

# Using Pathway Logic to Integrate Signal Transduction and Gene Expression Data

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

SRI International

Carolyn Talcott

Linda Briesemeister

Steven Eker Merrill Knapp

Patrick Lincoln Andy Poggio

Keith Laderoute

SRI International

Marti Jett Rasha Hammamieh

WRAIR

Pathway Logic (<http://pl.csl.sri.com/>) is an approach to the modeling and analysis of molecular and cellular processes based on rewriting logic. Pathway Logic (PL) models reflect the ways that biologists think about problems using informal models. They are curated from the literature, and written and analyzed using Maude, a rewriting-logic-based formalism. A Pathway Logic knowledge base includes data types representing cellular components such as proteins, small molecules, complexes, compartments/locations protein state, and post-translational modifications. Modifications can be abstract, just specifying being activated, bound, or phosphorylated, or more specific, for example, phosphorylation at a particular site. Collections of entities, treated as 'liquid' mixtures, are represented as multisets (unordered collections).

Rewrite rules describe the behavior of proteins and other components depending on modification state and biological context. Each rule represents a step in a biological process such as metabolism or intra/inter-cellular signaling. A specific model is assembled by specifying an initial state (called a dish): the cells, their components, and entities such as ligands in the supernatant.

The Pathway Logic Assistant (PLA) provides an interactive visual representation of PL models. Using PLA one can

- \* display the network of signaling reactions for a specified model;
- \* formulate and submit queries to find pathways, for example activating one protein without activating a second protein, or exhibiting a phenotype signature such as apoptosis;
- \* compare two pathways;
- \* compute and display the subnet relevant to one or more proteins;
- \* visualize gene expression data in the context of a network.

SRI in collaboration with WRAIR is developing a novel computational approach to identifying patterns of host cell transcription that serve as early markers of infection for disease diagnosis. The approach is based on integration of the Pathway Logic model of cellular signaling, and an information theoretic approach called test set generation, used to select genes of interest from gene expression data.

## The PL model of LPS - TLR4 Signaling

The model is generated from the PL knowledge base by specifying the relevant cell components (proteins, such as Tlr4 and small molecules, such as Lps) and computing the reachable (Figure 1).

How is signaling representing in the PL knowledge base? A cell and its ligands are represented as a term

ligands [cellType | locations]

