About Pathway Logic

http://pl.csl.sri.com

Carolyn Talcott
SRI International
January 9, 2008

PLAN

- Symbolic systems biology
- Pathway Logic Basics
- The Pathway Logic Assistant (PLA)
- Applications

SYMBOLIC SYSTEMS BIOLOGY

Symbolic systems biology is the qualitative and quantitative study of biological processes as integrated systems rather than as isolated parts.

Our initial goal for symbolic systems biology include:

- modeling causal networks of biomolecular interactions in a logical framework at multiple scales
- developing executable formal models that are as close as possible to domain expert's (biologists) mental models
- being able to compute with and analyze these complex networks
  - abstracting and refining the logical models
  - using simulation/deduction to compute/check postulated properties
  - making testable predictions about possible outcomes
  - using experimental results to update the models

Sampling of Symbolic/Executable Modeling Approaches

- Rewriting Logic + Temporal Logic (Pathway Logic!)
- Chemical Abstract Machine + Computation Tree Logic (BIOCHAM)
- Membrane calculi -- spatial process calculi / logics
  - BioAmbients / Brane calculus -- mobility of membranes
  - P Systems -- mobility of processes
- Statecharts + Live Sequence Charts
- Process Algebras + Stochastic Simulation & Probabilistic Model Checking
- Hybrid SAL -- hybrid (discrete + continuous) systems

PATHWAY LOGIC BASICS

About Rewriting Logic

Rewriting Logic is

- an extension of equational logic with local rewrite rules to model change over time (or logical deduction)
- a logic for executable specification and analysis of software systems, that may be concurrent, distributed, or even mobile.
- a (meta) logic for specifying and reasoning about formal systems, including itself (reflection!)

REWRITING

- Rewrite theory: (Signature, Labels, Rules)
- Signature: (Sorts, Ops, Eqns) -- an equational theory
  - Describe data types, structure of system state
- Rules have the form label : t => t' if cond
- Rewriting operates modulo equations
  - rules apply locally
  - generates computations (pathways)

ABOUT PATHWAY LOGIC

Pathway Logic (PL) uses rewriting logic to model biological processes as executable formal specifications with pathways as computations. Models can be queried

- using formal methods tools:
  - 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 updown stream of a given rule or component
  - find the subnet relevant to a goal (desired state)

PATHWAY LOGIC ORGANIZATION

A PL cell signaling model is generated from

- a knowledge base
- a cell state

A PL knowledge base consists of

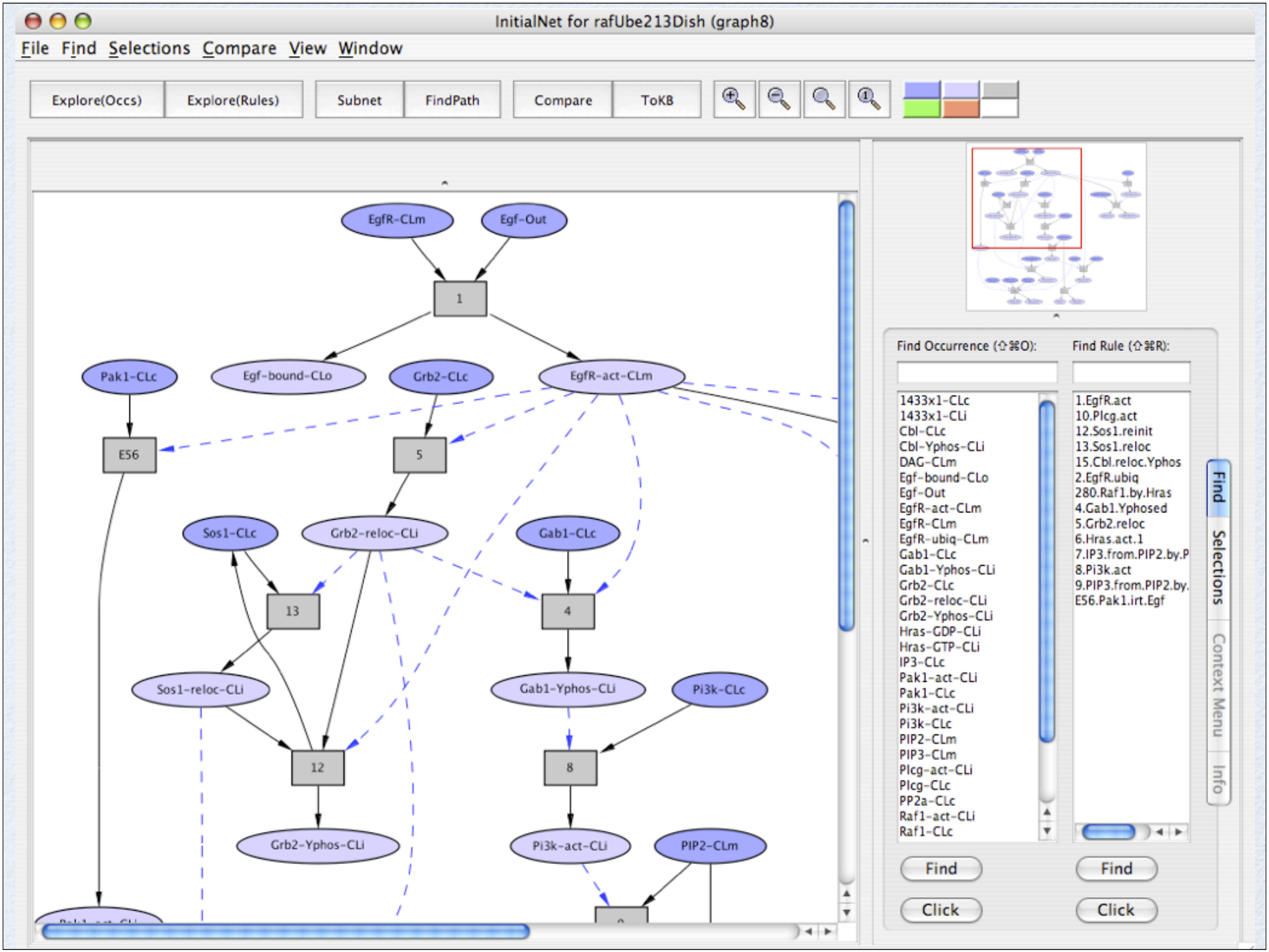
- Theops --- sorts and operations
- Components --- specific proteins, chemicals ...
- Rules --- biomolecular reactions / processes

A cell state is given by specifying

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

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)
  - SBML (xml based standard for model exchange)

