

# SYMBOLIC SYSTEMS BIOLOGY

---

## USING FORMAL METHODS TOOLS TO MODEL BIOLOGICAL PROCESSES

- <http://pl.csl.sri.com>

Carolyn Talcott  
SRI International  
SSFT May 2011



# PLAN

- Symbolic systems biology
- Pathway Logic
  - Modeling
  - Analyzing
- Case studies
- Choosing a formalism
- Future challenges



# BIOLOGICAL SYSTEMS

- Biological processes are complex
  - signaling, regulation, defense, ....
- Huge dynamics timescales: microseconds to years
- Spatial scales over 12 orders of magnitude
  - single protein to cell to organ to organism ...
- Oceans of experimental biological data generated
- Important intuitions captured in mental models that biologists build of biological processes
- How to build in silico models from all this?



# SYMBOLIC SYSTEMS BIOLOGY

- Symbolic -- represented in a **logical framework**
- Systems -- how things interact and work together, integration of multiple parts, viewpoints and levels of abstraction
- Goals:
  - Develop formal models that are as close as possible to domain expert's mental models
  - Compute with, analyze and reason about complex networks
  - New insights into / understanding of biological mechanisms



# EXECUTABLE FORMAL MODELS

- Describe system states and rules for change in a formal system
- From an initial state, derive a transition graph
  - nodes -- reachable states
  - edges -- rules connecting states
- Path in transition graph ~ computation/derivation
- Many kinds of analysis available



# PATHWAY LOGIC (PL) REPRESENTATION OF SIGNALING

<http://pl.csl.sri.com/>



# SIGNALING PATHWAYS

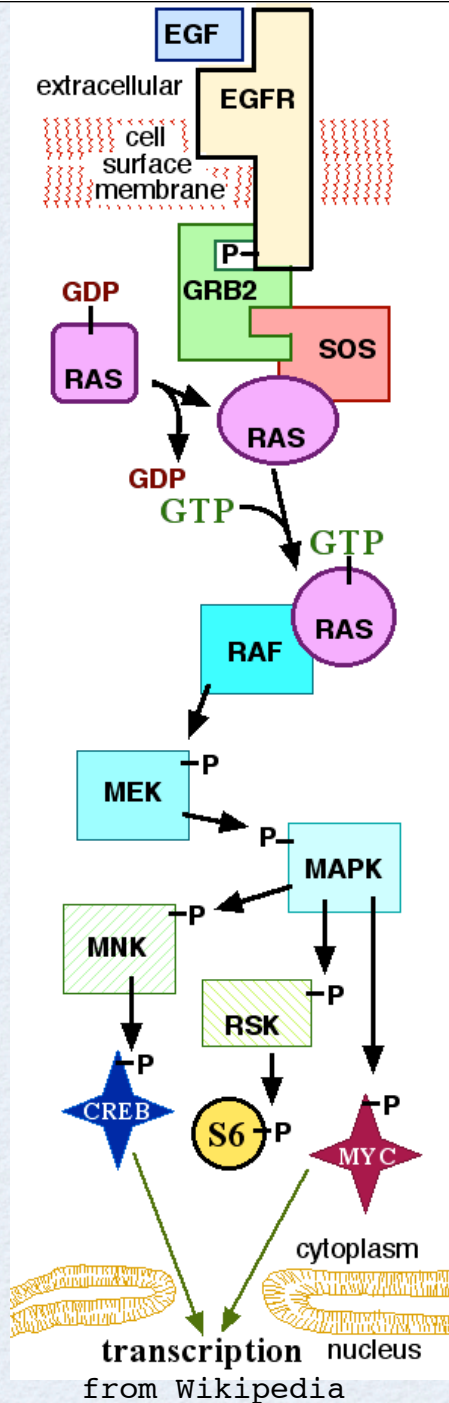
- Signaling pathways involve the modification and/or assembly of proteins and other molecules within cellular compartments into complexes that coordinate and regulate the flow of information.
- Signaling pathways are distributed in networks having stimulatory (positive) and inhibitory (negative) feedback loops, and other concurrent interactions to ensure that signals are propagated and interpreted appropriately in a particular cell or tissue.
- Signaling networks are robust and adaptive, in part because of combinatorial complex formation (several building blocks for forming the same type of complex), redundant pathways, and feedback loops.



## Egf stimulation of the Mitogen Activated Protein Kinase (MAPK) pathway.

Egf → EgfR → Grb2 → Sos1 → Ras → Raf1 → Mek → Erk

- Egf (EGF) binds to the Egf receptor (EgfR) and stimulates its protein tyrosine kinase activity to cause autophosphorylation, thus activating EgfR.
- The adaptor protein Grb2 (GRB2) and the guanine nucleotide exchange factor Sos1 (SOS) are recruited to the membrane, binding to EgfR.
- The EgfR complex activates a Ras family GTPase
- Activated Ras activates Raf1, a member of the RAF serine/threonine protein kinase family.
- Raf1 activates the protein kinase Mek (MEK), which then activates Erk (MAPK)
- ...





# REWRITING LOGIC

- Rewriting Logic is a logical formalism that is based on two simple ideas
  - states of a system are represented as elements of an algebraic data type
  - the behavior of a system is given by local transitions between states described by rewrite rules
- It is a logic for executable specification and analysis of software systems, that may be concurrent, distributed, or even mobile.
- It is also a (meta) logic for specifying and reasoning about formal systems, including itself (reflection!)



# ABOUT PATHWAY LOGIC

Pathway Logic (PL) is an approach to modeling biological processes as executable formal specifications (in Maude)

The resulting models can be queried

- using formal methods tools: given an initial state
  - execute --- find some pathway
  - search --- find all reachable states satisfying a given property
  - model-check --- find a pathway satisfying a temporal formula
- using reflection
  - find all rules that use / produce X (for example, activated Rac)
  - find rules down stream of a given rule or component
  - translate to alternative formalism and export



# PATHWAY LOGIC ORGANIZATION

A PL cell signaling model is generated from

- a knowledge base
- a cell state

A PL knowledge base consists of

- Theops --- sorts and operations
- Components --- specific proteins, chemicals ...
- Rules --- biomolecular reactions / processes

A cell state is given by specifying

- the proteins and other molecular components present
- the incoming signals (ligands)



# RULE 1: RECEPTOR BINDING

If a dish contains an EgfR ligand (?ErbB1L:ErbB1L) outside a cell with EgfR in the cell membrane then the ligand binds to exterior part of the receptor and the receptor is activated.

```
rl[1.EgfR.act]:  
  ?ErbB1L:ErbB1L  [CellType:CellType | ct {CLm | clm EgfR}]  
  =>  
  [CellType:CellType | ct {CLm | clm ([EgfR - act] : ?ErbB1L:ErbB1L)} ] .
```

Rule 1 applies to rasDish

```
PD(Egf [Cell | {CLm | EgfR PIP2}{CLi | [Hras - GDP] Src}  
    {CLc | Gab1 Grb2 Pi3k Plcg Sos1}])
```

with the match

```
?ErbB1L:ErbB1L := Egf  
clm := PIP2  
ct := {CLi | [Hras - GDP] Src} {CLc | Gab1 Grb2 Pi3k Plcg Sos1}
```

giving rasDish1

```
PD([Cell |  
    {CLm | ([EgfR - act] : Egf) PIP2}  
    {CLi | [Hras - GDP] Src}  
    {CLc | Gab1 Grb2 Pi3k Plcg Sos1}]) .
```



# HOW DO WE INFER RULES?

Rules are inferred from evidence, called datums, curated from the literature.

Consider the rule converting Rala-GDP to Rala-GTP in response to Egf stimulation

```
rl[1064.Rala.irt.Egf]:  
{EgfRC | [EgFR - act] : Egf Pi3k RalGds clm }  
{CLi | cli [Rala - GDP]}  
=>  
{EgfRC | [EgFR - act] : Egf Pi3k RalGds [Rala - GTP] clm }  
{CLi | cli } .
```

Evidence for this rule includes:

```
DID#05387: Rala[Ab] GTP[BD-PD] is increased irt Egf (tnr)  
cells: Cos7 in BMLS  
inhibited by: Wortmannin [chem] --- Pi3k Inhibitor  
inhibited by: LY294002 [chem] ""  
source: 15034142-Fig-5c
```

```
DID#12876: xRala[xAb]IP GTP/GDP[32Pi-TLC] is increased irt Egf  
cells: Cos1-xRalGds in BMS  
times: 0 1+ 2++ 3++ 4+ 5 min  
reqs: xRalGds [omission]  
inhibited by: xRalGds-C203S "membrane-binding-mutant" [substitution]  
comment: cells were pretreated with Vanadate 30 min before Egf treatment  
source: 9416833-Fig-2
```



THE PATHWAY LOGIC ASSISTANT  
(PLA)



# THE PATHWAY LOGIC ASSISTANT (PLA)

- Provides a means to interact with a PL model
- Manages multiple representations
  - Maude module (logical representation)
  - PetriNet (process representation for efficient query)
  - Graph (for interactive visualization)
- Exports Representations to other tools
  - Lola (and SAL model checkers)
  - Dot -- graph layout
  - JLambda (interactive visualization, Java side)

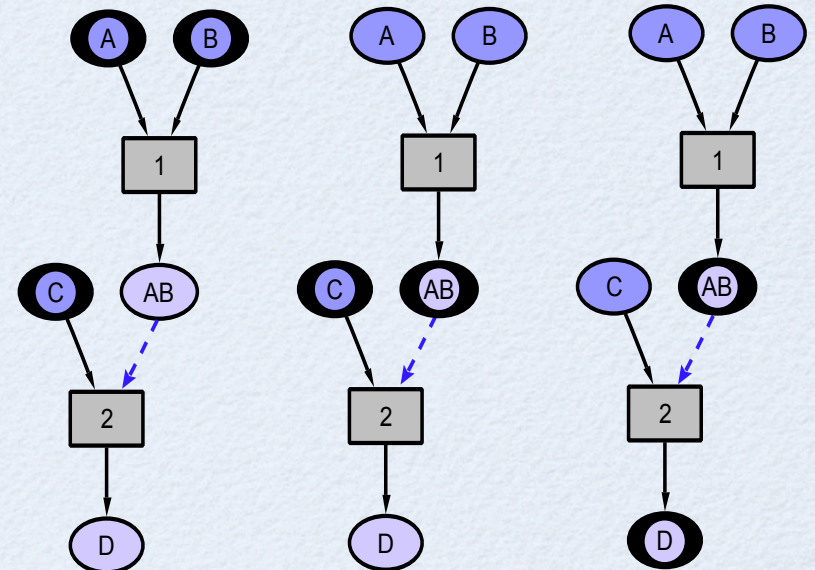


# ABOUT PETRI NETS

A Petri net is represented as a graph with two kinds of nodes:

- \* transitions/rules (reactions--squares)
- \* places/occurrences (reactants, products, modifiers--ovals)

A Petri net process has tokens on some of its places. A rule can fire if all of its inputs have tokens. Firing a rule moves tokens from input to output.