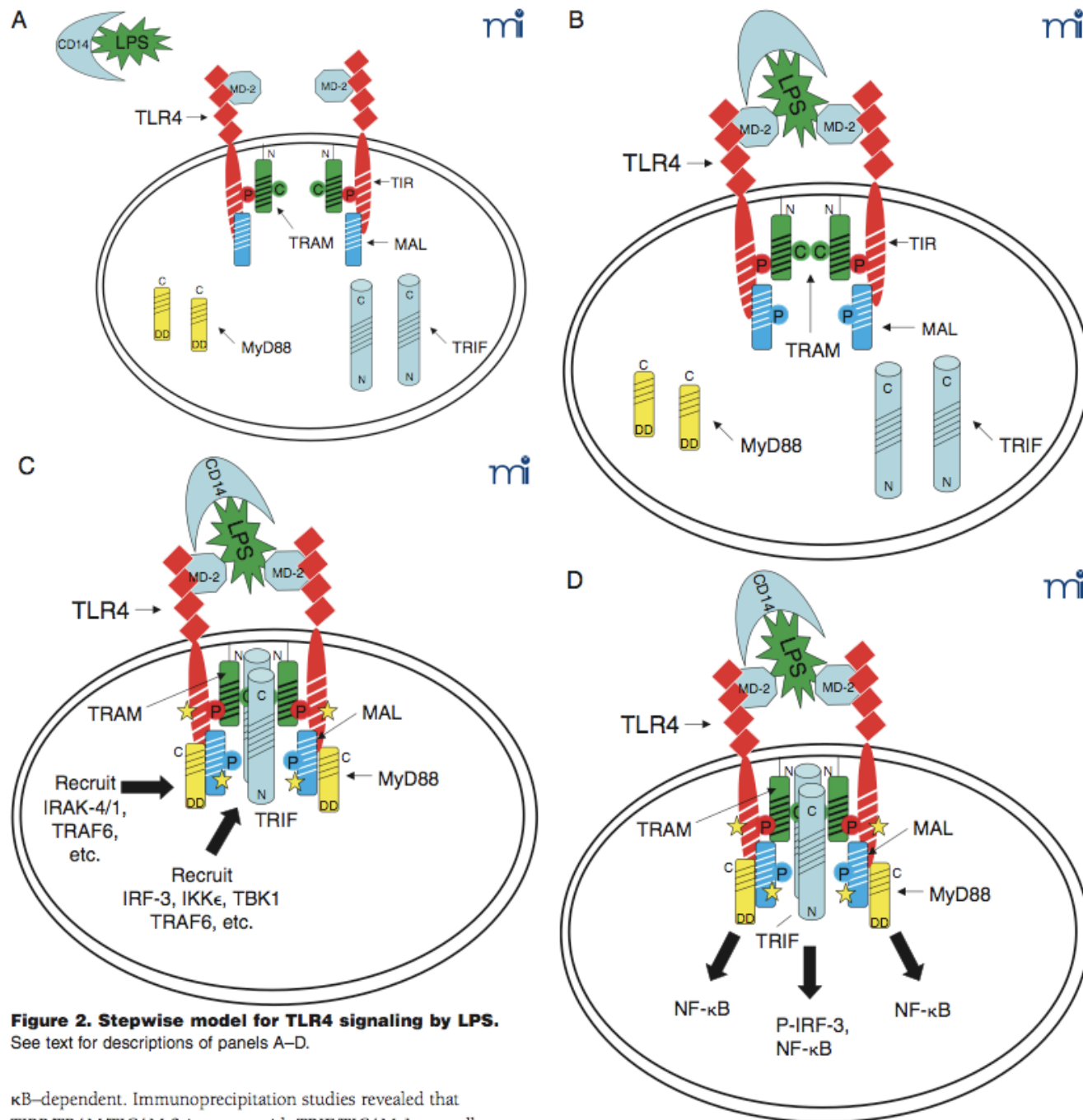
Each location has the form

{ locationName | components }

A signaling rule has the form

cellStateBefore => cellStateAfter





**Figure 2. Stepwise model for TLR4 signaling by LPS.**  
See text for descriptions of panels A–D.

$\kappa$ B-dependent. Immunoprecipitation studies revealed that TRIP/TRAM/TICAM-2 interacts with TRIF/TICAM-1, as well as

The cartoon shows initial steps in the LPS-TLR4 signaling pathway. Below are PL representation of some of these steps. (Cartoon names for components are shown in parens.)

Circulating LBP recognizes LPS in the plasma and brings it to CD14. This aids the loading of LPS onto the LPS receptor complex, which is composed of dimerized TLR4 receptors and two molecules of the extracellular adapter Ly96 (MD-2).

Activation of TLR4 causes the adaptor protein Tirap (Mal) to be recruited to the receptor complex [rl.208].

(The cartoon shows Mal being phosphorylated, but no evidence is presented.)

rl[208.Tirap.by.TLR4]:

{CLm | clm [TLR4 - act] }

{CLi | cli }

{CLc | clc Tirap } \*\*\*\* CLc -- the cell cytoplasm

=>

{CLm | clm [TLR4 - act] }

{CLi | cli [Tirap - reloc] }

{CLc | clc } .

Activated TLR4 also recruits Ticam2 (Tram) to the membrane [rl.859].

rl[859.Ticam2.by.TLR4]:

{CLm | clm [TLR4 - act]}

{CLi | cli}

{CLc | clc Ticam2}

=>

{CLm | clm [TLR4 - act]}

{CLi | cli [Ticam2 - reloc]}

{CLc | clc} .

Ticam2 (Tram) recruits Ticam1 (Trif) [rl.860]

rl[860.Ticam1.by.Ticam2]:

{CLm | clm [TLR4 - act]}

{CLi | cli [Ticam2 - reloc]}

{CLc | clc Ticam1}

=>

{CLm | clm [TLR4 - act]}

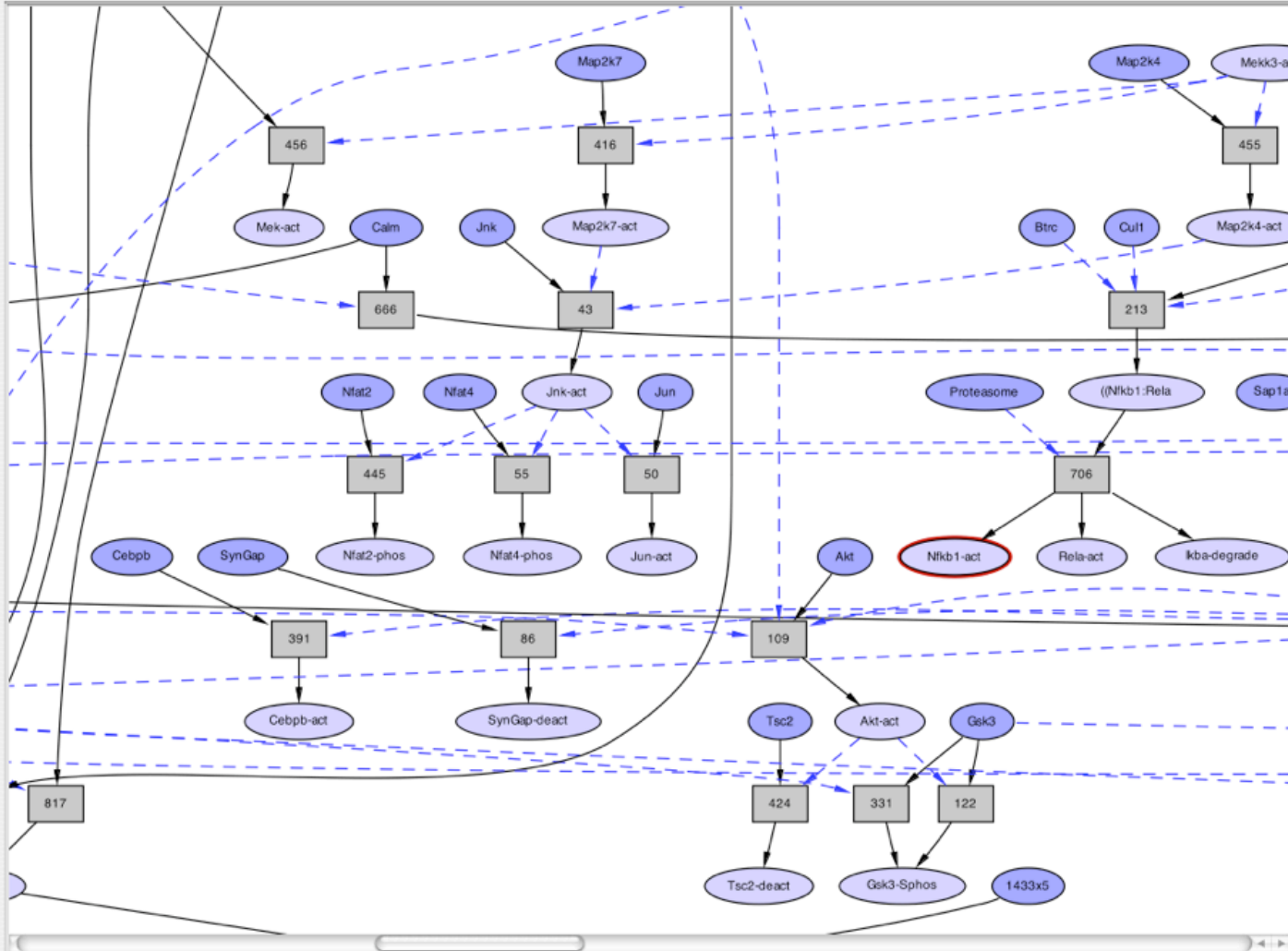
{CLi | cli [Ticam2 - reloc] [Ticam1 - reloc]}

{CLc | clc } .

Path (LoLA)

Path (Maude)

Subnet



Find Occurrence (⇧#O):

- Nfk1-act-NUC
- Map2k6-CLC
- Map2k7-act-CLC
- Map2k7-CLC
- Marcks-CLI
- Mef2-act-NUC
- Mef2-NUC
- Mek-act-CLC
- Mek-act-CLI
- Mek-CLC
- Mekk3-act-CLC
- Mekk3-CLC
- Mkk5-act-CLC
- Mkk5-CLC
- Mlc-CLC
- Mlc-phos-CLC
- Mnk-act-CLC
- Mnk-CLC
- Msk-act-NUC
- Msk-NUC
- Mtor-act-CLC
- Mtor-CLC
- Myd88-CLC
- Myd88-reloc-CLI
- Myk-act-CLC
- Myk-CLC
- (Nemo:(Ikk1:Ikk2-act))-f
- (Nemo:(Ikk1:Ikk2))-CLC
- Nfat2-CLC
- Nfat2-phos-CLC
- Nfat4-CLC
- Nfat4-phos-CLC
- Nfk1-act-NUC

Find Rule (⇧#R):