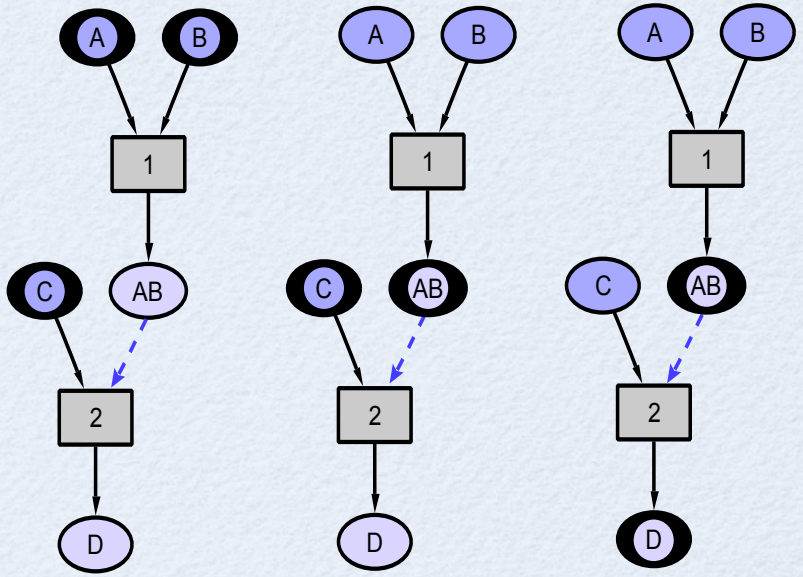

InitialNet for rafUbe213Dish (graph8)
File Find Selections Compare View Window
Explore(Occs)
Explore(Rules)
Subnet
FindPath
Compare
ToKB
EgfR-CLm
Egf-Out
1
Pak1-CLc
Egf-bound-CLo
Grb2-CLc
EgfR-act-CLm
E56
5
Sos1-CLc
Grb2-reloc-CLi
Gab1-CLc
13
4
Sos1-reloc-CLi
Gab1-Yphos-CLi
Pi3k-CLc
12
8
Grb2-Yphos-CLi
Pi3k-act-CLi
PIP2-CLm
Find Occurrence (⇧⌘O):
Find Rule (⇧⌘R):
1433x1-CLc
1433x1-CLi
Cbl-CLc
Cbl-Yphos-CLi
DAG-CLm
Egf-bound-CLo
Egf-Out
EgfR-act-CLm
EgfR-CLm
EgfR-ubiq-CLm
Gab1-CLc
Gab1-Yphos-CLi
Grb2-CLc
Grb2-reloc-CLi
Grb2-Yphos-CLi
Hras-GDP-CLi
Hras-GTP-CLi
IP3-CLc
Pak1-act-CLi
Pak1-CLc
Pi3k-act-CLi
Pi3k-CLc
PIP2-CLm
PIP3-CLm
Plcg-act-CLi
Plcg-CLc
PP2a-CLc
Raf1-act-CLi
Raf1-CLc
1.EgfR.act
10.Plcg.act
12.Sos1.reinit
13.Sos1.reloc
15.Cbl.reloc.Yphos
2.EgfR.ubiq
280.Raf1.by.Hras
4.Gab1.Yphosed
5.Grb2.reloc
6.Hras.act.1
7.IP3.from.PIP2.by.P
8.Pi3k.act
9.PIP3.from.PIP2.by.
E56.Pak1.irt.Egf
Find
Click
Find
Click
Find
Selections
Context Menu
Info

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.

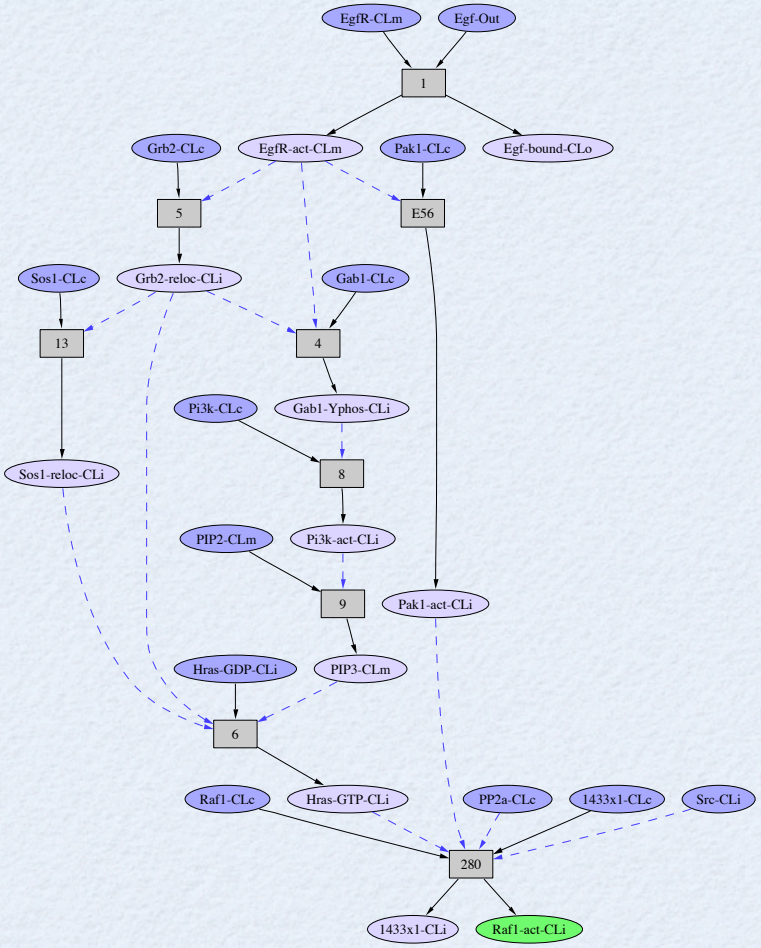


An execution is a sequence of rule firings.
A pathway is represented as an execution subgraph.

