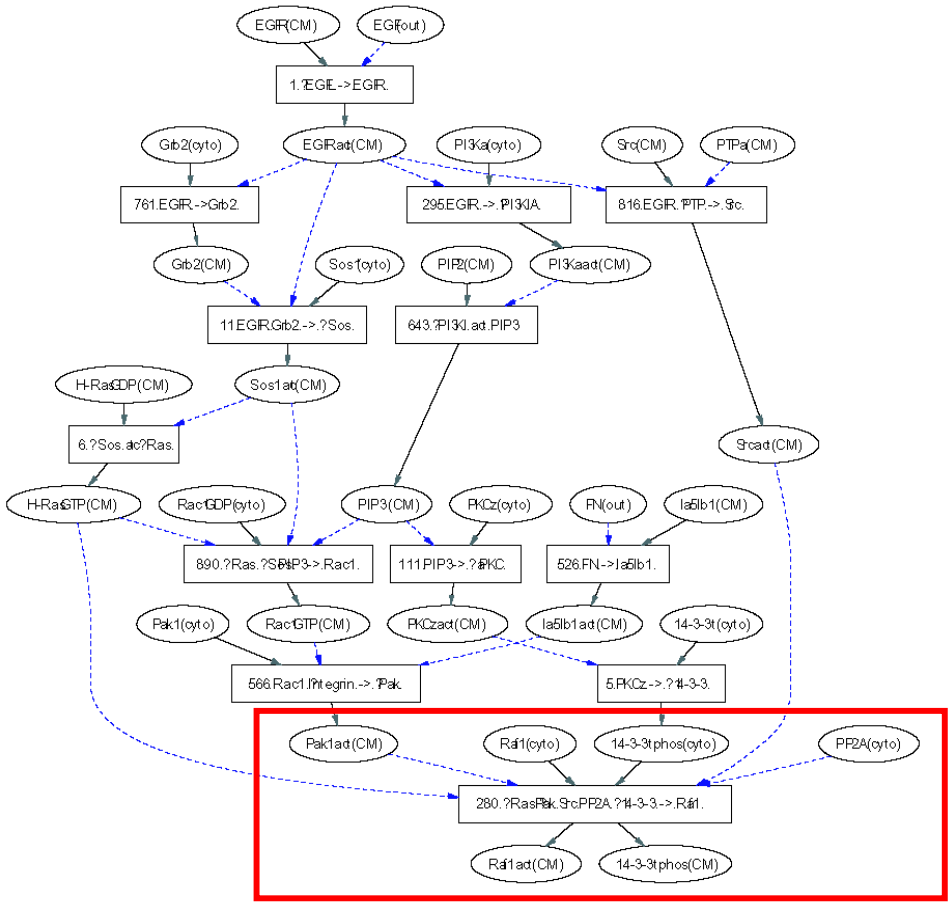
- But what about critical biochemical features such as phosphorylation sites and protein functional domains?
- Level I networks do not include such features for two reasons:

  - Data are not currently available for all the protein-protein interactions in the network.
  - The number of possible interactions is extremely large.

- We have extracted RULE 280 from the ErbB1 network to show how a Level I rule can be expanded into Level II rules …