An execution is a sequence of rule firings.

A pathway is represented as an execution subgraph.



# rasNet

Rule instances relevant to Hras activation

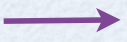
Hras activated



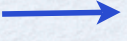
Parallel paths



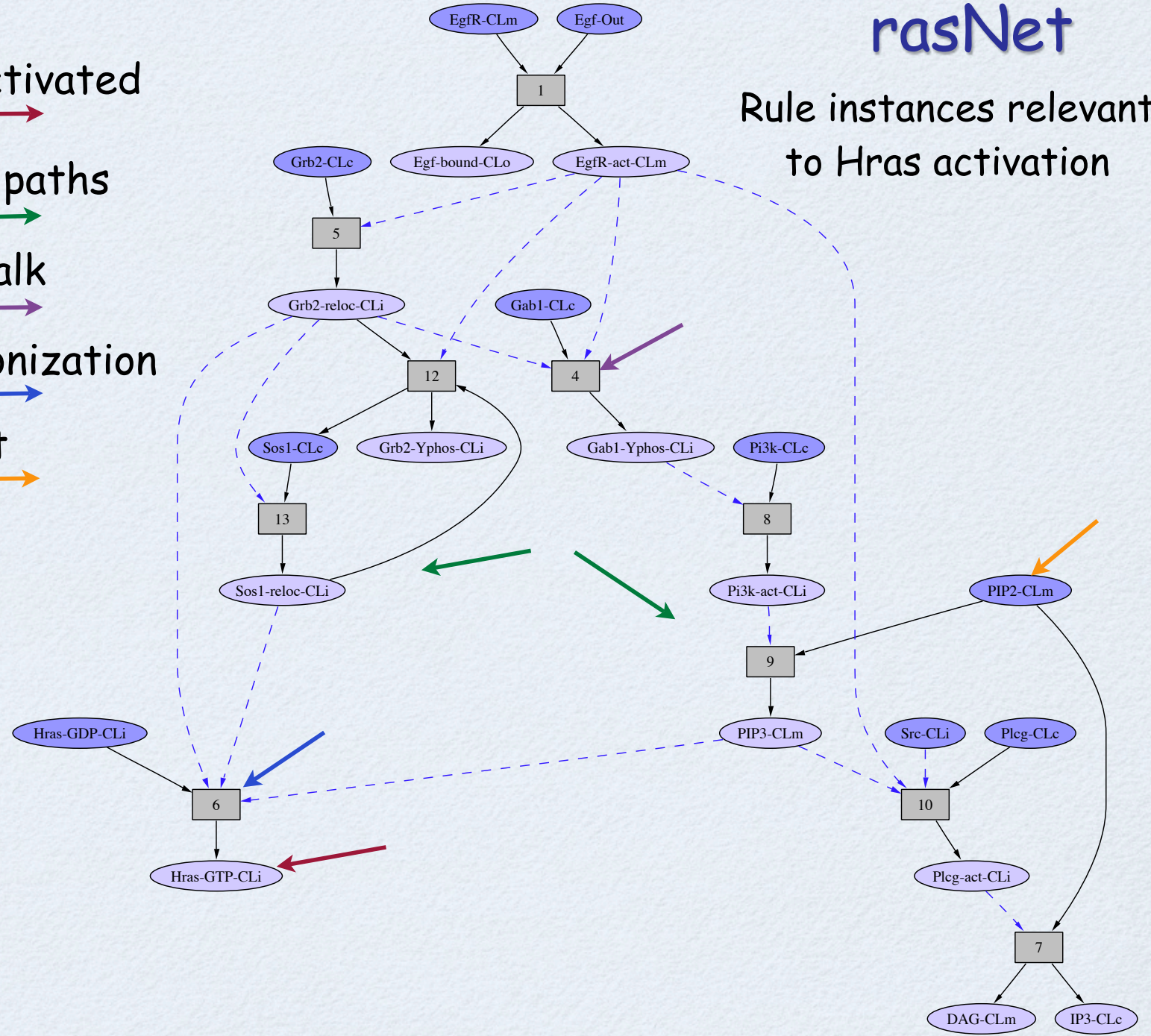
Cross talk



Synchronization



Conflict



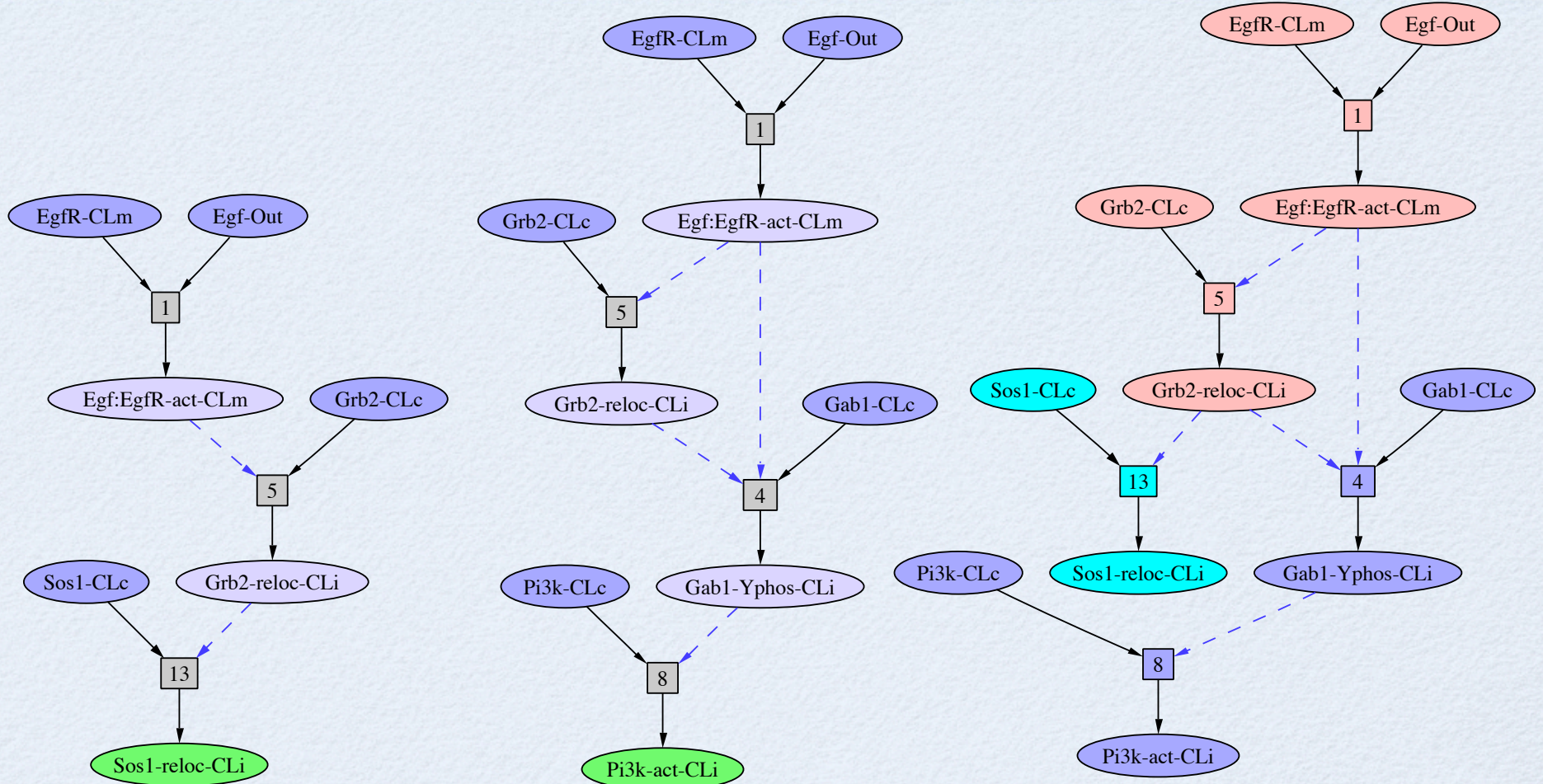


# A SIMPLE QUERY LANGUAGE

- Given a Petri net with transitions  $P$  and initial marking  $O$  (for occurrences) there are two types of query
  - subnet
  - findPath - a computation / unfolding
- For each type there are three parameters
  - $G$ : a goal set---occurrences required to be present at the end of a path
  - $A$ : an avoid set---occurrences that must not appear in any transition fired
  - $H$ : as list of identifiers of transitions that must not be fired
- subnet returns a subnet containing all (minimal) such pathways (using backward and forward collection)
- findPath returns a pathway (transition list) generating a computation satisfying the requirements (using model checking on the negation).