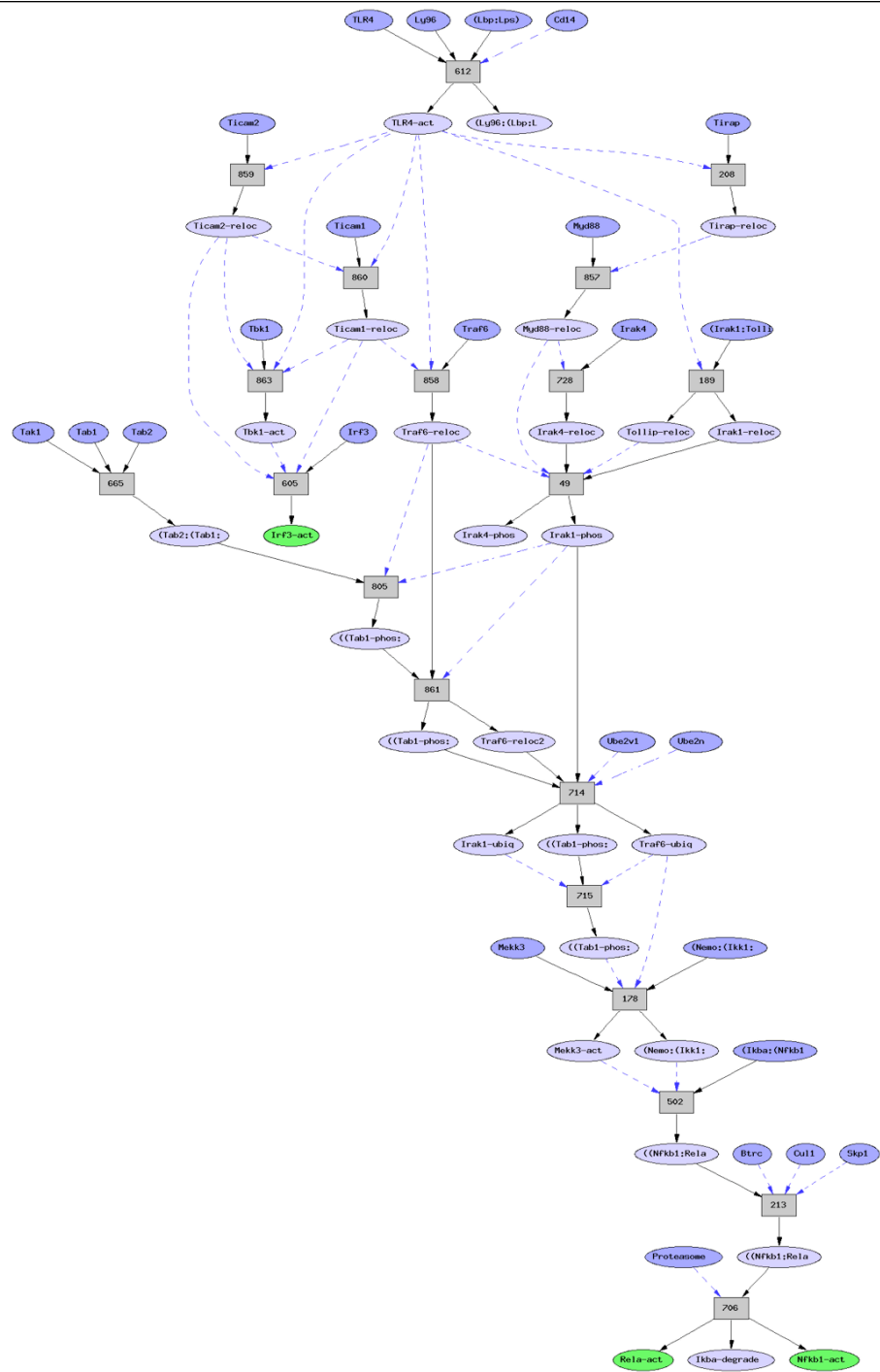
- 5.1433t.by.PkcZ
- 6.Hras.on.1
- 11.Sos.on
- 15.Tiam.on
- 17.Plce1.by.Rap1
- 19.Plce1.by.Ras
- 27.PA.by.PLD
- 37.RalGds.by.Ras
- 37.RalGds.by.Ras
- 38.Elmo.by.Dock
- 43.Jnk.on
- 49.Irak1.Irak4.ph
- 50.Jun.on
- 55.Jnk.inhib.Nfat
- 58.Srf.by.Limk
- 59.Adducin.by.Rc
- 63.Atf2.by.P38
- 64.Elk1.by.P38
- 65.ddit3.by.P38
- 69.Erk5.by.Mkk5
- 70.Tbk1.Traf2.co
- 81.Rafb.by.Rap1
- 84.IP3.by.Plc
- 84.IP3.by.Plc#1
- 86.SynGap.xby.C
- 107.PIP2.from.PI
- 108.Pdk1.by.PIP3
- 109.Akt1.on.1
- 110.Vav1.on
- 111.aPkc.on
- 122.Gsk3.xby.Ak
- 131.Chl.by.Pvk2

Find Selections Context Menu Info

Find Click

Find Click

**Figure 1. The LPS-Signaling network viewed using the Pathway Logic Assistant.** In the upper right is a thumbnail sketch of the full network. The area in the red rectangle is shown magnified in the main frame. Each numbered box represents a signaling rule. The ovals connected by arrows going into the box represent components of the *cellStateBefore* part of the rule (reactants). The ovals connected by arrows coming from the box represent components of the *cellStateAfter* part of the rule (products). An oval connected with a dashed arrow represents components that are unchanged in the reaction. Dark colored ovals represent components initially present.