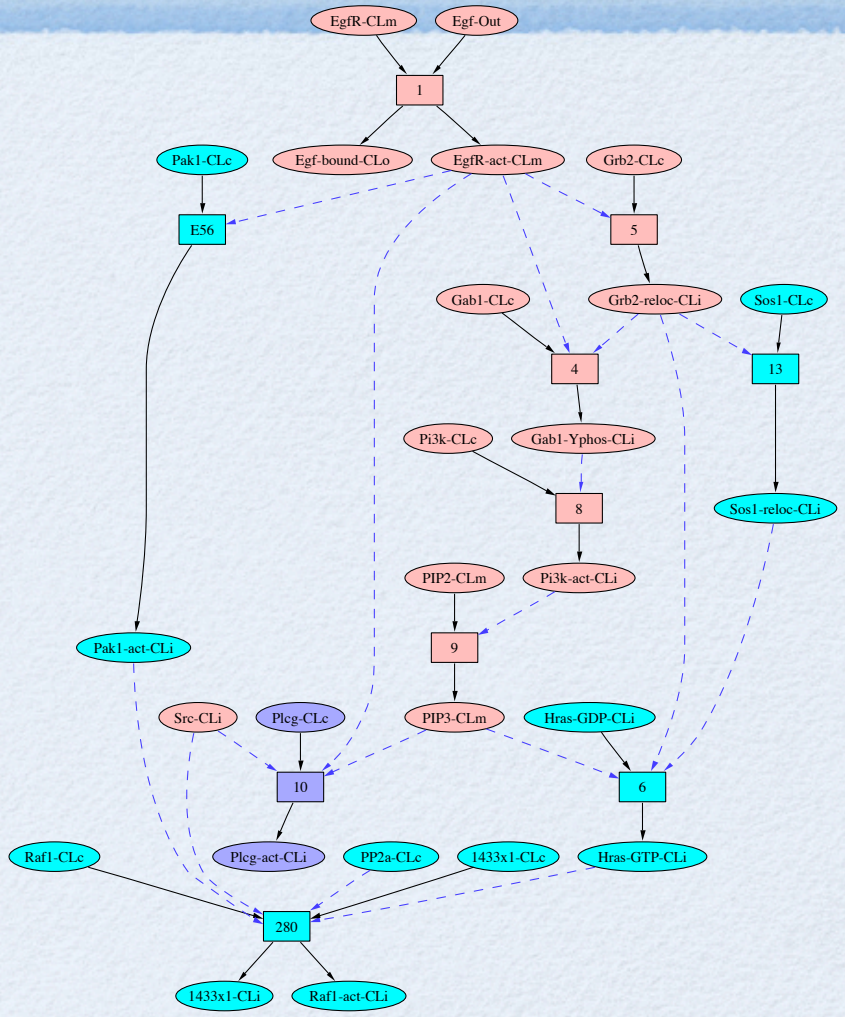
A SIMPLE QUERY LANGUAGE

- 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
- findPath returns a pathway (transition list) generating a computation satisfying the requiremments.
- subnet returns a subnet containing all (minimal) such pathways.

PATHWAY EXAMPLES



Activating Raf

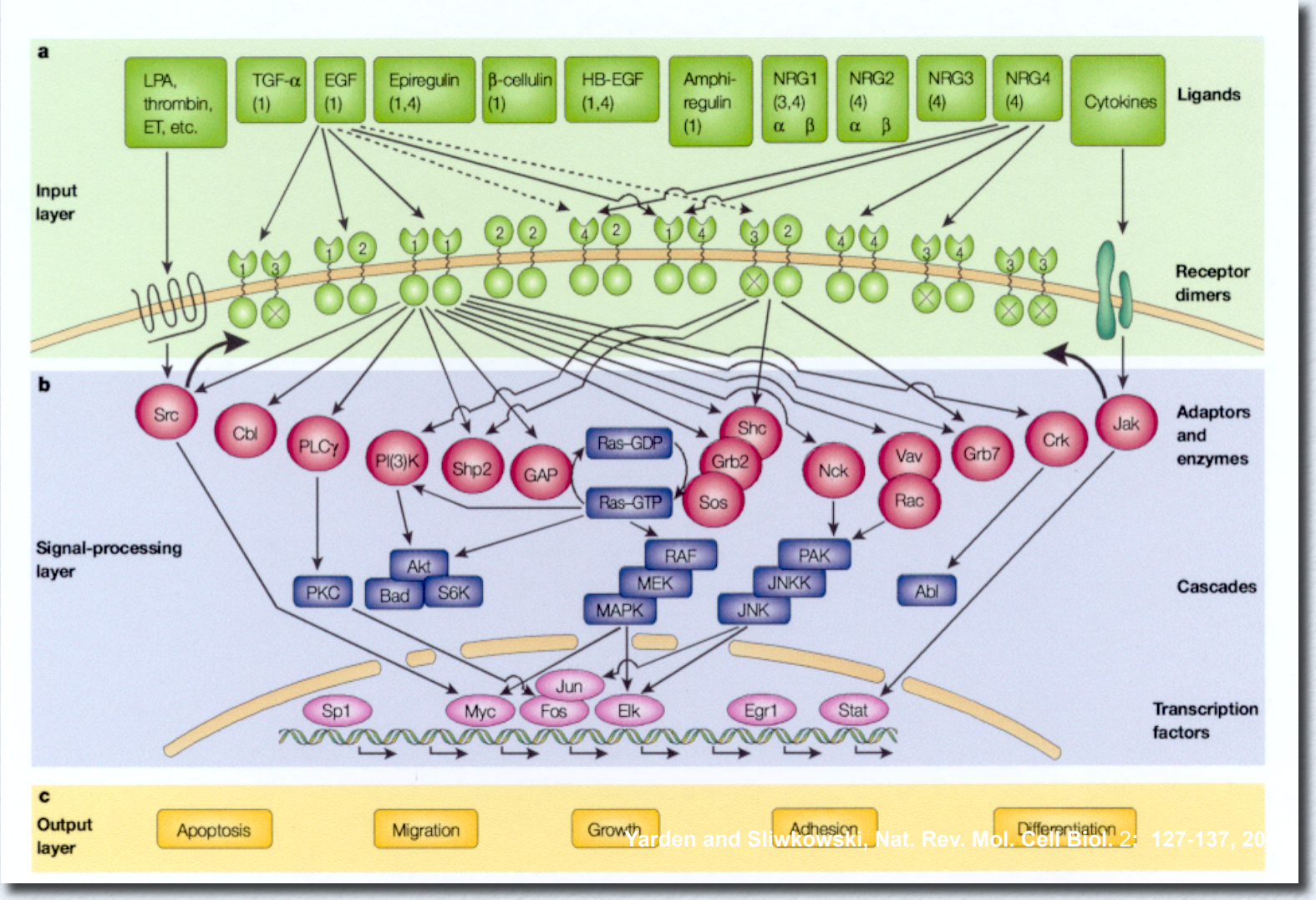


Comparing Raf and Plcg Pathways

Model of EGF stimulation

(by Merrill Knapp)

THE ERBB NETWORK (CARTOON FORM)