# PATHWAYS IN RAS NET



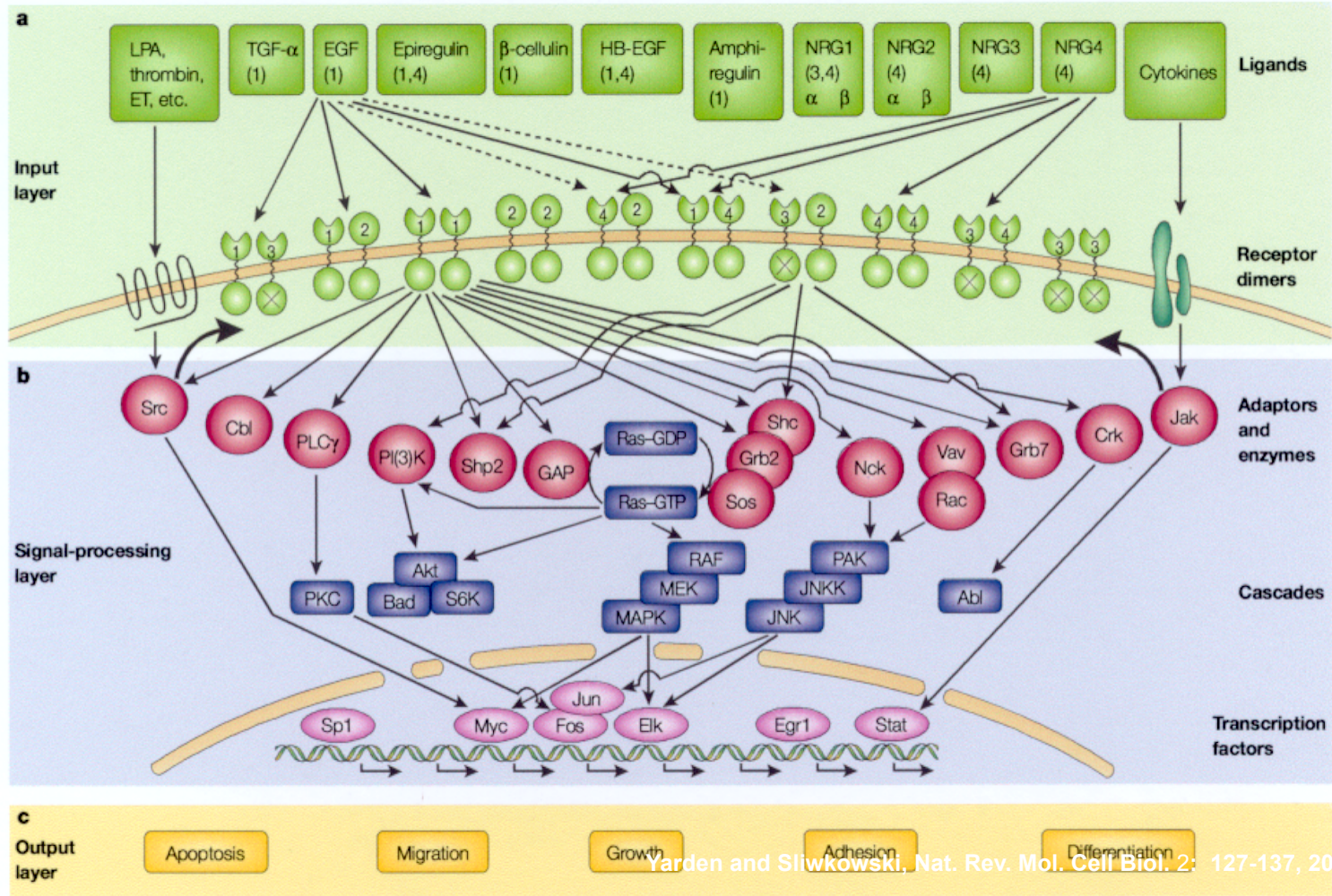


# MODEL OF EGF STIMULATION

(by Merrill Knapp)