**Figure 2. a pathway in the LPS signaling network that turns on the Ifnb gene (by activating Nfkb1, Rela, and Irf3).**

The pathway is represented as a network of reactions, which if fired as they become enabled lead to the goals.

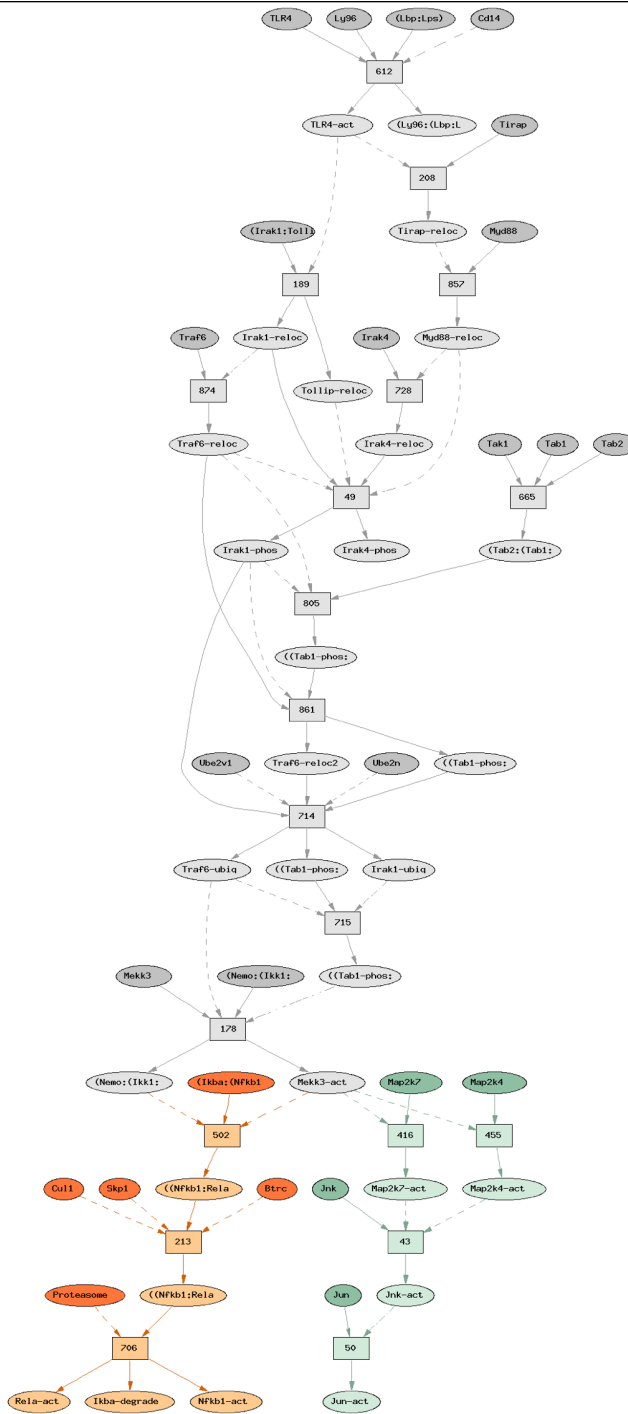


**Figure 3. A comparison of two pathways activating Jun in the LPS signaling network.**

Orange is the pathway stimulated by TLR2.

Green is the pathway stimulated by TLR4.

The gray part is shared by both pathways.



**Figure 4. A comparison of two pathways stimulated by TLR4.**

Orange is the activation of Nfkb1.

Green is the activation of Jun.

The gray part is shared by both pathways.

## Mapping host-response gene expression data onto the LPS network.

The data (provided by the Marti Jett Lab, WRAIR): blood cells collected from 67 donors were exposed to 8 pathogens and gene expression data was collected for 1185 genes at several (3-5) time intervals using microarrays.

Test set generation methods were used to chose a subset of at most 400 *signature* genes capable of discriminating amongst the 29 conditions (pathogen, exposure time).