Yarden and Sliwkowski, Nat. Rev. Mol. Cell Biol. 2: 127-137, 20

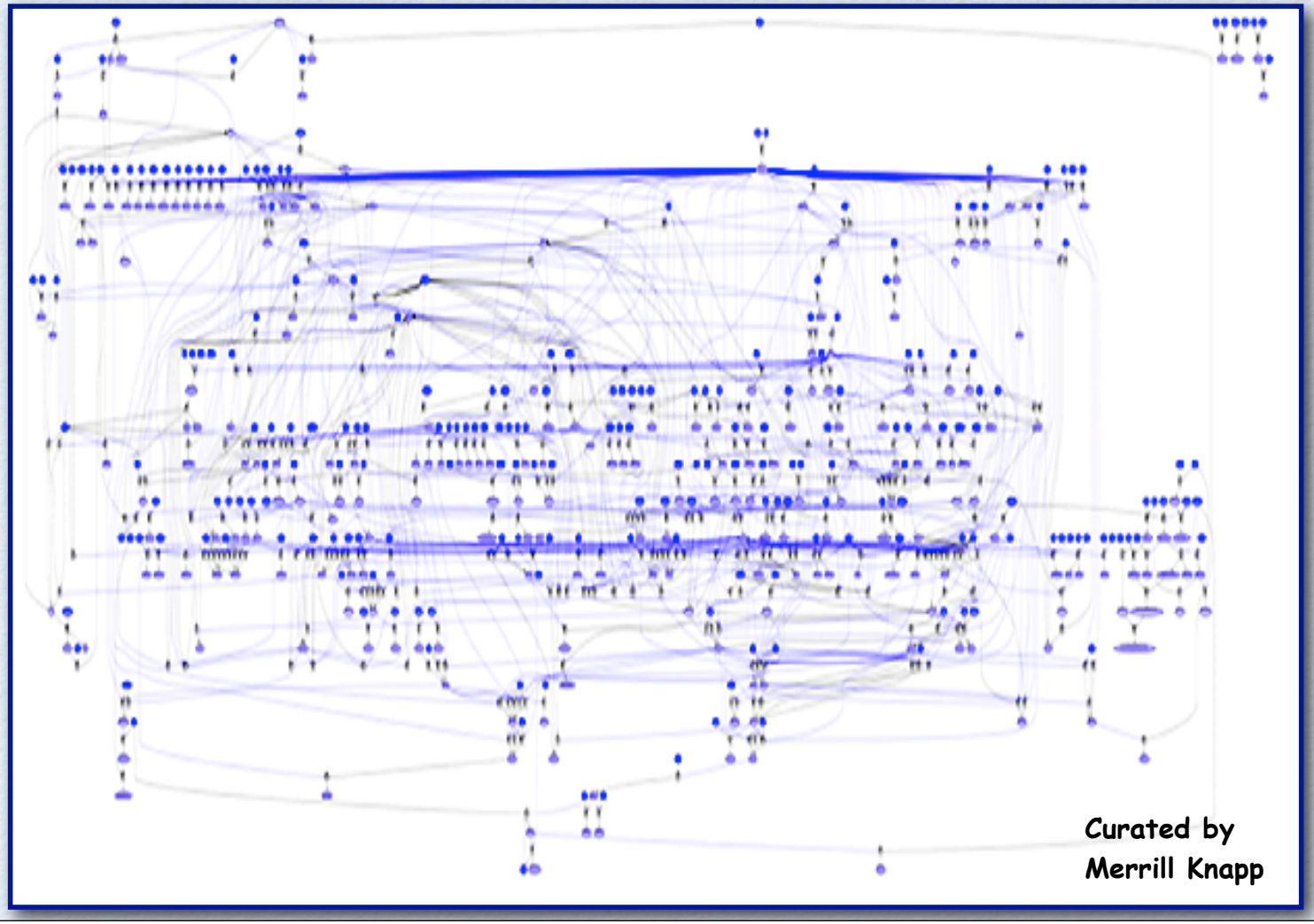
CANONICAL EGF-ERK PATHWAY

Egf ➔ EgfR ➔ Grb2 ➔ Sos1 ➔ a Ras family member ➔ Raf1 ➔ MEK1/2 ➔ ERK1/2

- Egf binds to the EGF receptor (EgfR), stimulates its protein tyrosine kinase activity, causing autophosphorylation (and activation).
- a complex containing the adaptor protein Grb2 and the guanine nucleotide exchange factor Sos1 docks to the activated EgfR.
- the Sos1-containing EgfR complex activates a Ras family GTPase,
- the activated Ras protein activates Raf1, a member of the RAF serine/threonine protein kinase family.
- Raf1 then activates the dual-specificity protein kinases Mek1 and/or Mek2 (MEK1/2),
- MEK1/2 then activates Erk1 and/or Erk2 (ERK1/2).

PL EGF MODEL

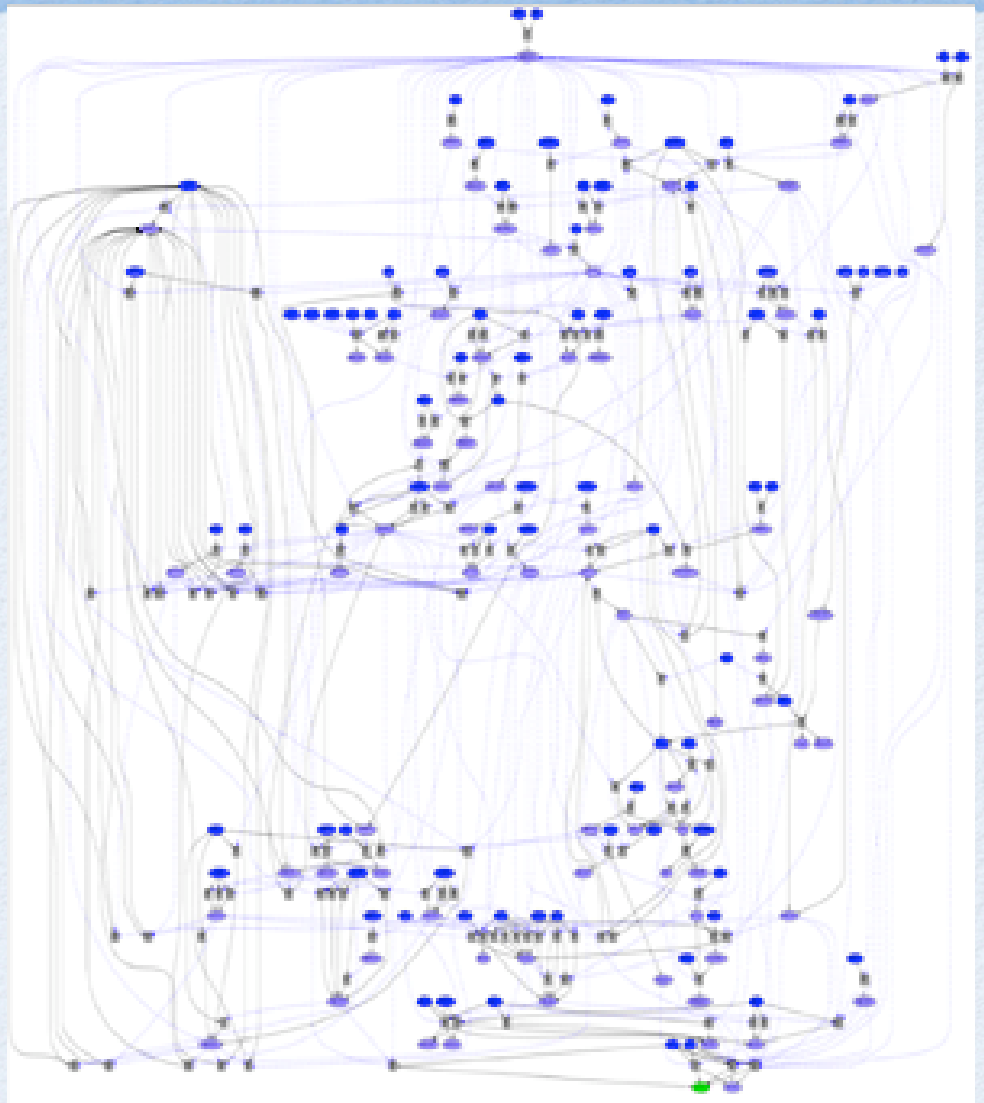
Events that could occur in response to Egf