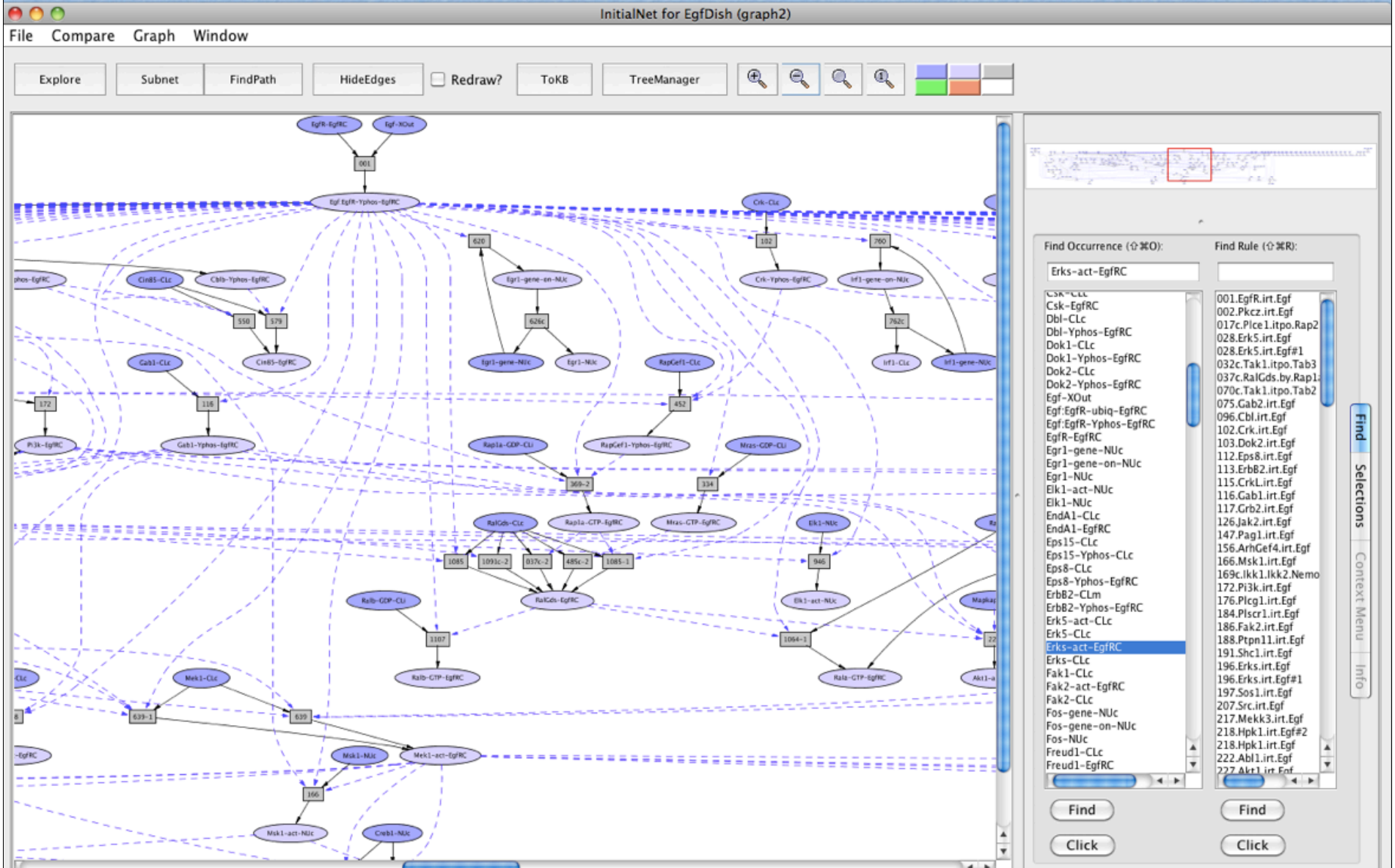


# THE ERBB NETWORK (CARTOON FORM)





# PL EGF MODEL

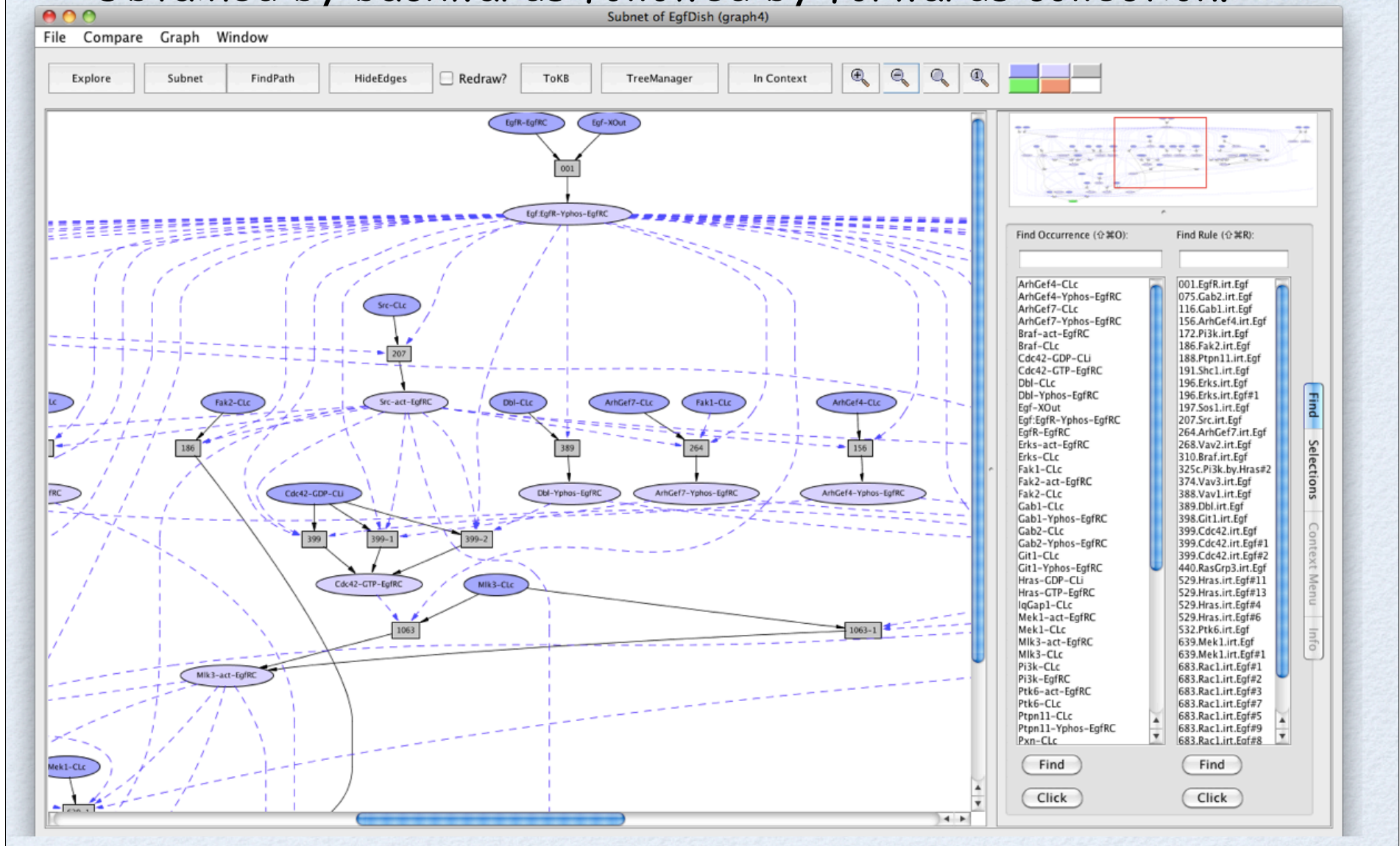




# SUBNET RELEVANT TO ERK ACTIVATION

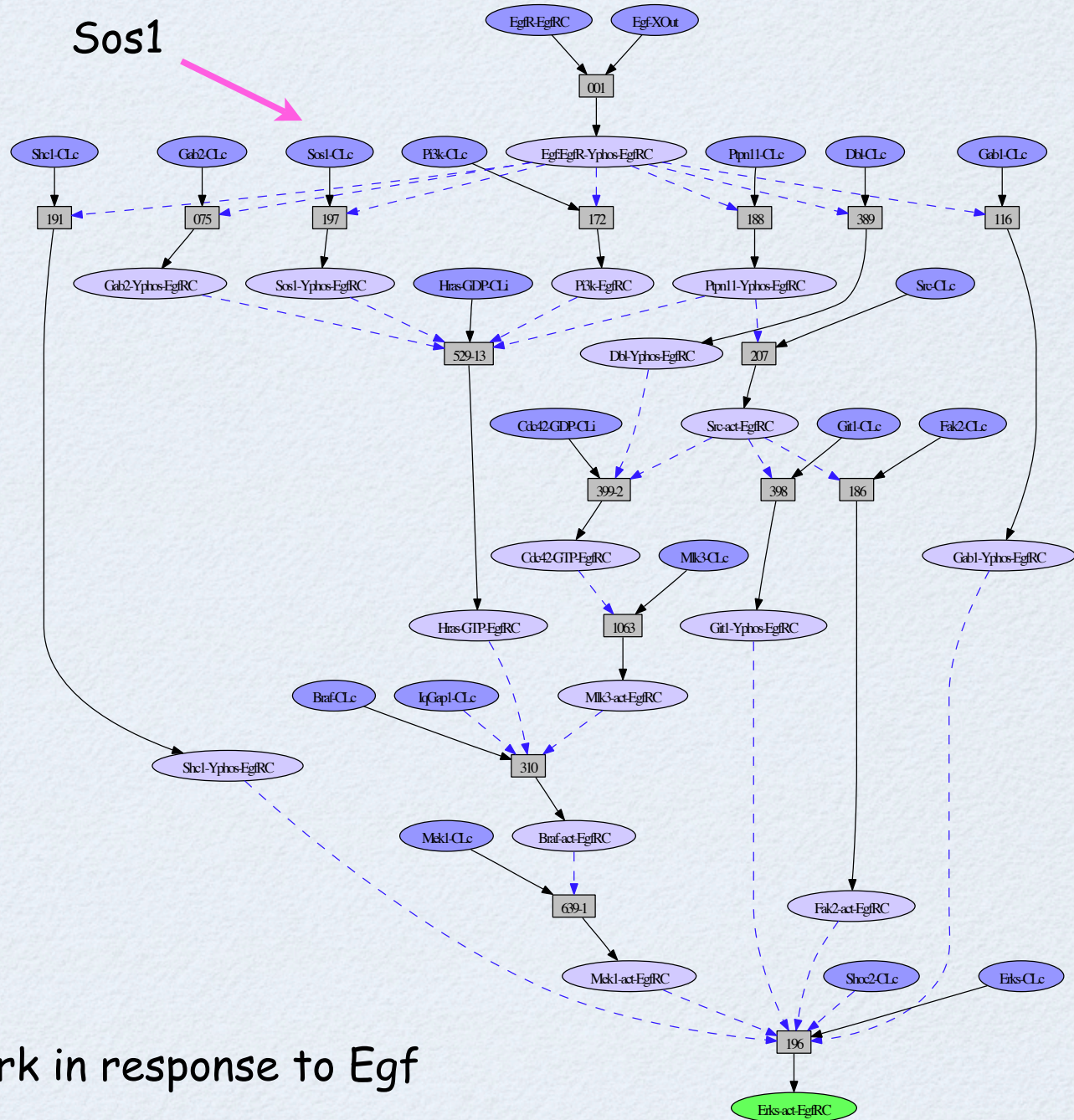
Contains all pathways leading to activation of Erk.

Obtained by backwards followed by forwards collection.



