

FCModeler: Dynamic Graph Display and Fuzzy Modeling of Regulatory and Metabolic Maps

Julie Dickerson¹, Zach Cox¹ and
Andy Fulmer²

¹*Iowa State University* and ²*Proctor &
Gamble.*

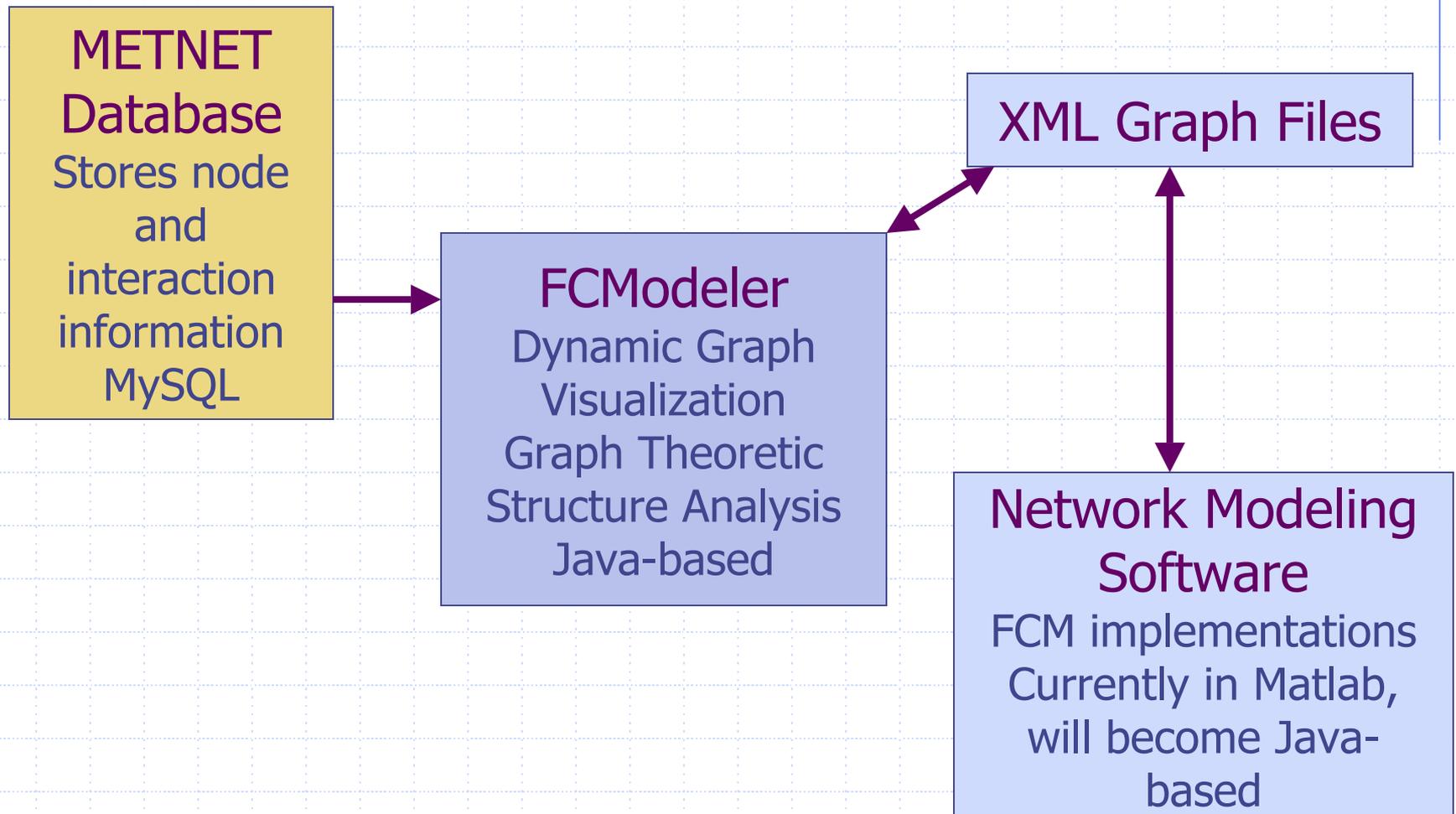
FCModeler Goals

- ◆ Capture the intuitions of biologists and provide a modeling framework for assessing large amounts of information
- ◆ Test effects of hypotheses
- ◆ Use graph theoretic approaches to analyze network structure and behavior and search for critical paths in the network
- ◆ Develop simulation tools based on fuzzy methods that model changes in the network

Key Components of FCModeler

- ◆ Dynamic graph visualization package written in Java
- ◆ Network analysis using graph theoretic methods
- ◆ Network modeling using fuzzy cognitive maps

Software Structure



Current Efforts

◆ FCModeler Development

- Incorporation of different levels of information such as functional links
- Finding critical points in interacting pathways
- Network analysis and parallel paths between nodes
- Visualization of graphs in 3D virtual reality systems

Acknowledgements

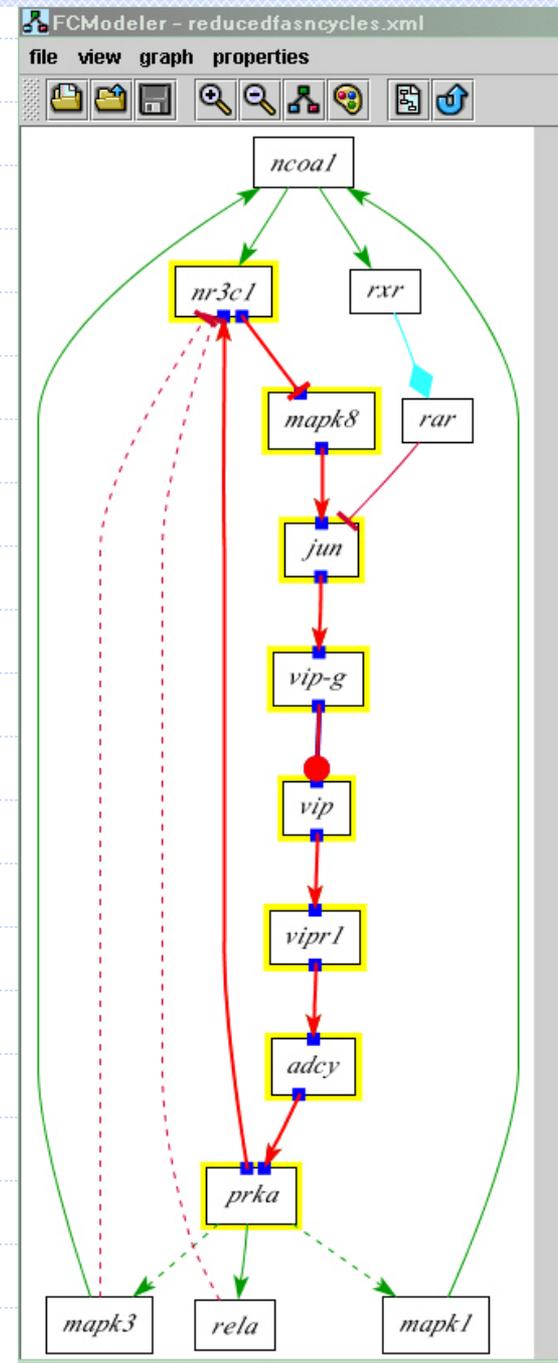
- ◆ This work has been supported by
 - Proctor and Gamble Corporation
 - National Science Foundation
 - Iowa State University Plant Sciences Institute
- ◆ For more information see:
www.eng.iastate.edu/~julied/research