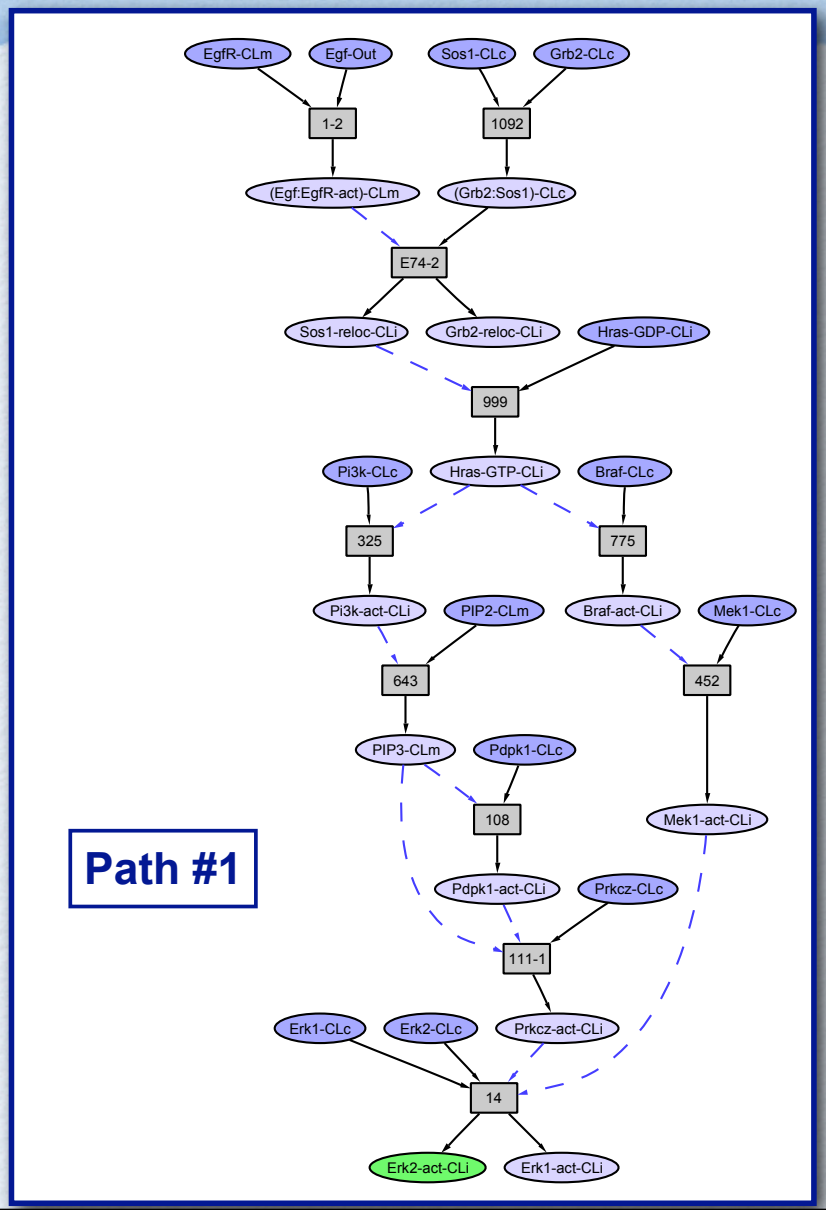
SUBNET RELEVANT TO ERK ACTIVATION

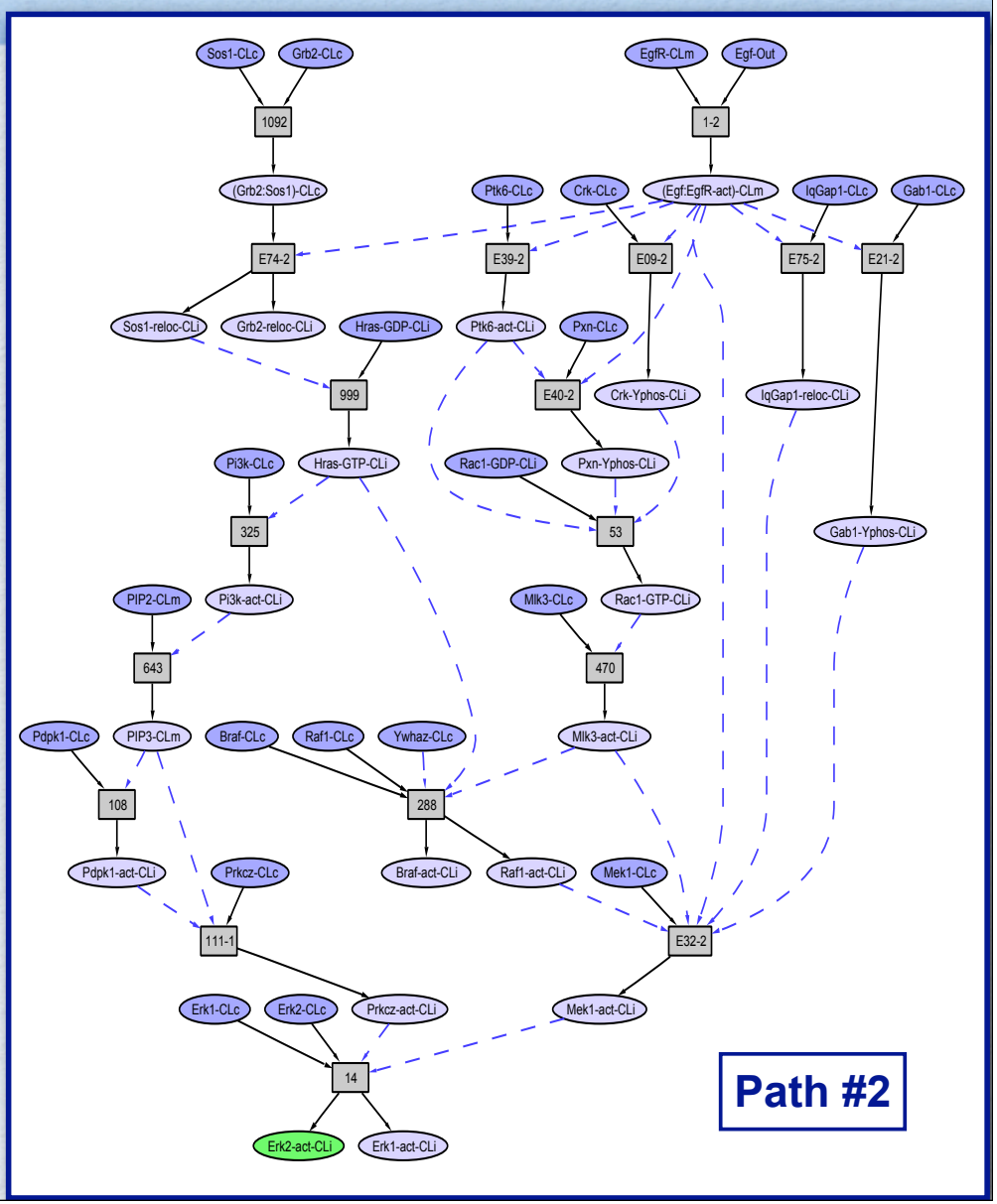
Subnet containing all pathways leading to activation of Erk.