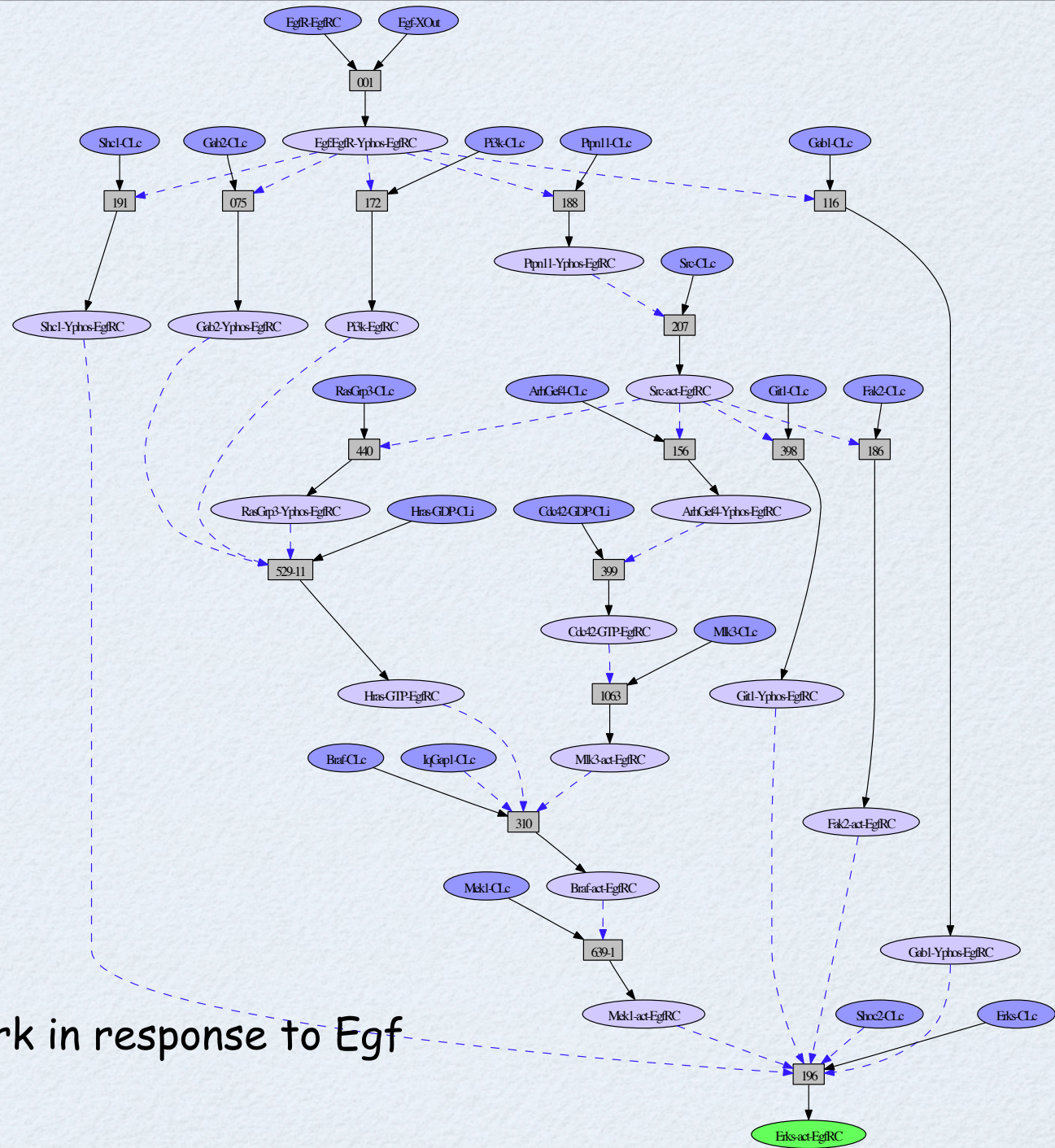


Sos1



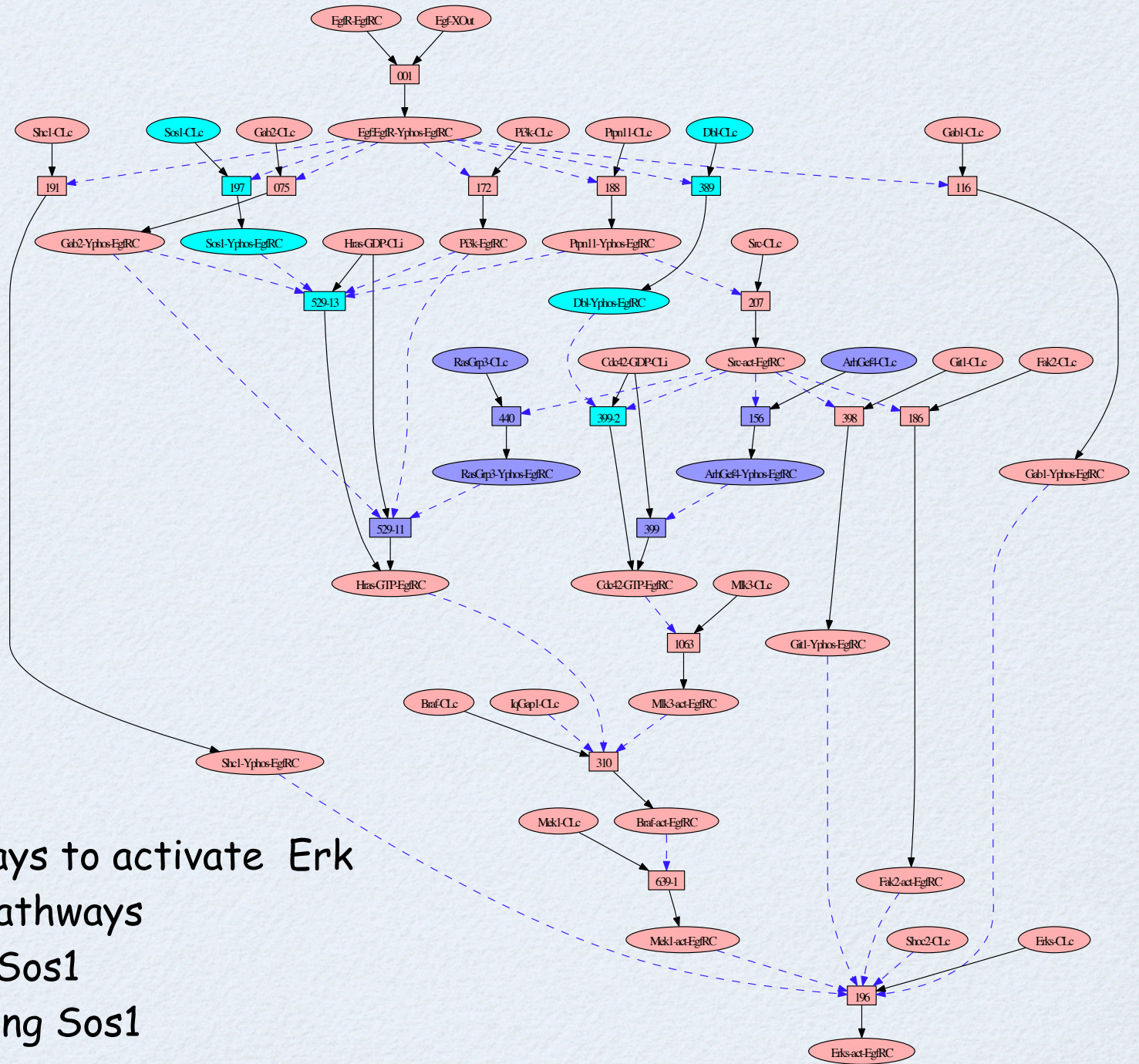
Activation of Erk in response to Egf





Activation of Erk in response to Egf  
avoiding Sos1





Comparing ways to activate Erk  
 Pink - both pathways  
 Cyan -- with Sos1  
 Blue -- avoiding Sos1





SLEEP

(with MaryAnn Greco and Merrill Knapp)



# THE QUESTION

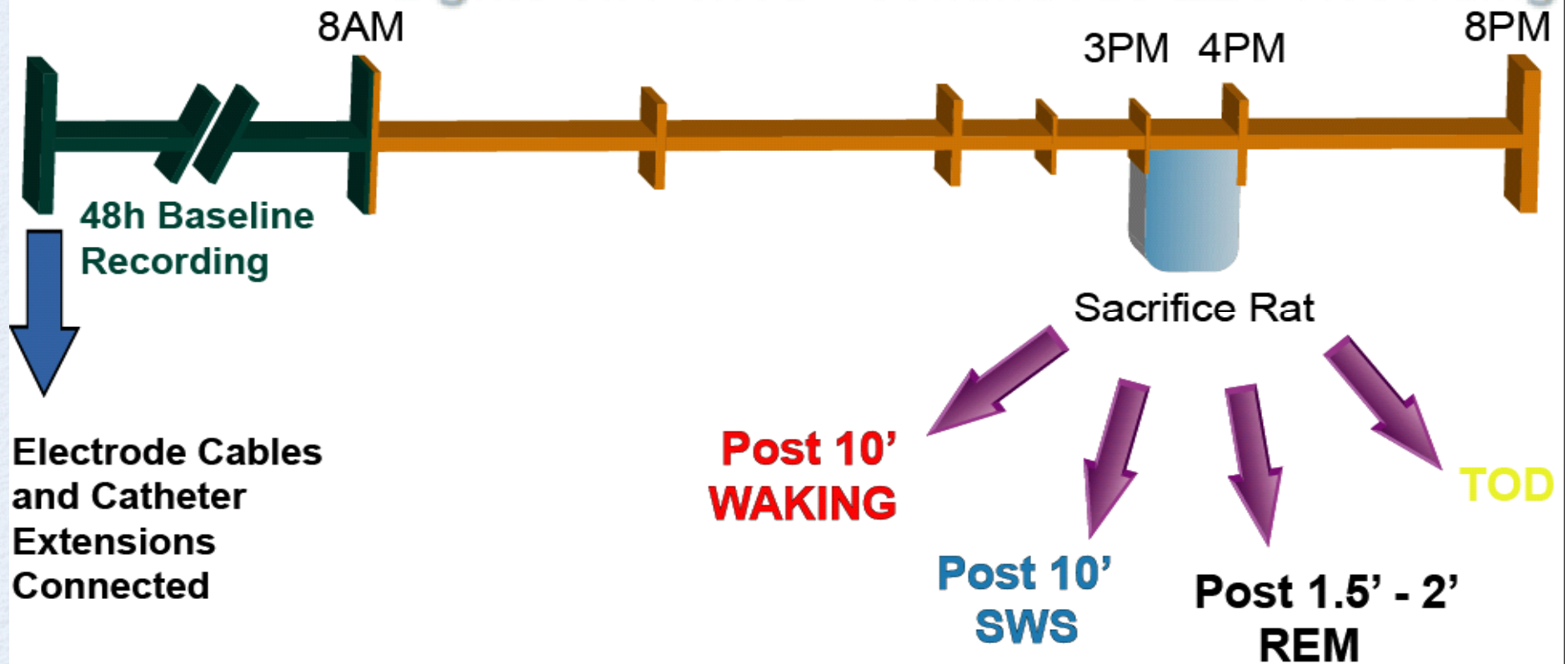
- What is the function of sleep?
- What are your cells doing when you sleep? vs awake?
- Rat model -- proteomics from different organs at different sleep states