Dynamic Graph Visualization

- ◆ Display graph data using different layouts
- ◆ Select subgraphs for further analysis
- ◆ Visualize simulation results

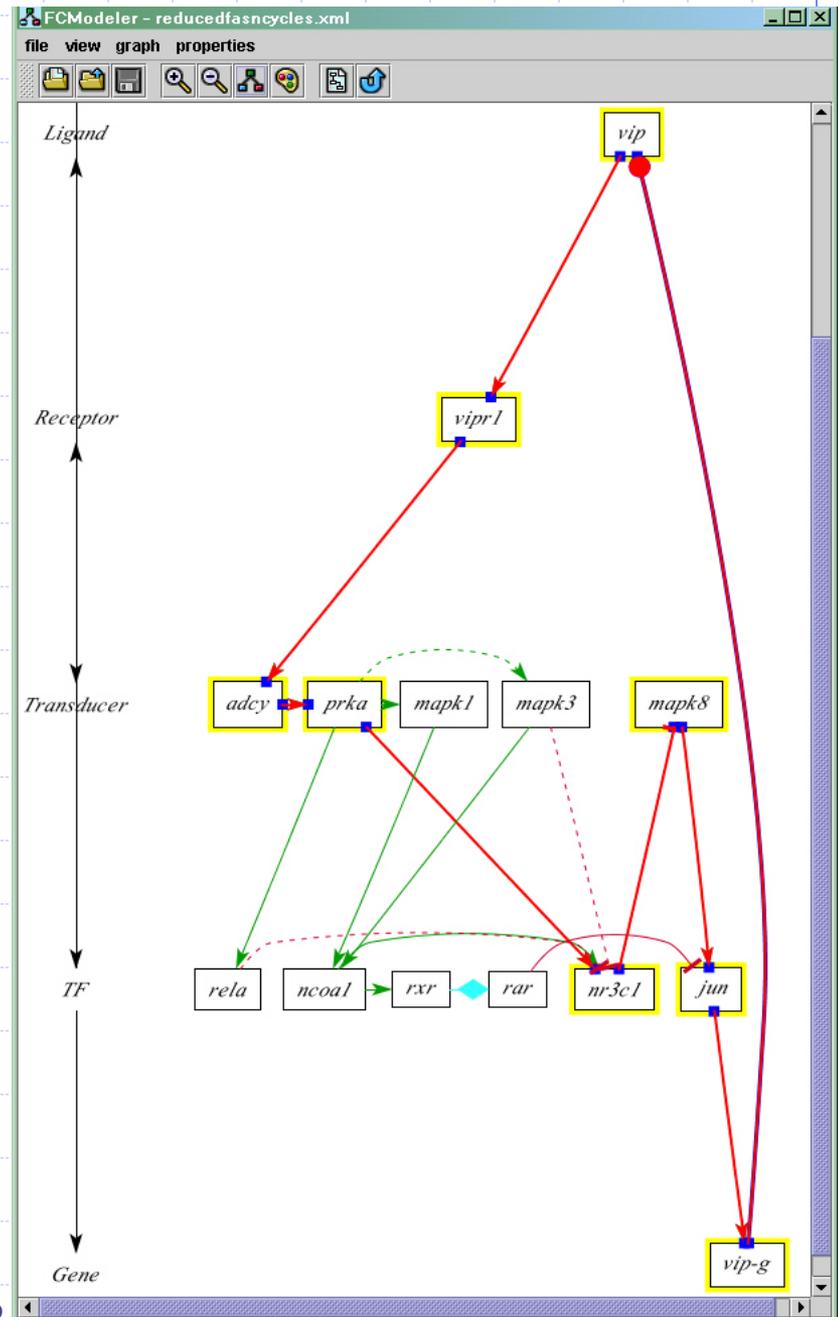
Graph Layout, *Dot*

- ◆ *dot* from *Graphviz* package - AT&T research labs (<http://www.research.att.com/sw/tools/graphviz/>)
- ◆ Developed for noncyclic graphs.
- ◆ Good for hierarchical data, handles large graphs well