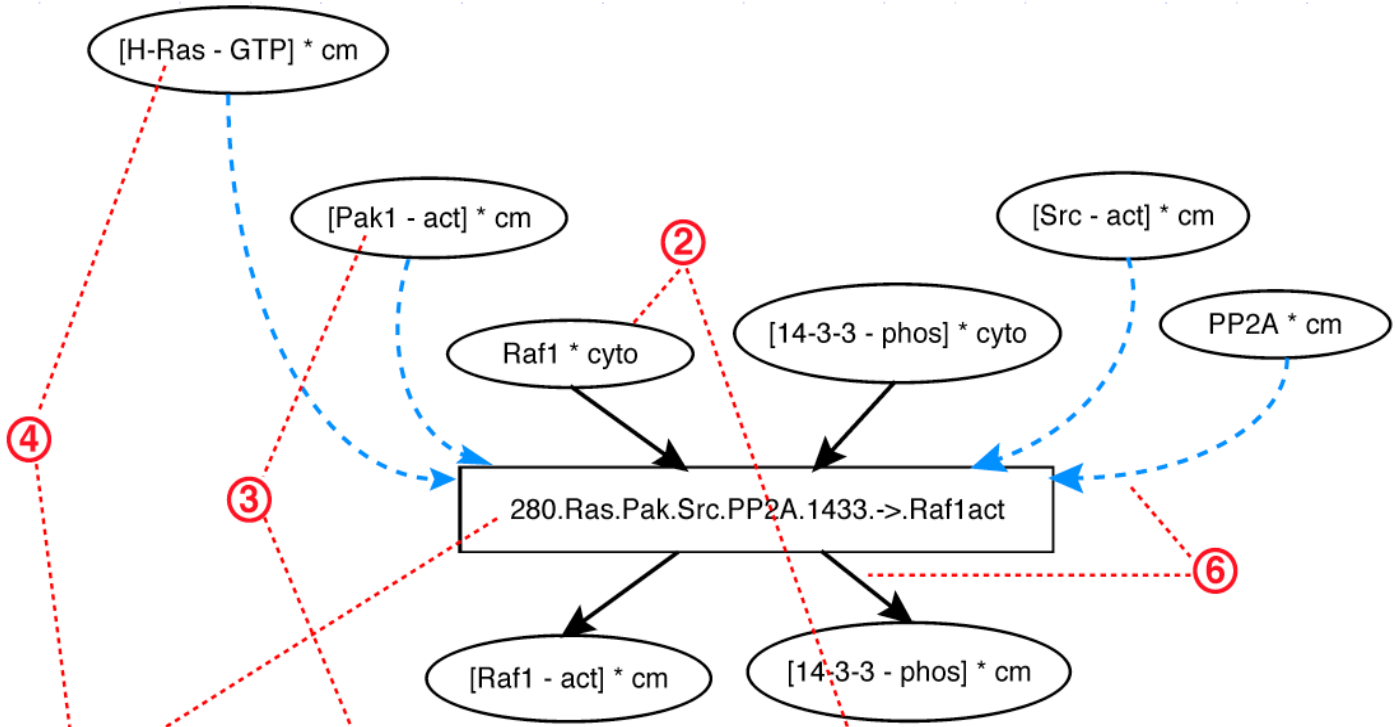
Translation of Level I Rule 280:

If the cell membrane contains an activated Ras, Pak, and Src and the cytoplasm contains Raf-1, PP2A, and a phosphorylated 14-3-3, then Raf-1 will translocate from the cytoplasm to the membrane and become activated.



Part of the EGF receptor (ErbB1) network leading to activation of the Raf-1 kinase

Level I Rule 280 Expanded (Some Hidden Maude Code)