Obtained by backwards followed by forwards collection



POSSIBLE PATHWAYS TO ERK



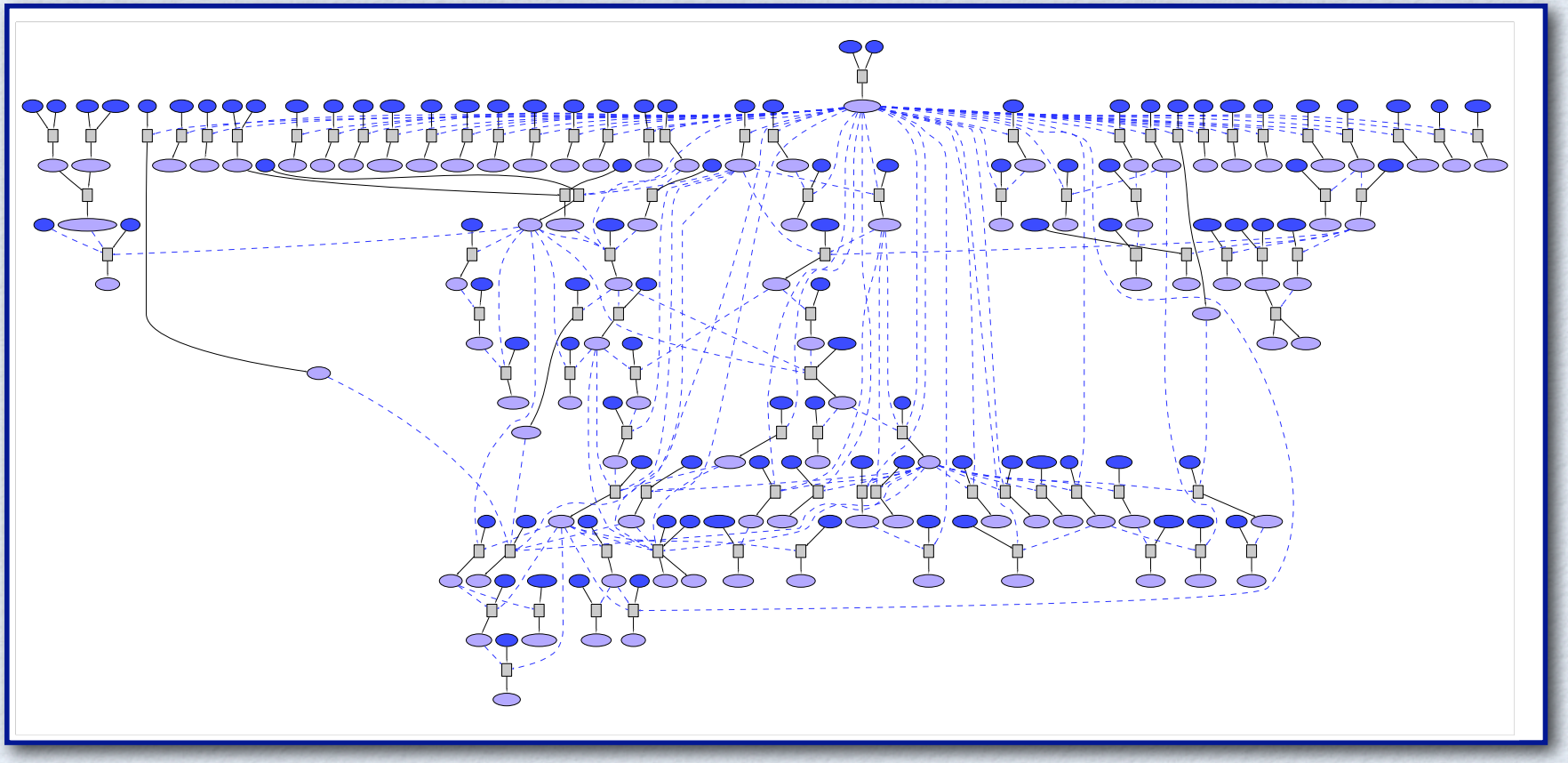


CONSTRAINING THE EGF MAP

The idea is to go from all possible pathways to a plausible set, given the context.

- a list of 85 protein state changes demonstrated experimentally to occur in response to a short stimulus with Egf were set as goals and a set of concurrent paths were produced by PLA. This subnet ensures that the paths used to reach chosen goals are mutually compatible.
- (reachability of all of the goals is also a test of the model)
- Egf Rules, with requirements specific to Egf signaling, were given preference over Common Rules