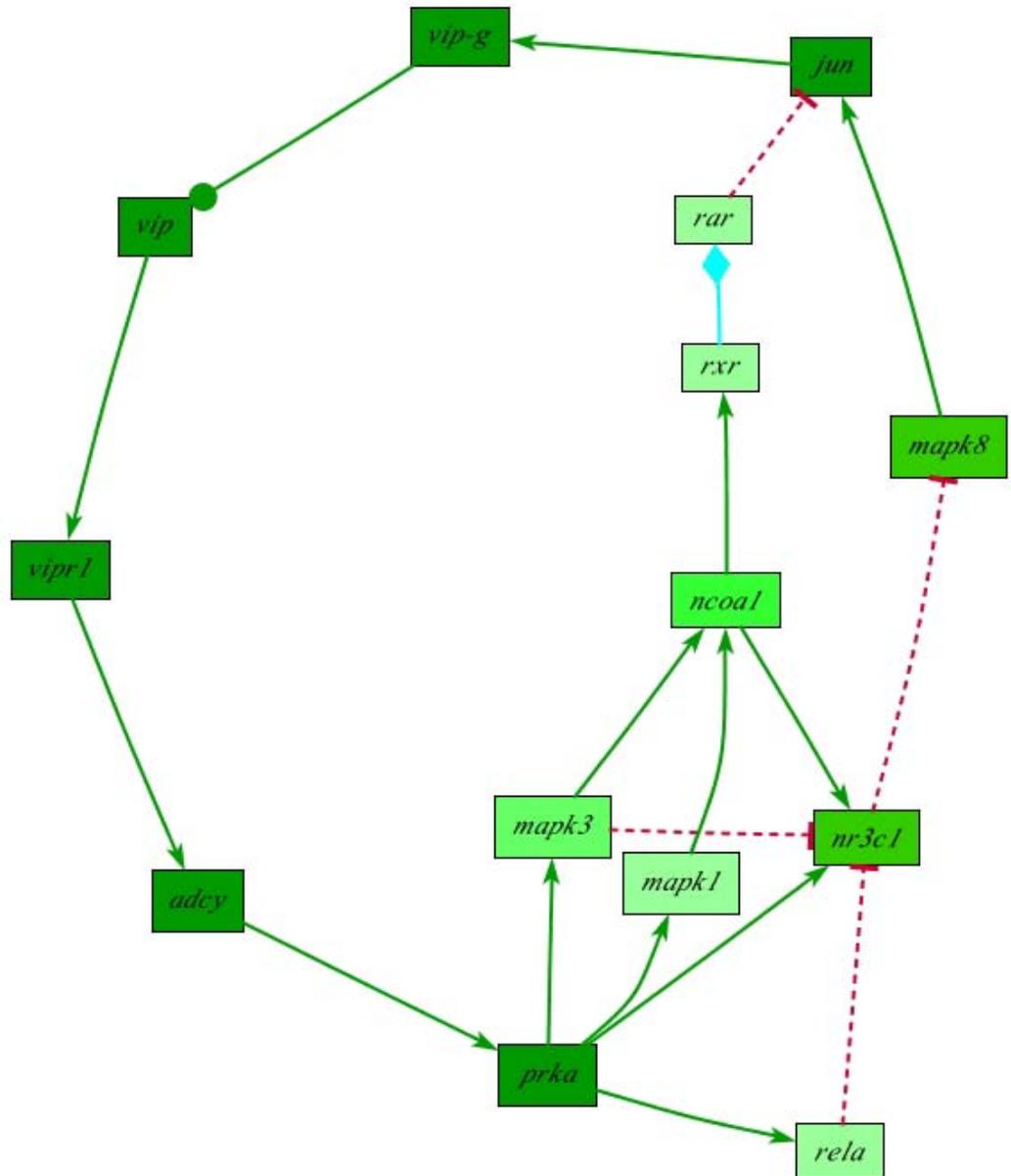
Ranks and Clusters

- ◆ Elements can be separated into different levels based on cellular location, type, etc
- ◆ More intuitive
- ◆ Slower layout, doesn't always converge



GEM Layout

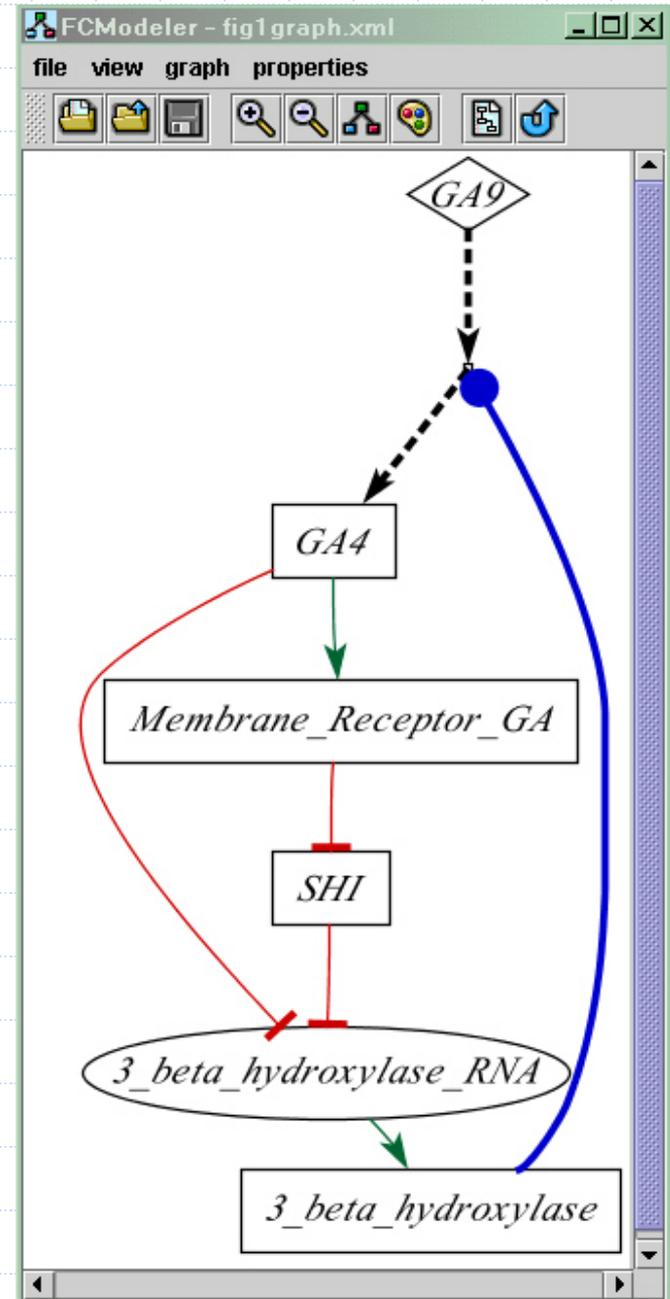
- ◆ Spring Embedder based layout,
- ◆ Places attractive gravitational force at the barycenter of the layout.
- ◆ Nodes are attracted to adjacent nodes to it and repelled by other node
- ◆ Shows cyclic structures well



Based on "A Fast Adaptive Layout Algorithm for Undirected Graphs" by Frick, Ludwig and Mehldau
ISMB 2002