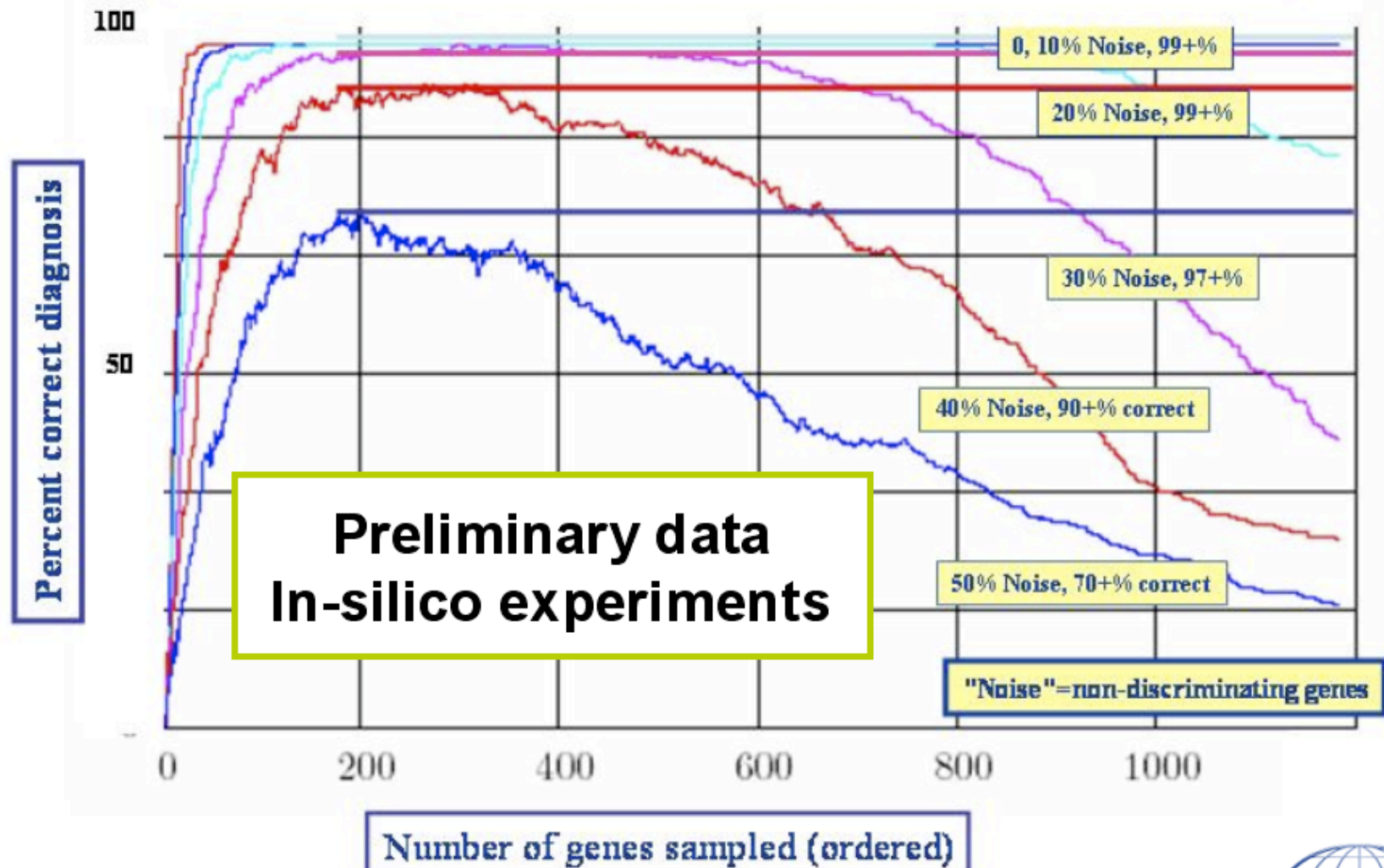
## **Finding signature genes.**

Step 1: assign regulation types to each gene, sample pair---a non-empty subset of the three possibilities 'up, down, same' was used based on results from 4 methods for normalization and thresholding.

Step 2: define a distance measure between conditions---sum over all genes the distance between regulation types.

Step 3: choose a subset of genes that maximizes the minimum distance (to maximize the number of errors that can be handled).

## IDENTIFICATION OF GENES THAT DISCRIMINATE AMONG BIOTHREAT PATHOGENIC AGENTS



Data from Jett at WRAIR





What role do the proteins coded by the signature genes play in the network? The top 250 signature genes were mapped onto the LPS network.

1. Find the proteins that appear in the network.
2. Find the subnet of reaction rules connected to these proteins (Figure 5a,b)
3. Compute the subnet impacted by these proteins (reaction rules that are unreachable in their absence) (Figure 6)