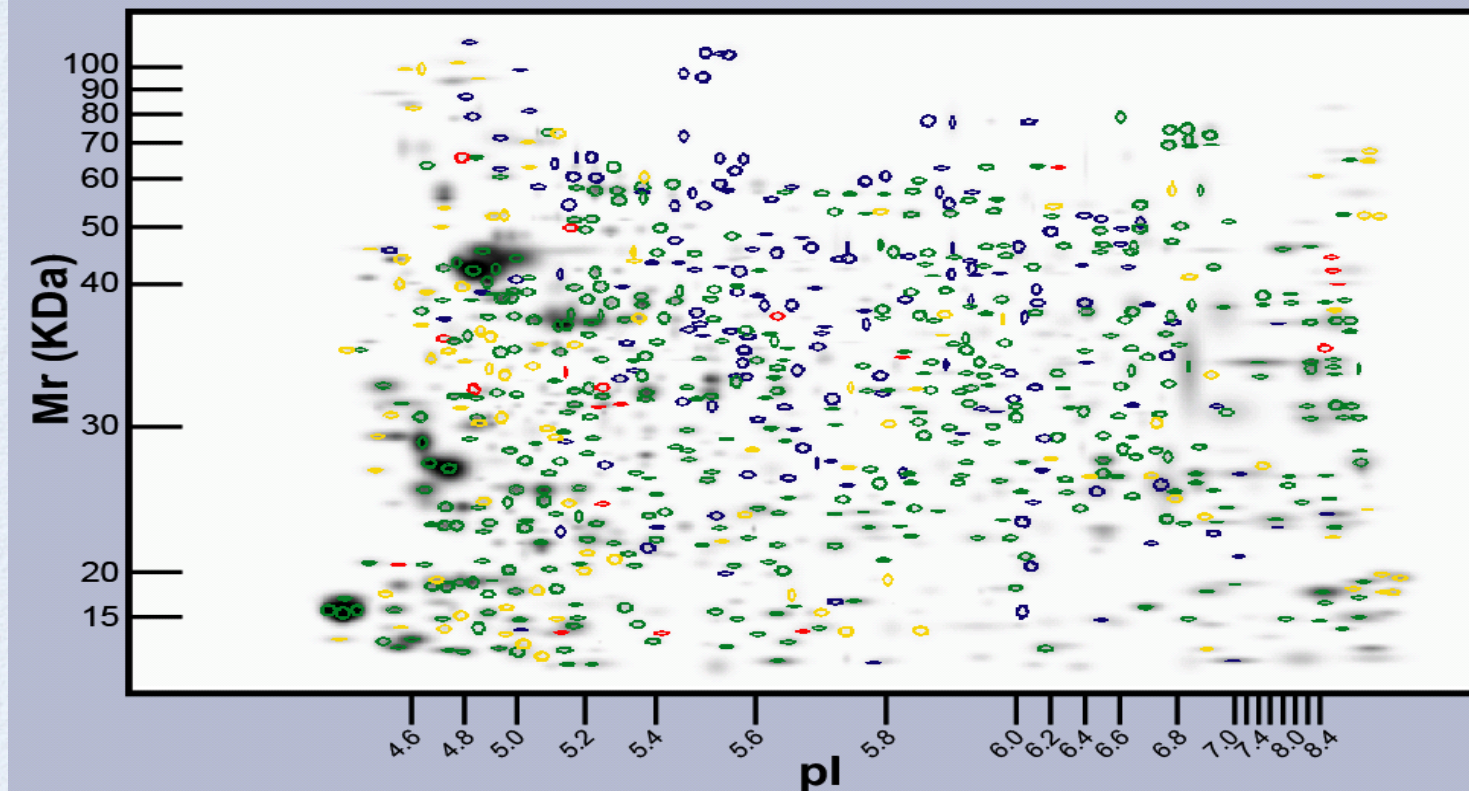
# NATURAL SLEEP PARADIGM

**Lights on Period - Continuous EEG Recording**





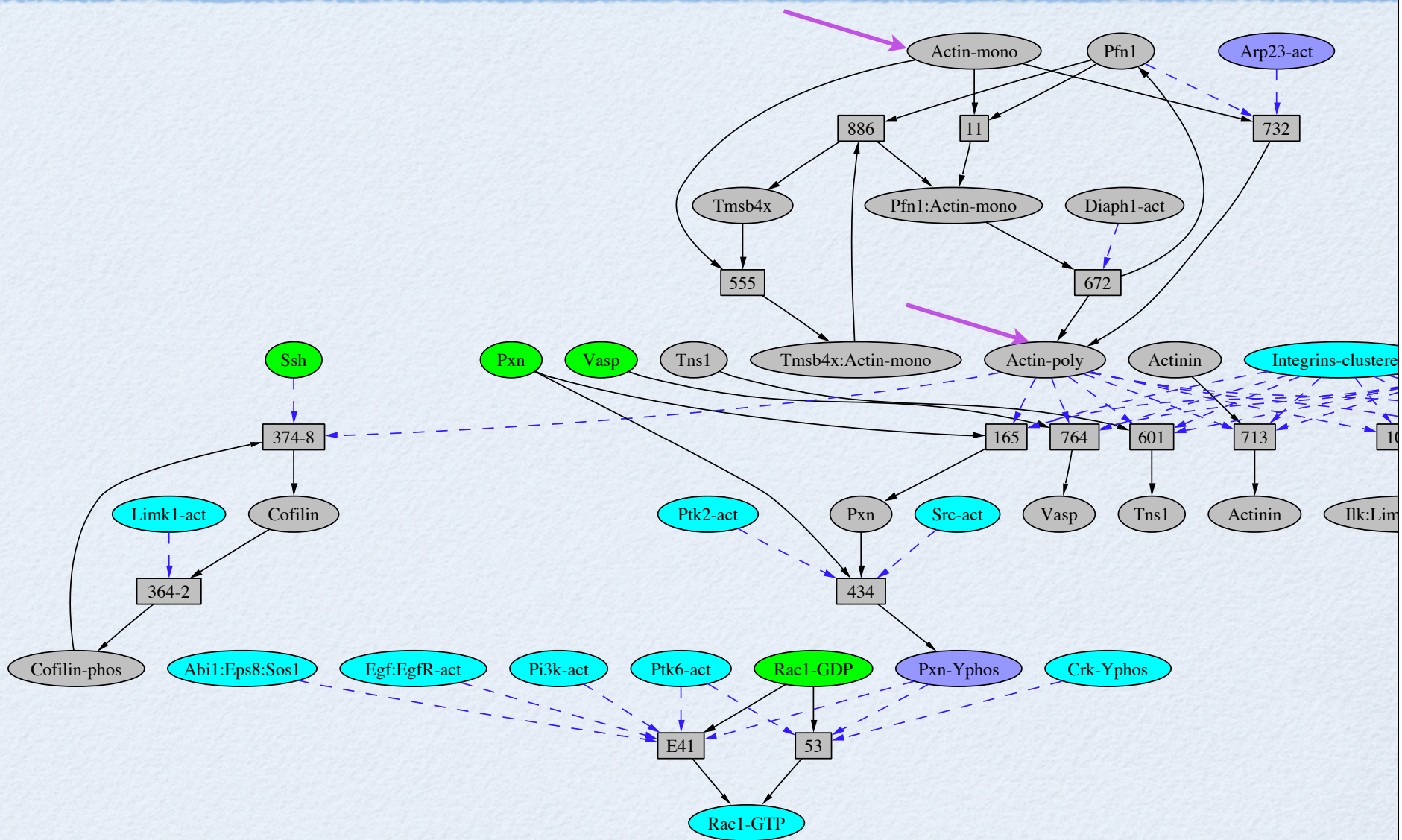
# 2D MASTER GEL



Proteins unique to different states were identified  
Those modeled in PL included Actin and Rhob  
Use the PLA explorer to find signaling connections