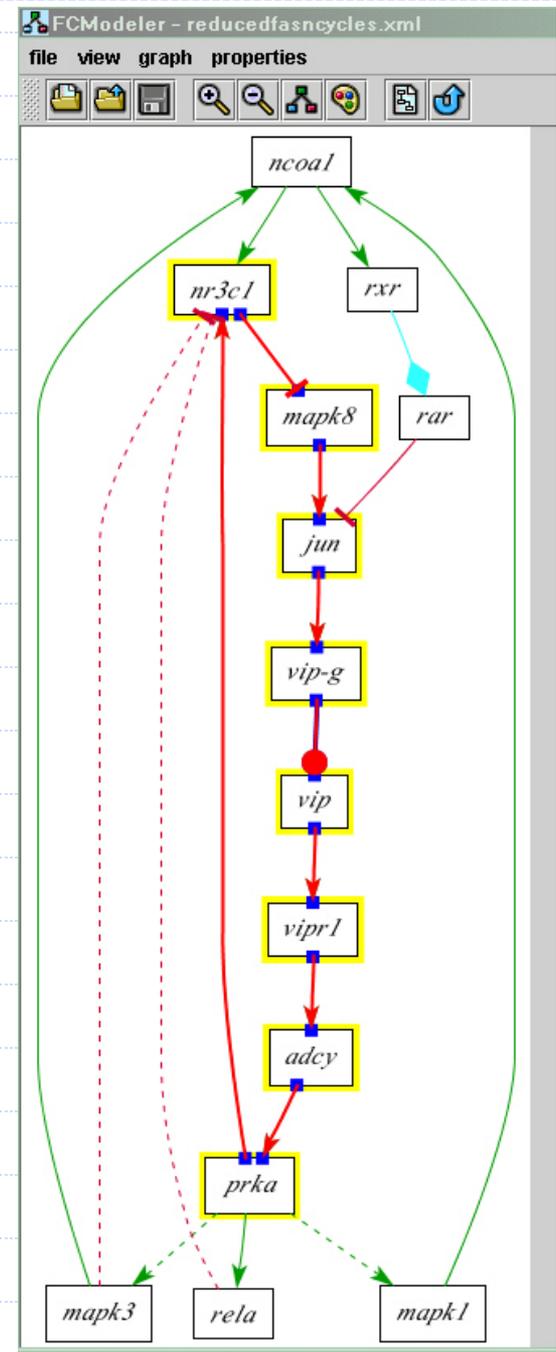
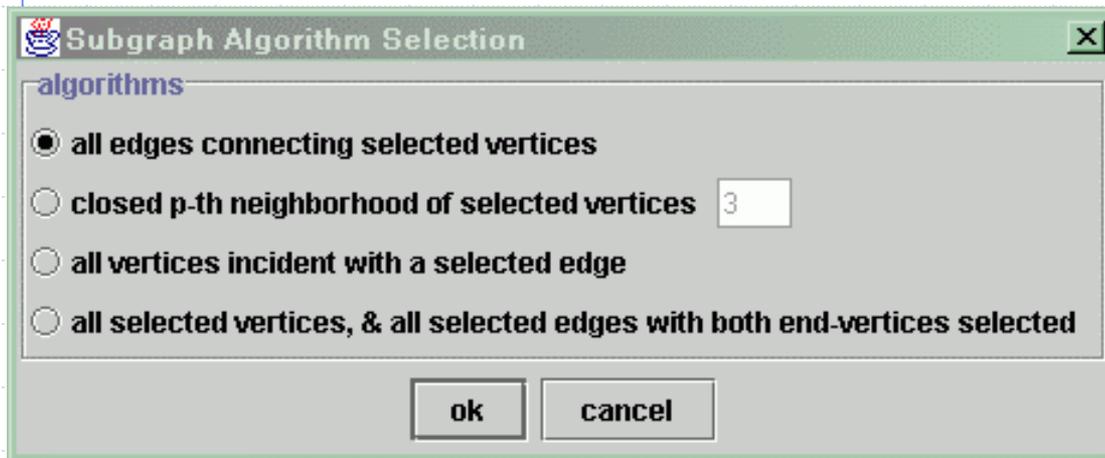
Edge and Node Mappings

- ◆ User settable mappings for edges and nodes
 - Color
 - Width
 - Arrowhead type
 - Node shape
- ◆ User can move nodes and edges and store the coordinates for future reference



Subgraph Selection

- Selectable subsets for further analysis based on vertices, cycles, etc



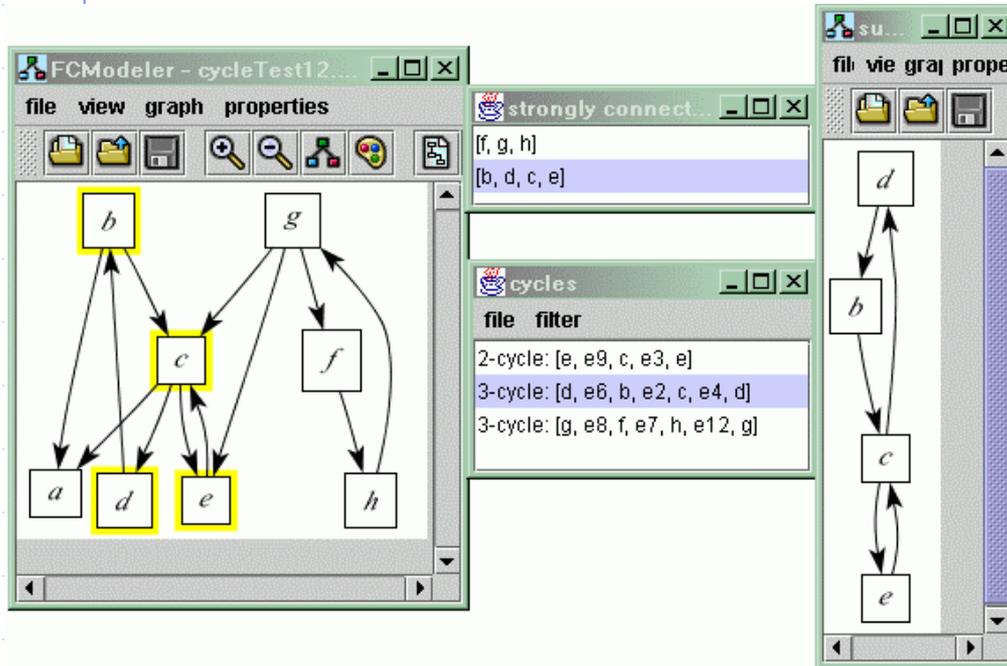
Dynamic Graph Visualization

- ◆ Java™ interface that reads and displays data from a mySQL (www.mysql.com) database of links and nodes.
- ◆ Based on DIVA package for node/edge features www-cad.eecs.berkeley.edu/diva/index.html
- ◆ Uses Java 1.4 or higher

Graph Theoretic Analysis

- ◆ Strongly Connected Components
- ◆ Cycle Analysis and Self Organizing Maps
- ◆ Alternate Path Analysis

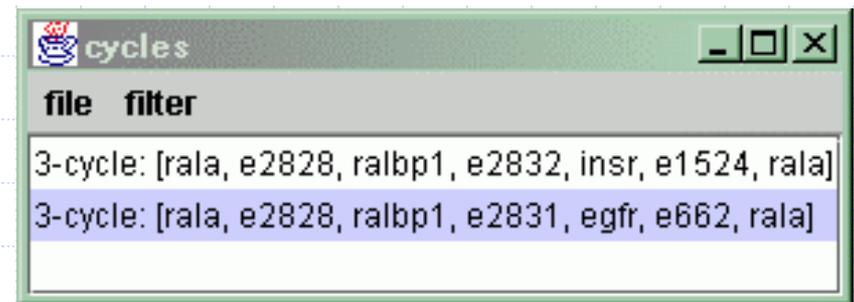
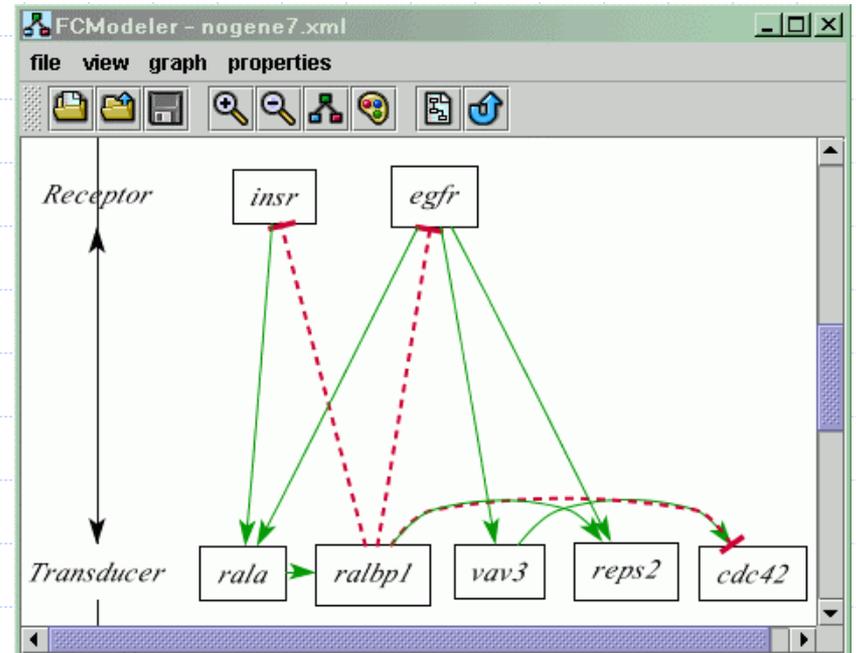
(SCC) Strongly Connected Components