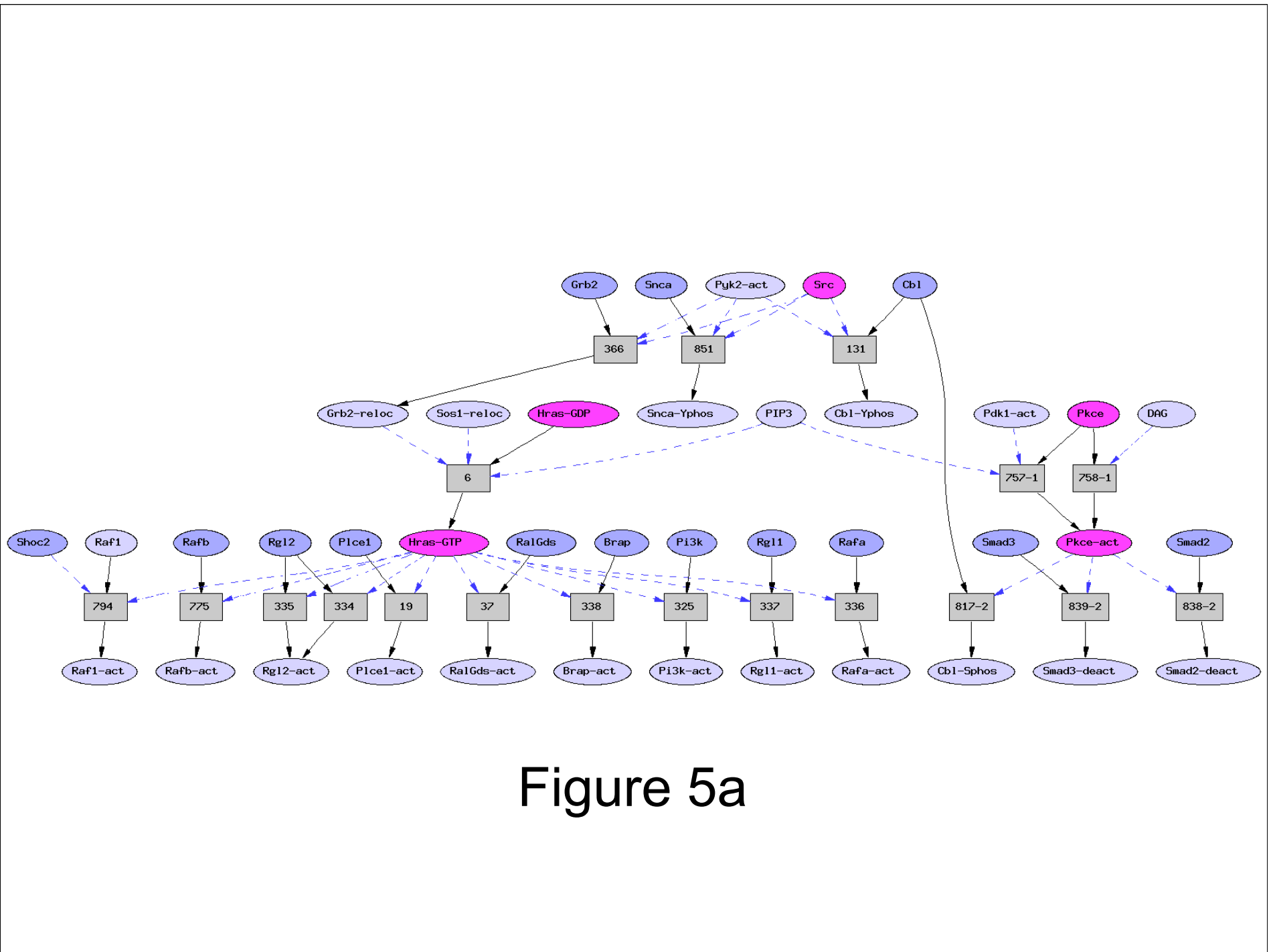


Figure 5a

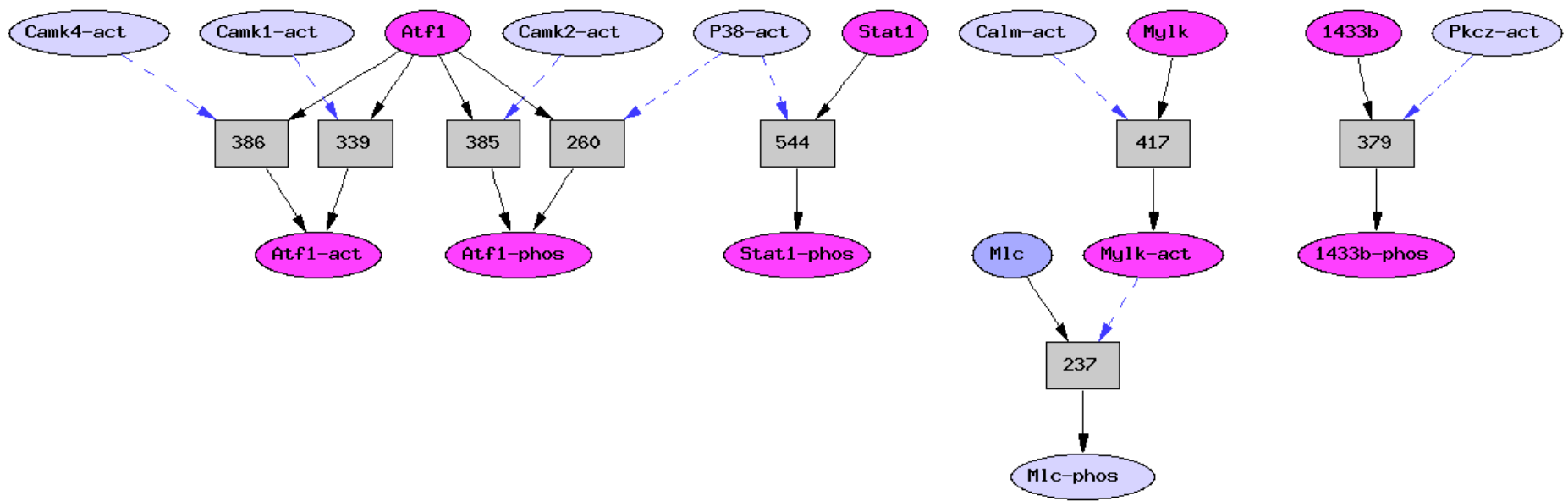