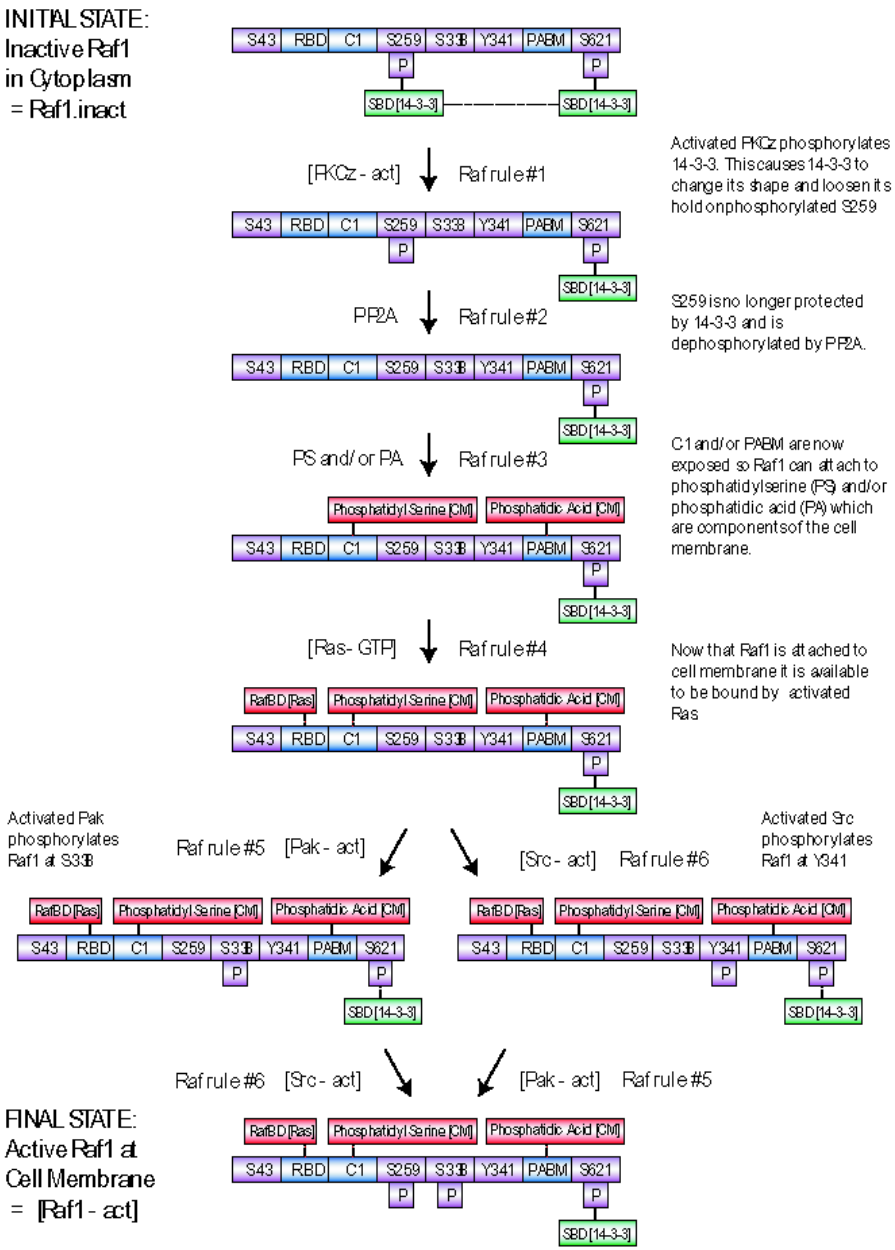

```
crl[280.Ras.Pak.Src.PP2A.14-3-3.->.Raf1act]:
{CM | cm [?Ras - GTP] [?Pak - act] [Src - act] {cyto Raf1 [?14-3-3 - act] PP2A }} =>
{CM | cm [?Ras - GTP] [?Pak - act] [Src - act] [Raf1 - act] [?14-3-3 - act] {cyto PP2A}}
 if ?Ras S:Soup := N-Ras K-Ras H-Ras [metadata "21192014, 22122191, 21864215"] .
```

Level I Rule 280 Expanded—How to Read a Petri Net Rule

- Level I rule 280 is shown as a Petri net representation:

1. The rectangle represents a biochemical transition described by the rule. The Maude text inside is a string copied from the square brackets on the first line of the rule code.
2. Ovals represent "occurrences" processed by the rule. An occurrence consists of a component (e.g., a protein), an attribute (e.g., activated, GTP bound, or phosphorylated), and a location (e.g., cell membrane or cytoplasm).
3. Question marks in Maude represent variables. The expression [?Pak - act] means that any isozyme of Pak is activated. The PetriNet viewer replaced the variable expression with the constant (Pak1) used in rule 280.
4. Variables can be conditional. Here, H-Ras, K-Ras, or N-Ras but not R-Ras or M-Ras will activate the rule.
5. Metadata provide information that is important for the rule but not used to generate the network. Here, the references used in writing rule 280 are listed by their MedLine ID's.
6. Solid black arrows show that an occurrence is changed by a rule (e.g., Raf-1 has a new attribute (act) and a new location (cm, cytoplasm). Blue dotted arrows show that an occurrence is necessary for the rule but is not changed when the rule is "fired".

A Level II Rule



This picture shows how the activation of Raf-1 (in Rule 280) will appear at Level II.