- ◆ Every node is reachable from every other node in the same SCC.
- ◆ In metabolic pathway graphs, the SCC's show which components have bi-directional paths between them.
- ◆ Only nodes in an SCC can be in cycles

Cycle Search

- ◆ Elementary circuits, or cycles, are a series of nodes that start and end in the same place.
- ◆ Preliminary results have found existing pathways in the map as well as new relationships and oversights in the models.
- ◆ Cycles in a graph show how one node can affect another



Cycle Similarity Metrics

- ◆ Many overlapping cycles, need to be grouped
- ◆ Sample Metrics

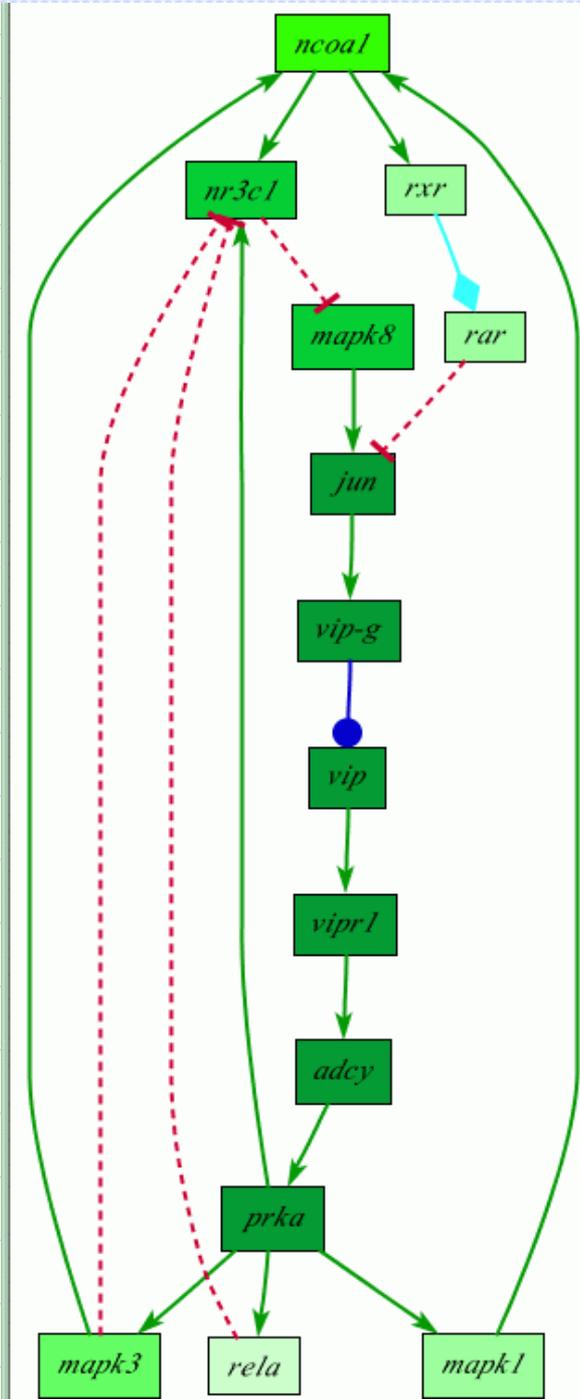
- Common Elements

$$d(A, B) = \frac{\text{card}(A \cap B)}{\text{card}(A \cup B)}$$

- Overlap (subsethood)

$$S(A, B) = \frac{\text{card}(A \cap B)}{\text{card}(B)}; \quad S(B, A) = \frac{\text{card}(A \cap B)}{\text{card}(A)}$$

- Edit Distance



Cycle Similarity Analysis

```

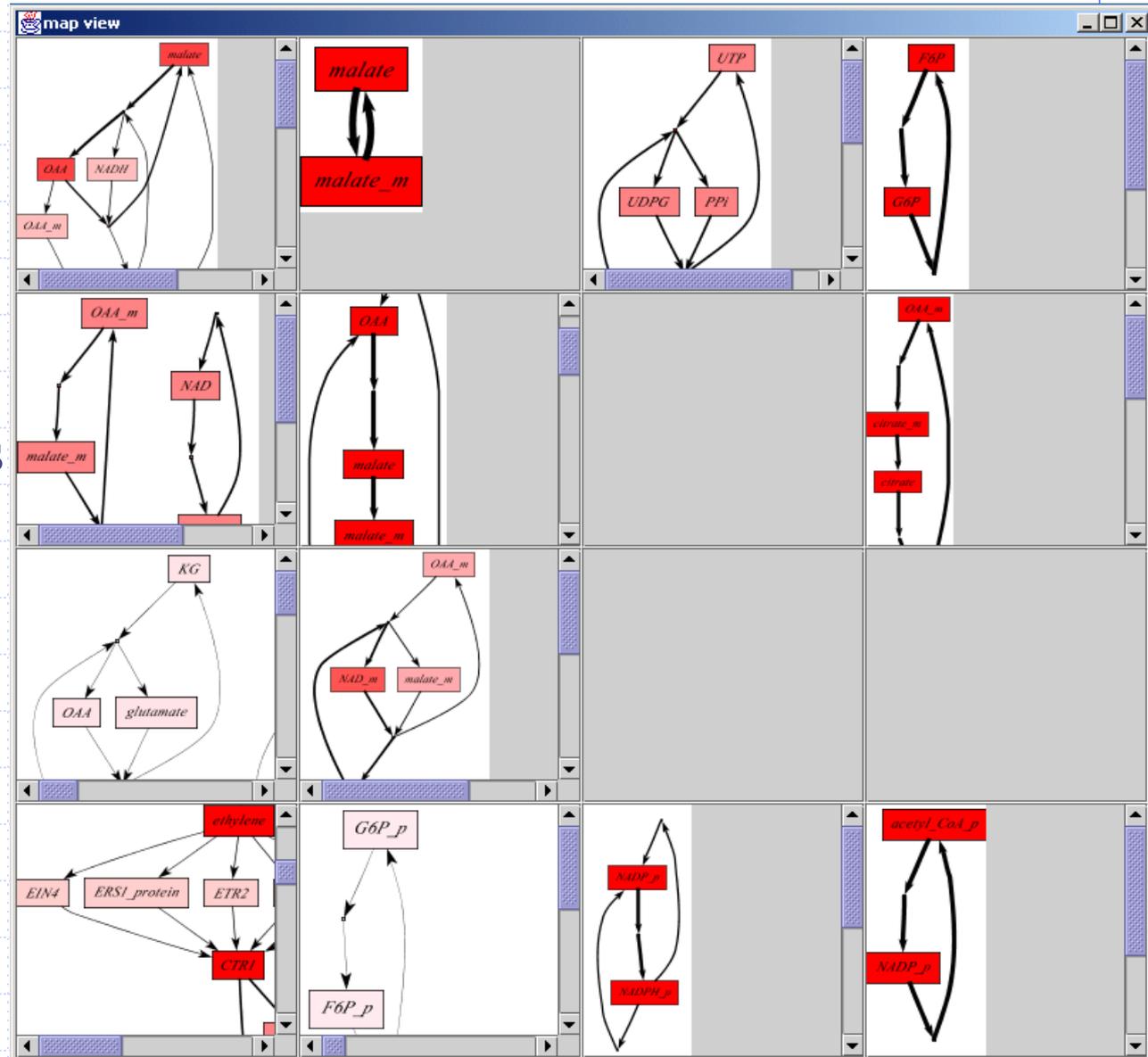
cycles
file filter
8-cycle: [prka, e41, nr3c1, e93, mapk8, e424, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka]
9-cycle: [mapk3, e24, nr3c1, e93, mapk8, e424, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e111, mapk3]
10-cycle: [mapk3, e98, ncoa1, e37, nr3c1, e93, mapk8, e424, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e111, mapk3]
10-cycle: [rar, e440, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e111, mapk3, e98, ncoa1, e103, nxr, e334, rar]
10-cycle: [rar, e440, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e109, mapk1, e97, ncoa1, e103, nxr, e334, rar]
10-cycle: [mapk1, e97, ncoa1, e37, nr3c1, e93, mapk8, e424, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e109, mapk1]
9-cycle: [rela, e43, nr3c1, e93, mapk8, e424, jun, e409, vip-g, e278, vip, e279, vipr1, e180, adcy, e76, prka, e112, rela]
    
```