Figure 5b

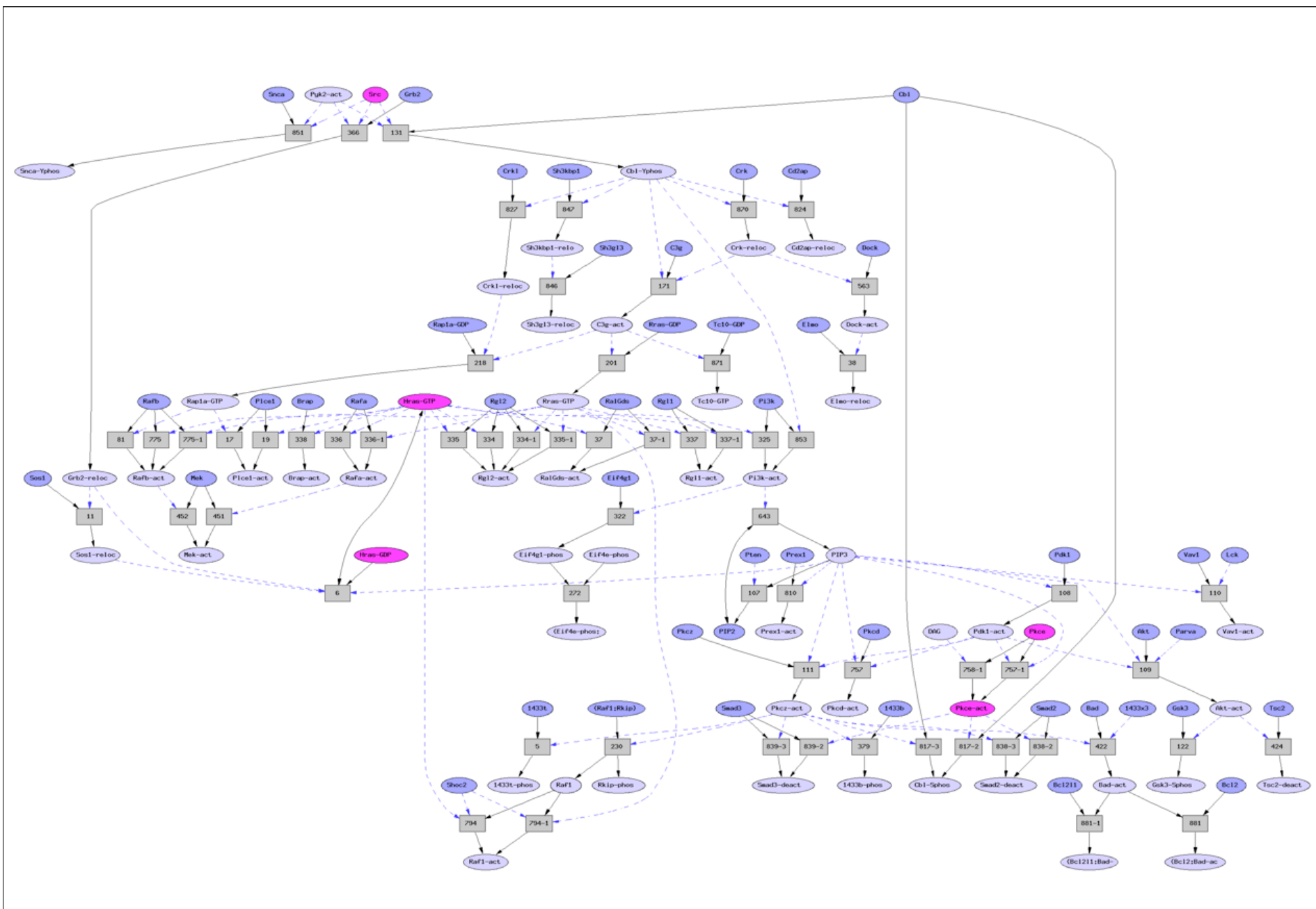


Figure 6

Figure 5a

Figure 5b

Figure 6