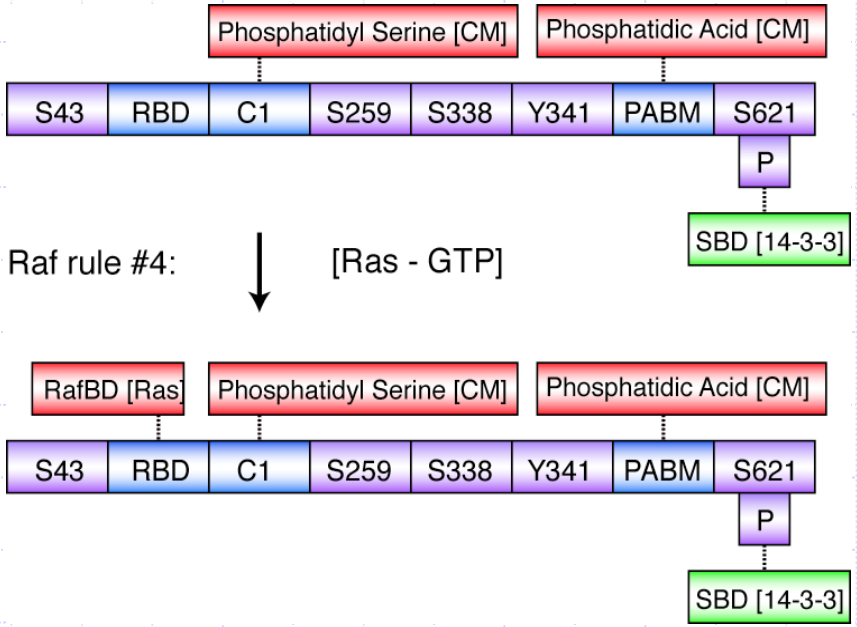
The picture is a hand drawn prototype showing how the computer will interpret level II code (work in progress).

A user will be able to access this information from the Level I Petri net. When more detail about a particular rule is desired, the user will be able to click on the rule rectangle and this type of diagram will be displayed (any suggestions?).