Common Element Similarity Metric

Cycle Number	1	2	3	4	5	6	7
1	0	0.89	0.80	0.50	0.50	0.80	0.89
2	0.89	0	0.90	0.58	0.46	0.73	0.80
3	0.80	0.90	0	0.67	0.54	0.82	0.73
4	0.50	0.58	0.67	0	0.82	0.54	0.46
5	0.50	0.46	0.54	0.82	0	0.67	0.46
6	0.80	0.73	0.82	0.54	0.67	0	0.73
7	0.89	0.80	0.73	0.46	0.46	0.73	0

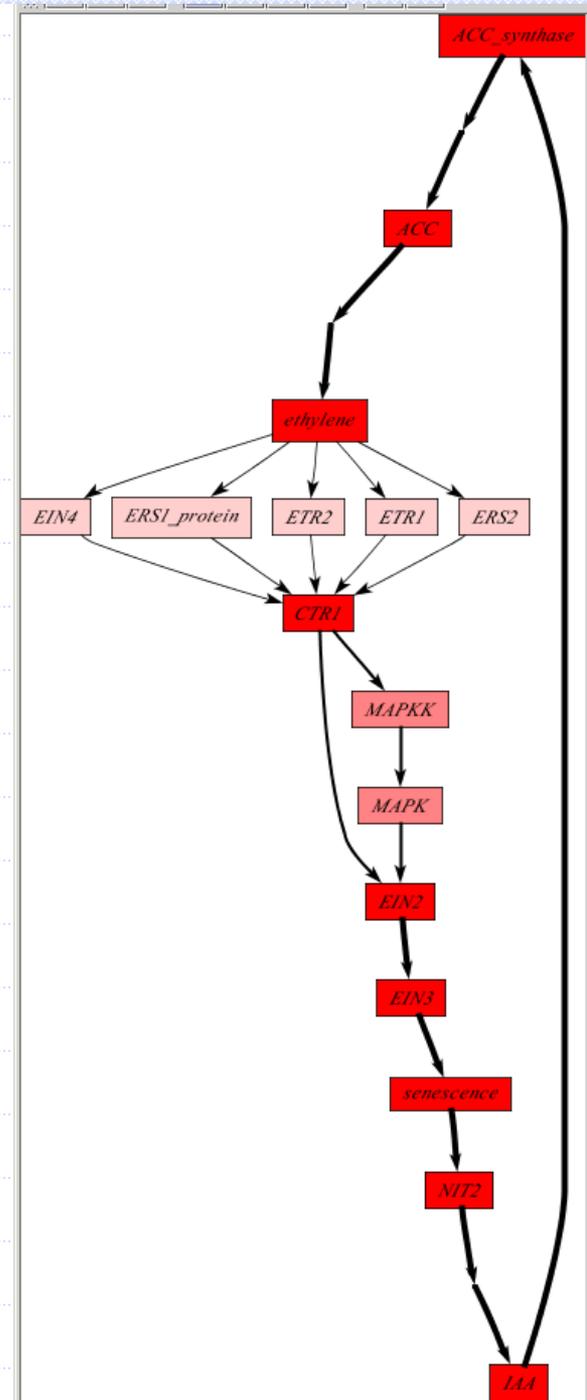
Clustering Cycles

Each graph shows the model of the corresponding map unit, which is the generalized median of the cycles assigned to that map unit.