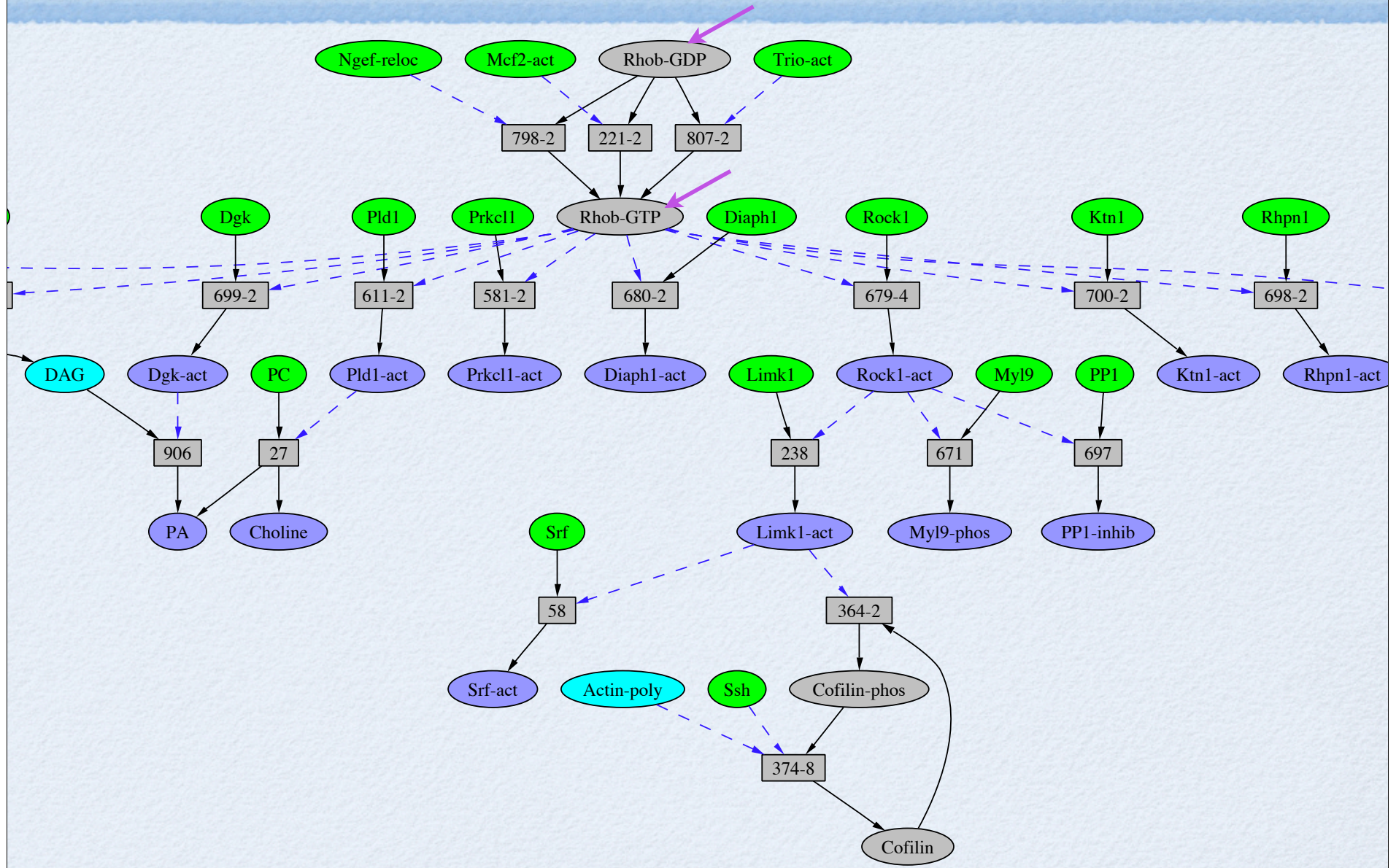


# EXPLORING PLKB FROM ACTIN



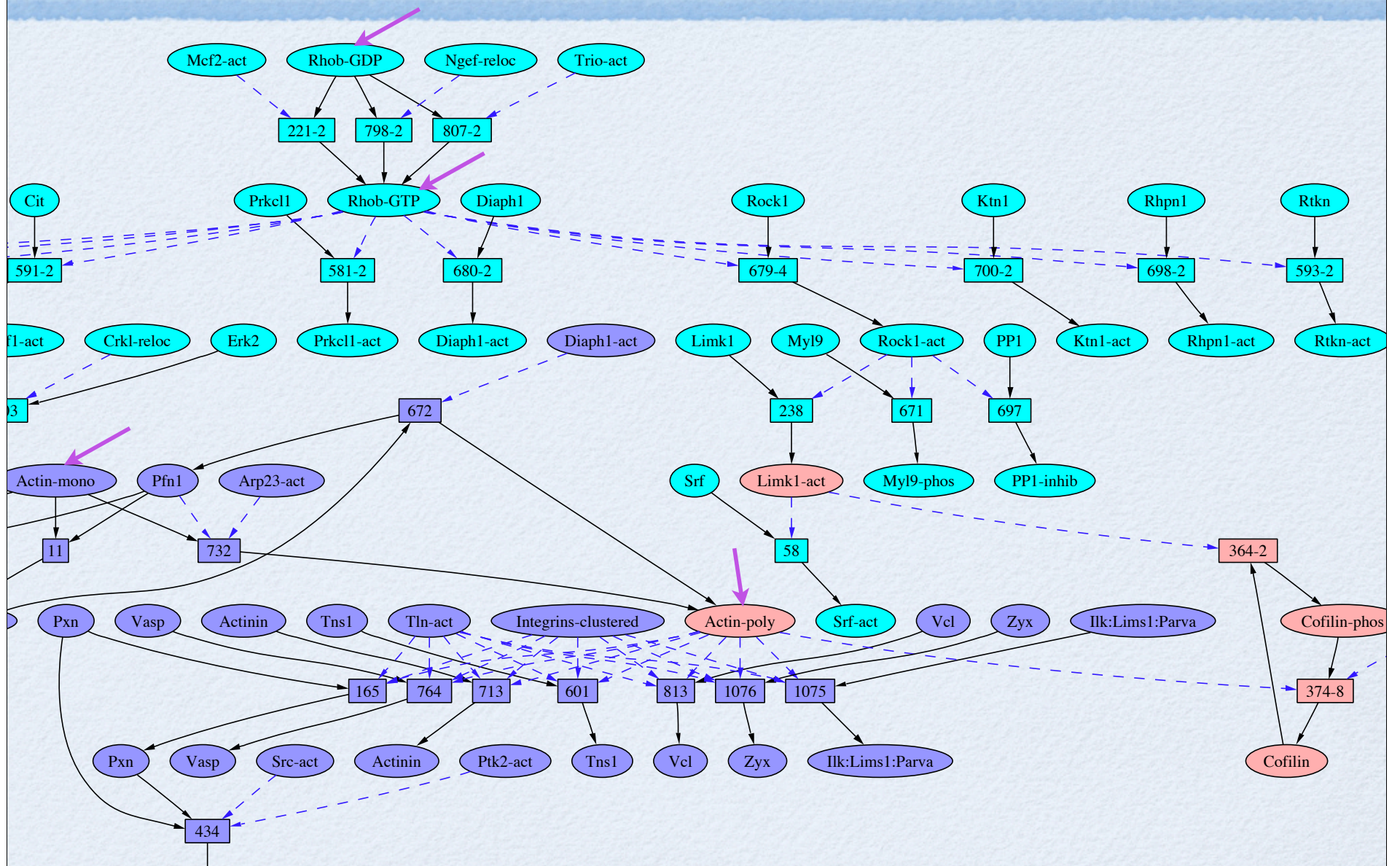


# EXPLORING PLKB FROM RHOB



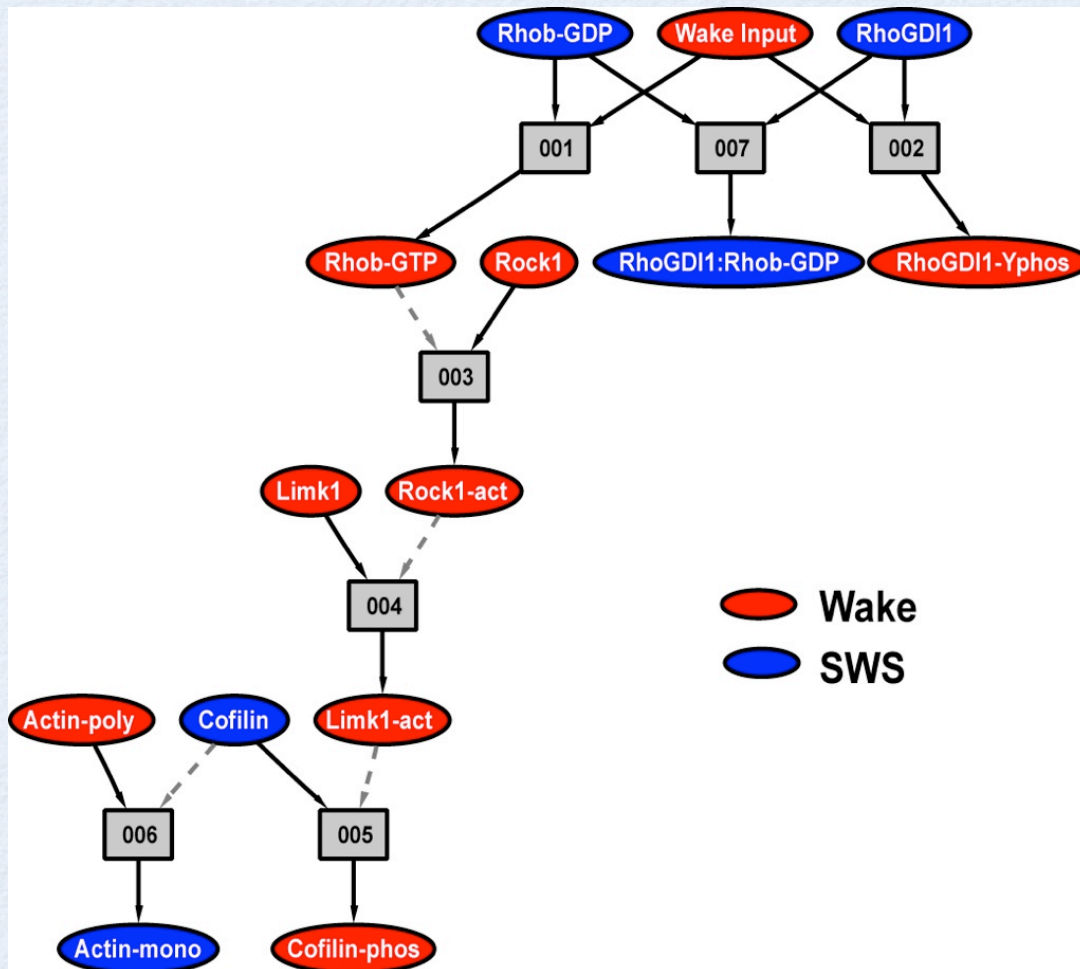


# COMPARING THE EXPLORE NETS





# A HYPOTHETICAL MODEL PATHWAY RELATING STATE AND SYNAPTACTIC PLASTICITY



Wake state:

unknown signal(s)

=> phosphorylation of Rock1

=> activation of Limk1

=> phosphorylation of cofilin

=> increase in polymerized actin

(Phosphorylated cofilin is unable to depolymerize actin)

SWS:

RhoGDI1 binds RhoGDP

(is not phosphorylated)

=> Rock1, Limk1, and cofilin would

not be phosphorylated and

=> actin depolymerization

=> decrease in synaptic weight

TODO -- test the hypothesis





MODELING METABOLISM  
(work of Malabika Sarker)



# THE PROBLEM

- Identify candidate drug targets in mycobacteria
- Idea: integrate screening data, molecular structure models, and metabolic models (using symbolic system biology!)
- Initial steps
  - curation of PL model of mycolic acid synthesis (including drug action)
  - importing PGDBs into PL

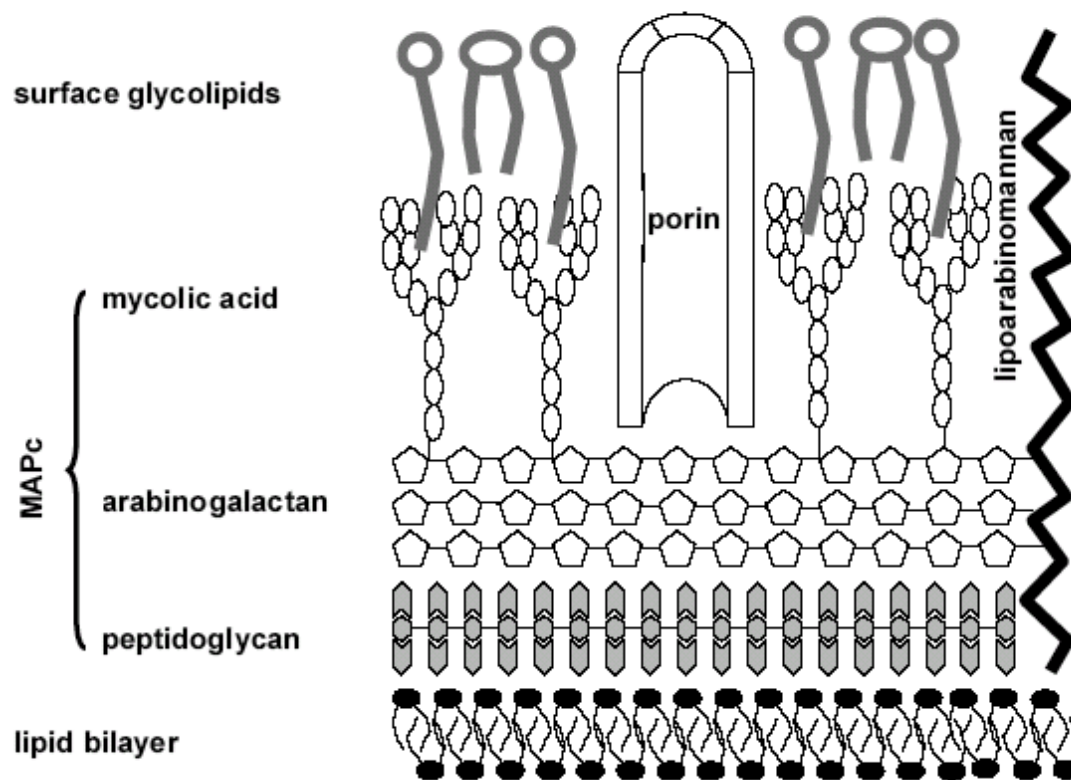


# WHY MYCOLIC ACID

Mycolate biosynthesis enzymes are essential for survival of Mycobacteria---excellent drug targets