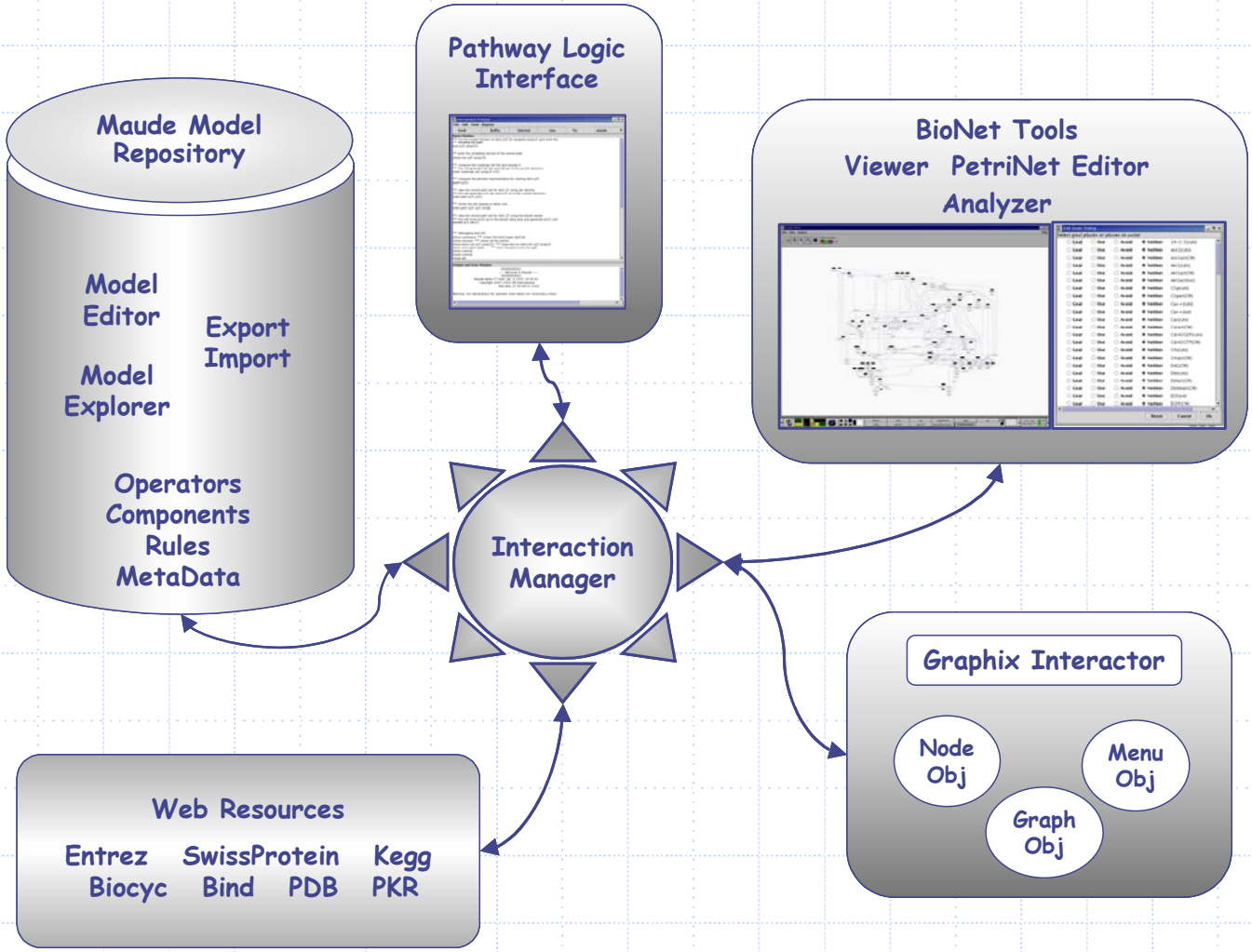
Some Level II Maude Code (in case you're interested)

```
rl[Raf1#4.Ras]:
  {CM | cm PS PA
       [?Ras - GTP]
       [Raf1    | (S 43), (S 259), (S 338), (Y 341),
                  (S 621 - phos - bound), (C1 - bound), (PABM - bound),
                  RBD, raf1:Atts]
       [14-3-3a | (SBD - bound), (DMD - bound), 1a:Atts]
       [14-3-3b |  SBD         , (DMD - bound), (T 141 - phos)]
       e((Raf1,  (S 621)), (14-3-3a,SBD))
       e((Raf1,   C1    ), b(PS))
       e((Raf1,   PABM  ), b(PA))
       e((14-3-3a,DMD   ), (14-3-3b,DMD))
  {cyto }}
=>
  {CM | cm PS PA
       [?Ras    | GTPbound, (RafBD - bound)]
       [Raf1    | (S 43), (S 259), (S 338), (Y 341),
                  (S 621 - phos - bound), (C1 - bound), (PABM - bound),
                  (RBD - bound), raf1:Atts]
       [14-3-3a | (SBD - bound), (DMD - bound), 1a:Atts]
       [14-3-3b |  SBD         , (DMD - bound), (T 141 - phos)]
       e((Raf1,   (S 621)), (14-3-3a,SBD))
       e((Raf1,    C1    ), b(PS))
       e((Raf1,    PABM  ), b(PA))
       e((14-3-3a, DMD   ), (14-3-3b,DMD))
       e((Raf1,    RBD   ), (?Ras,    RafBD ))
  {cyto }} .
```

… and the pictorial representation (binding of GTP-bound Ras to Raf-1)



Future Directions: The Pathway Logic Assistant



The Pathway Logic Assistant—making Pathway Logic accessible to biologists and biomedical researchers

References

1) W. Kolch. The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. Biochem J 404: 3-9, 2002.

2) S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and K. Sonmez. Pathway Logic: Symbolic analyses of biological signaling. *In* Pacific Symposium on Biocomputing, 2002, pp. 400-12.

3) Steven Eker, Merrill Knapp, Patrick Lincoln, Keith Laderoute, and Carolyn Talcott. Pathway Logic: Executable models of biological networks. *In* 4th International Workshop on Rewriting Logic and its Applications, 2002.

4) C. Talcott, S. Eker, M. Knapp, P. Lincoln, and K. Laderoute. Pathway Logic modeling of protein functional domains in signal transduction. *In* Pacific Symposium on Biocomputing, 2004 (http://psb.stanford.edu/psb-online/).

5) Maude Websites: http://maude.cs.uiuc.edu
http://www.csl.sri.com