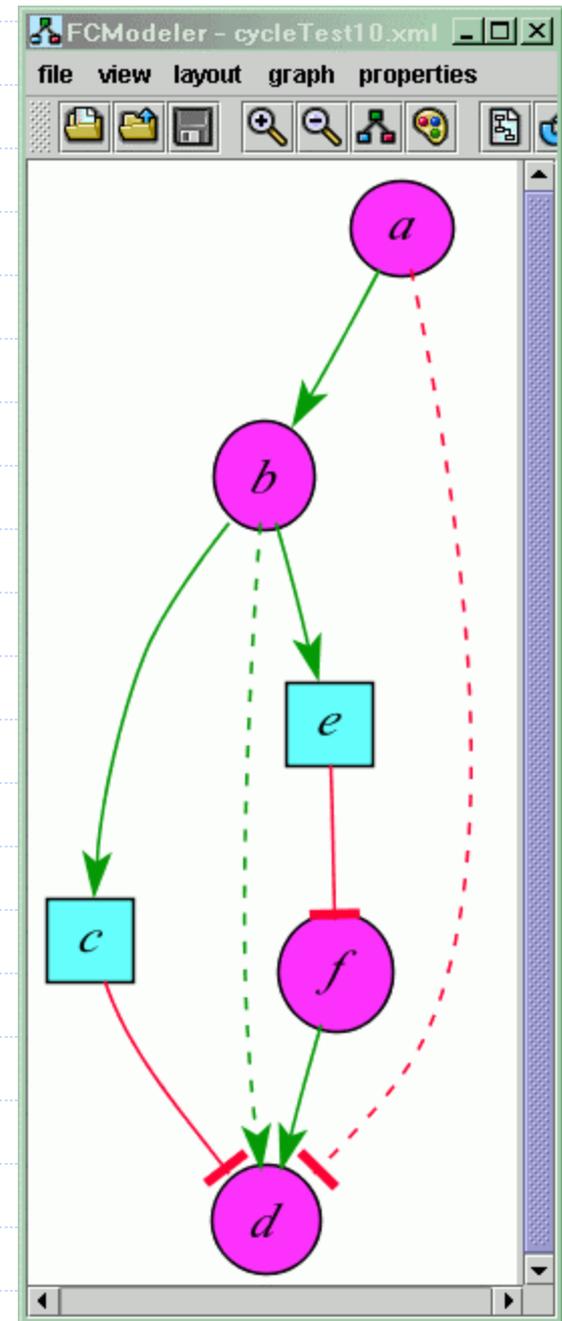
Results of cycle clustering

- Similar cycles with overlap are found
- Shows parallel paths in the data structure



Alternate Paths

- ◆ Often partial information is known (e.g. a inhibits d)
- ◆ Checks for other paths that go from a to d and checks overall effect
 - $a \rightarrow b \rightarrow c \rightarrow d$ has a net inhibitory effect
- ◆ Helps check for consistency (e.g. b activates d), alternate paths only inhibit d so information is missing

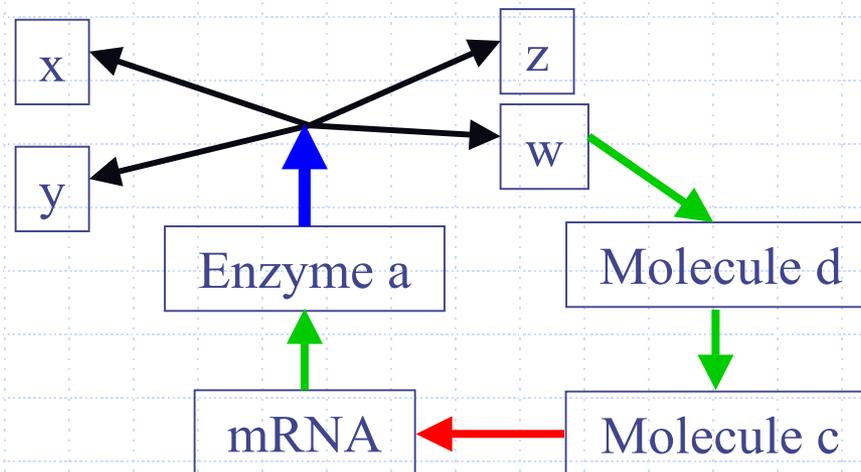


Metabolic Network Modeling

- ◆ **Metabolic networks** model net accumulation of biomolecules in organisms.
- ◆ **Regulatory networks** modulate the action of metabolic networks
 - **Nodes** represent specific biochemicals such as proteins, RNA, and small molecules, or stimuli, such as light, heat, or nutrients.
 - **Links** show interactions between nodes

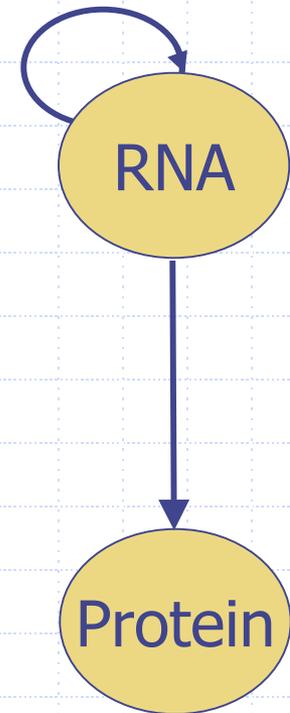
Link Types

- ◆ **Conversion link** (black arrow), a node is converted into another node, and used up in the process.
- ◆ **Regulatory link** (green and red arrows), node activates or deactivates another node, not used up.
- ◆ **Catalytic link** (blue arrows) an enzyme that enables a chemical conversion and not used up in the process.



Regulatory Link Models

- ◆ Edge weights $\{-1/N, 0, 1/N\}$, where N is the number of nodes going into the link
- ◆ As more information becomes known about the underlying biology, the functional link models will be updated.
- ◆ Regulatory input nodes have self-feedback to keep the nodes on until they are inhibited.

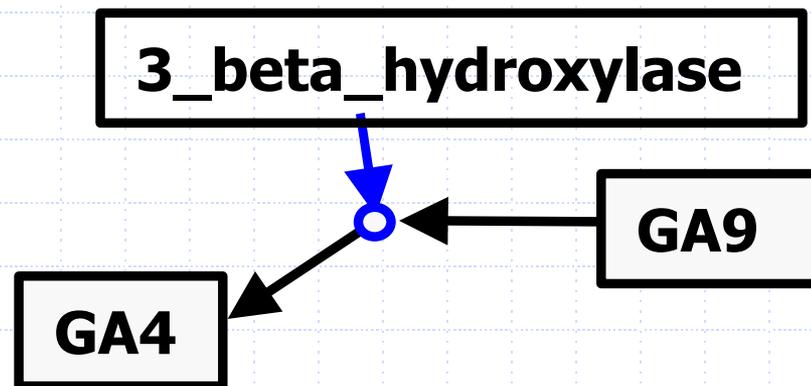


Conversion Link Models

- ◆ For causal relationships between nodes the link is modeled the same as a regulatory link
- ◆ When information about the rate of change in a reaction is available:
 - Simple difference equations can model the gradually rising and falling levels of the nodes
 - Step size depends on the reaction rate and the relationship between the nodes

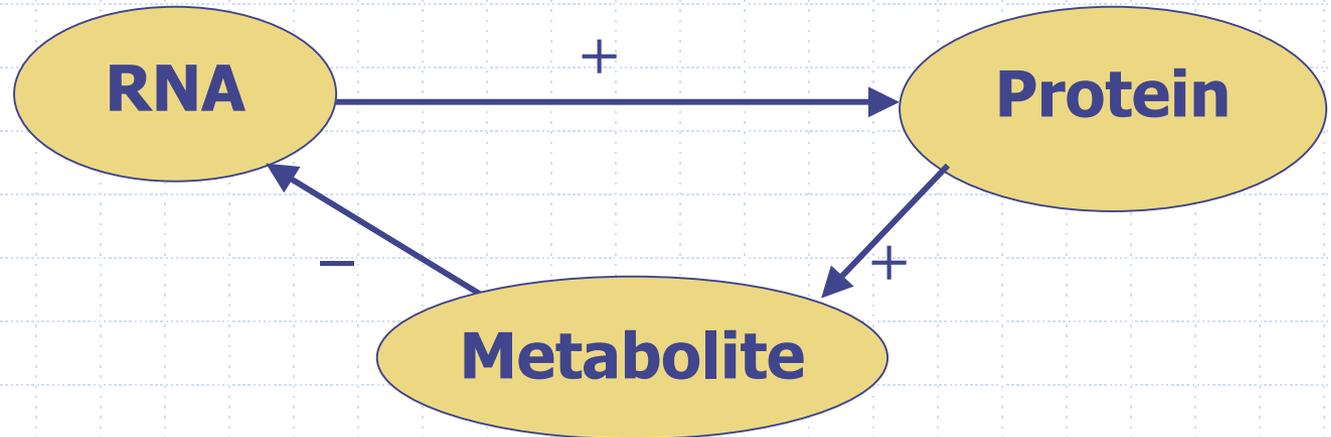
Catalyzing Link Models

- ◆ Catalyzed reactions add a dummy node that acts upon a conversion link. This allows one link to modify another link.
- ◆ Another method is an augmented edge matrix that operates on the edges between the nodes.