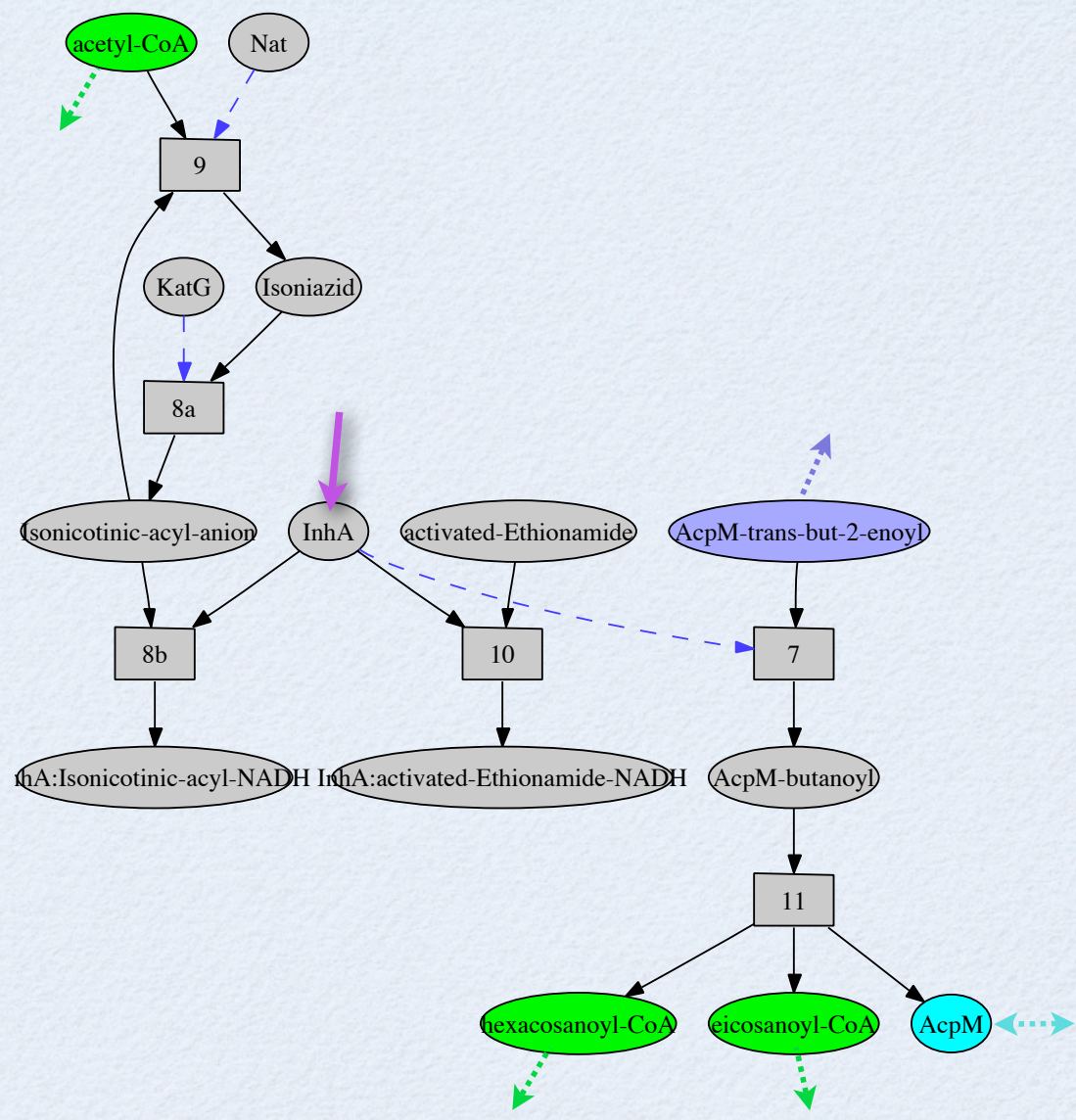
Isoniazid(INH)/Ethionamide(ETH)/Triclosan(TRC) --| InhA

## Mycobacterial Cell Wall





# MYCOLIC ACID FRAGMENT SHOWING INHIBITION OF INHA







# CHOOSING A FORMALISM



# REQUIREMENTS

- What do you want to represent?
- What questions do you want to ask?
- What material/data is available, what must be estimated/hypothesized?
- How are you going to validate?
- What are you familiar with?
- What formalism features:
  - Language primitives, semantic models, tools
  - model analysis/validation vs system analysis



# WHAT TO MODEL?

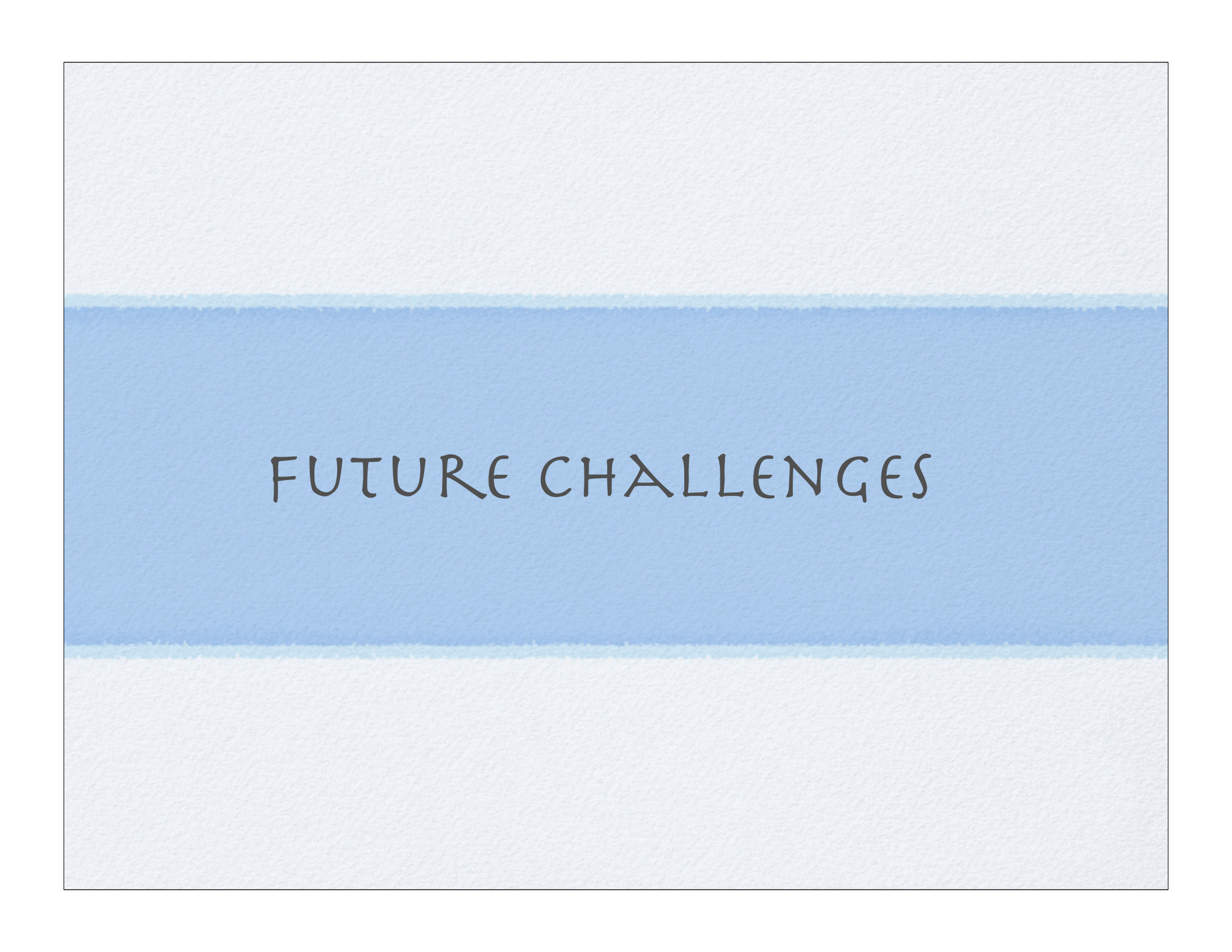
- Species: Individual, concentration, population
- Structure: compartments, binding, location
- Process:
  - non-determinism -- computation trees
  - stochastic / probabilistic -- MC
  - deterministic -- ODE
- Quantitative: kinetics, dynamics,
- Qualitative: interactions, causal relations ...
- Effects of perturbation
  - Knockout / KnockIn / Mutations
  - Stimulus / Stress



# WHAT QUESTIONS?

- Examples:
  - All the ways to phosphorylate Erk?
  - How fast can signal get to nucleus?
  - Is Pi3k required for activation of Raf?
- Categories:
  - Reachability
  - Information/material flow
  - Dynamical features
    - steady states, stable states, oscillations
    - competition/race conditions/interference
  - Kinetic profiles





# FUTURE CHALLENGES



# FUTURE CHALLENGES BIOLOGY RELATED

- Integration of signaling and metabolic networks
- Host response
  - integrating host and pathogen
- Integrating models from different sources
  - different levels of detail
    - phos, Yphos, phos(Y 234), ...
  - different representation choices



# FUTURE CHALLENGES ALGORITHMS

- Adding semiquantitative information
- Integrate probabilistic / stochastic reasoning
- Algorithms to discover meaningful subnets
- Finding all pathways
- Visualizing 10s/100s of pathways



# FUTURE CHALLENGES MAGIC

- Automating curation
  - from text
  - from figures
- Inferring rules
  - what changes (one step)
  - in what



???