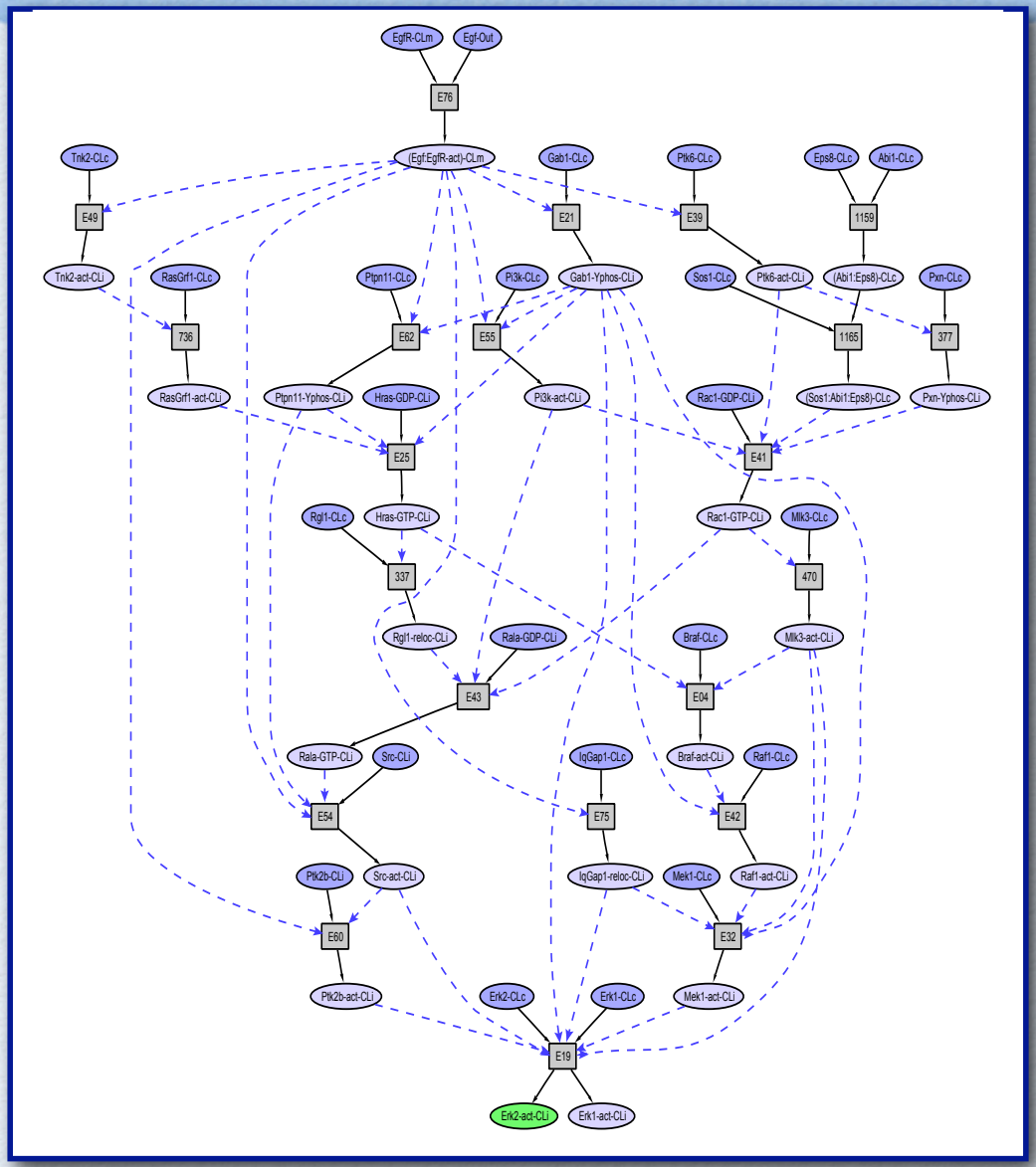
THE CONSTRAINED EGF MAP



ACTIVATION OF ERK IRT EGF

The path leading to activation of Erk1/2 in the constrained Egf network.

This path exists in the context of all the other experimental observations,



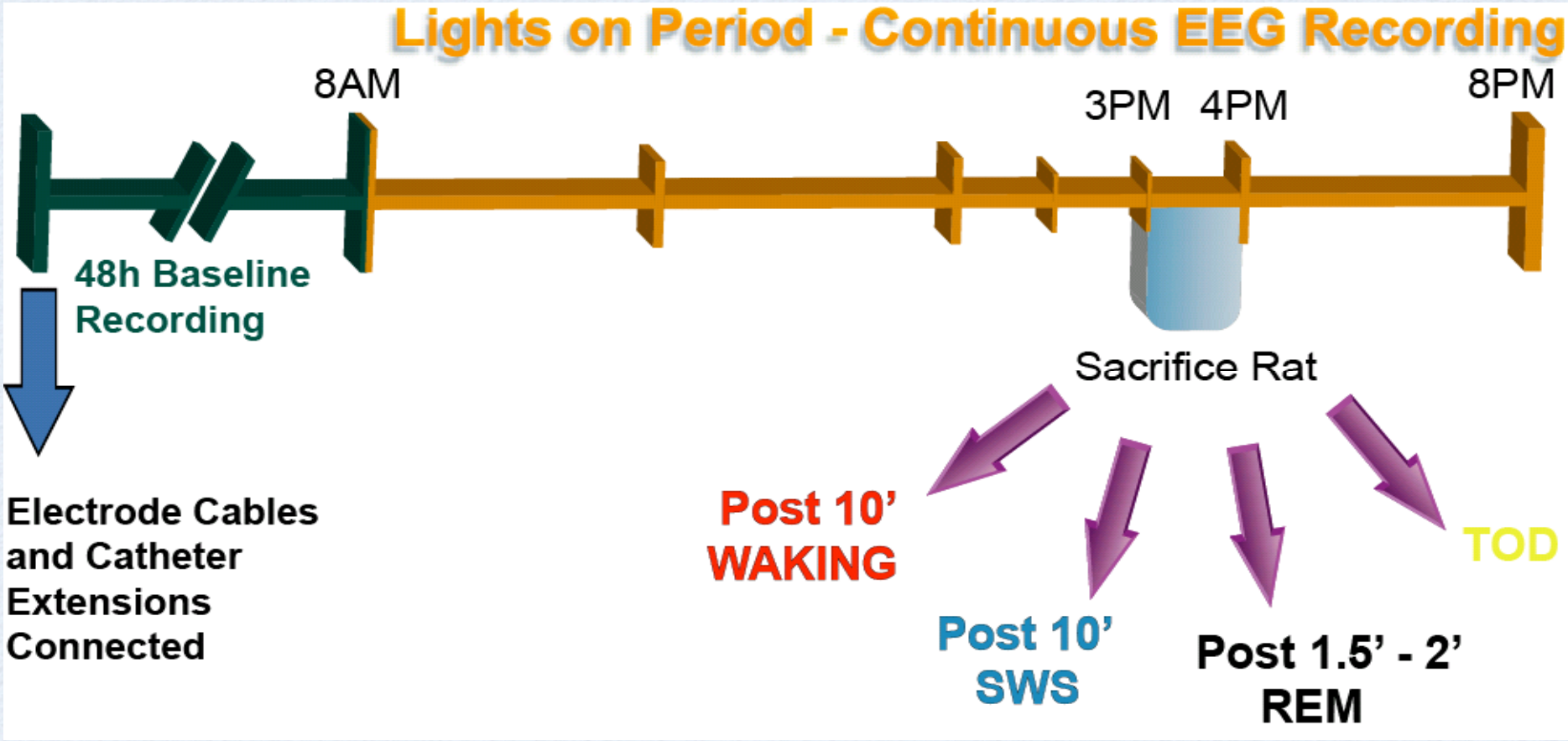
SLEEP

(with MaryAnn Greco and Merrill Knapp)