Fuzzy Cognitive Maps

- ◆ Fuzzy Cognitive Maps show interactions between different variables
 - Fuzzy signed digraphs represent causal flow between objects or concepts
 - Constructed using expert knowledge or neural learning

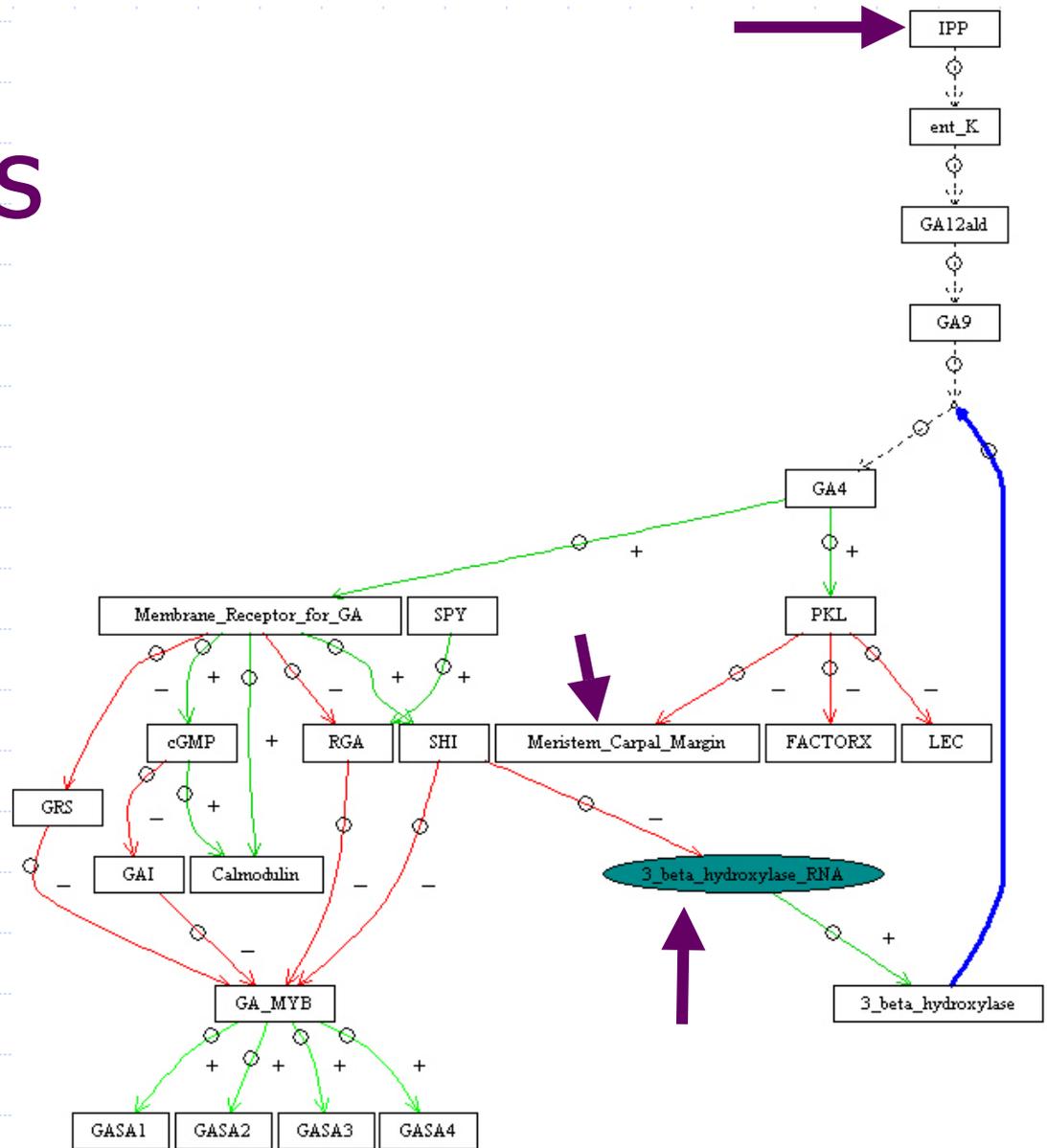


Gibberellin Example

- ◆ Create metabolic map of system using expert knowledge
- ◆ Model the activation and regulation of Gibberellin (GA)
- ◆ Metabolites are held constant until inhibited by Gibberellin

Assumptions

- ◆ RNA assumed active until inhibited
- ◆ Precursor small molecule IPP present
- ◆ Metabolites and other factors active unless inhibited



FCM of Gibberellin Activation

- ◆ Model shows the regulation and homeostatic control of Gibberellin(GA):
 - Regulates its own production
 - Controls the activity of other biomolecules
- ◆ Oscillation of GA levels directs the generation of biomolecules that are implicated in the formation of new cellular proliferation centers, (meristems)
- ◆ Many key features of this model, including timing, can be tested experimentally by globally monitoring temporal profiles of mRNA, protein, and metabolite

Metabolic Networks

- ◆ Metabolic networks form the basis for the net accumulation of biomolecules in organisms.
- ◆ Regulatory networks modulate the action of metabolic networks, leading to physiological and morphological changes.
- ◆ Modeling tool represents the interactions within and between these networks
 - **Nodes** represent specific biochemicals such as proteins, RNA, and small molecules, or stimuli, such as light, heat, or nutrients.
 - **Links** show interactions between nodes

Simple Fuzzy Cognitive Maps

- ◆ Graph edges are $\{-1,0,1\}$. Used when the direction of causality is agreed on, but not its degree
- ◆ Concepts either occur or do not occur
- ◆ Can test out hypotheses
- ◆ Concepts are usually summed then thresholded to get the next state

$$\mathbf{C}(t_{n+1}) = S[\mathbf{C}(t_n)\mathbf{E}]$$

$$S(\mathbf{y}) = \frac{1}{1 + e^{-c(\mathbf{y}-\mathbf{T})}}$$

Nested FCMs

- ◆ If more information is known about the links between concepts, then more detailed functional links can be used to combine information
 - Differential equations
 - Fuzzy approximators