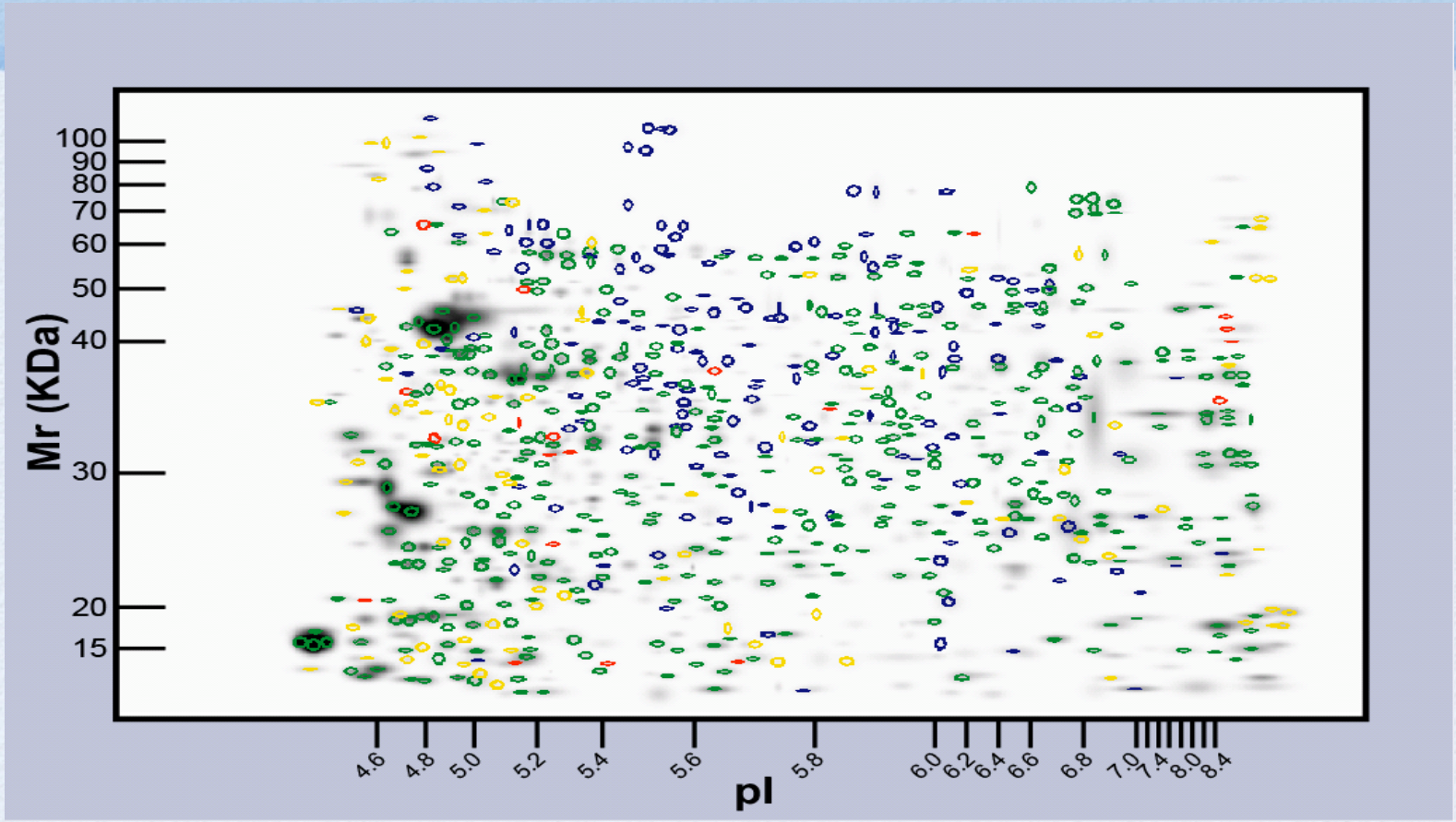
THE QUESTION

- 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



2D MASTER GEL

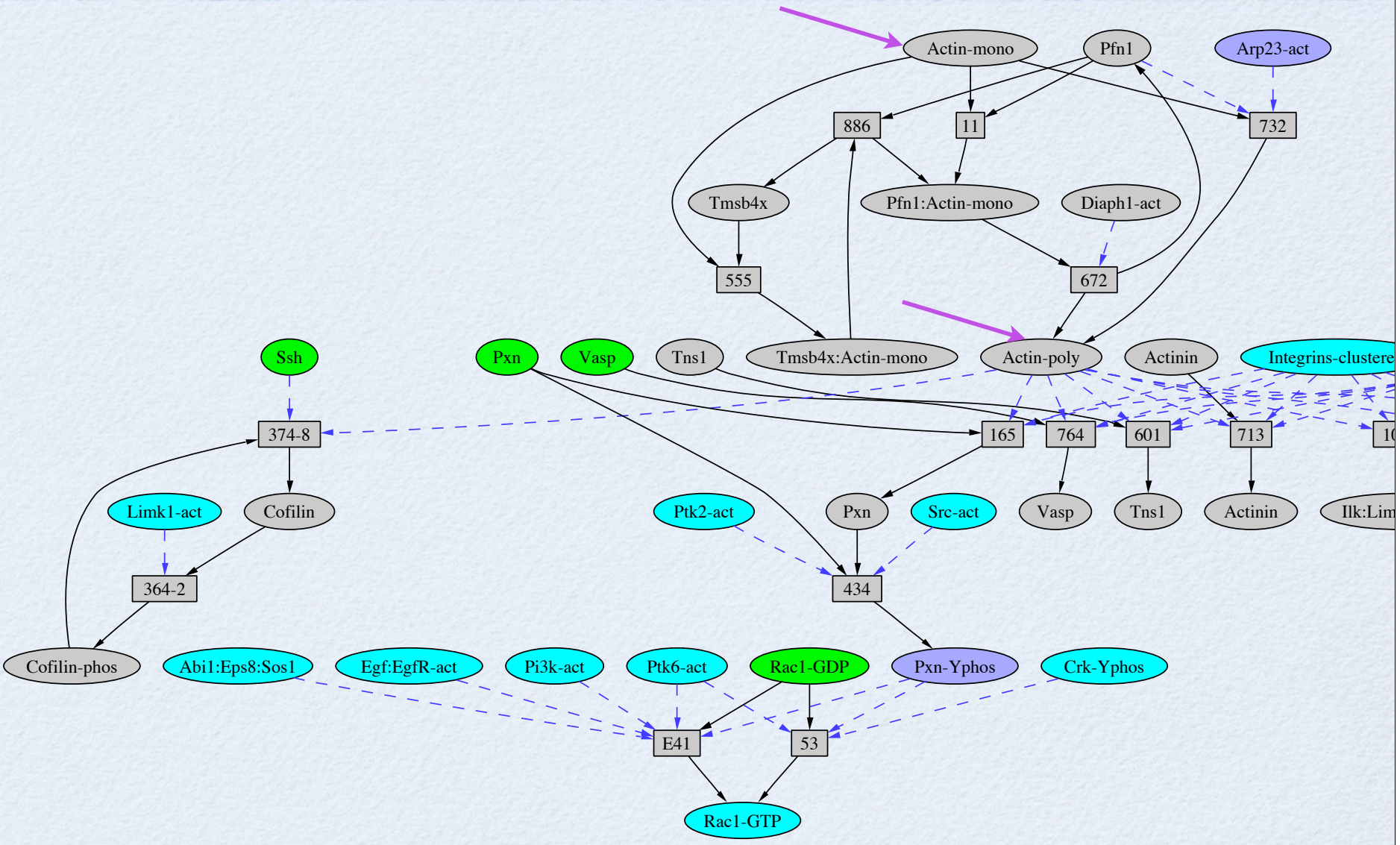


Proteins unique to different states were identified

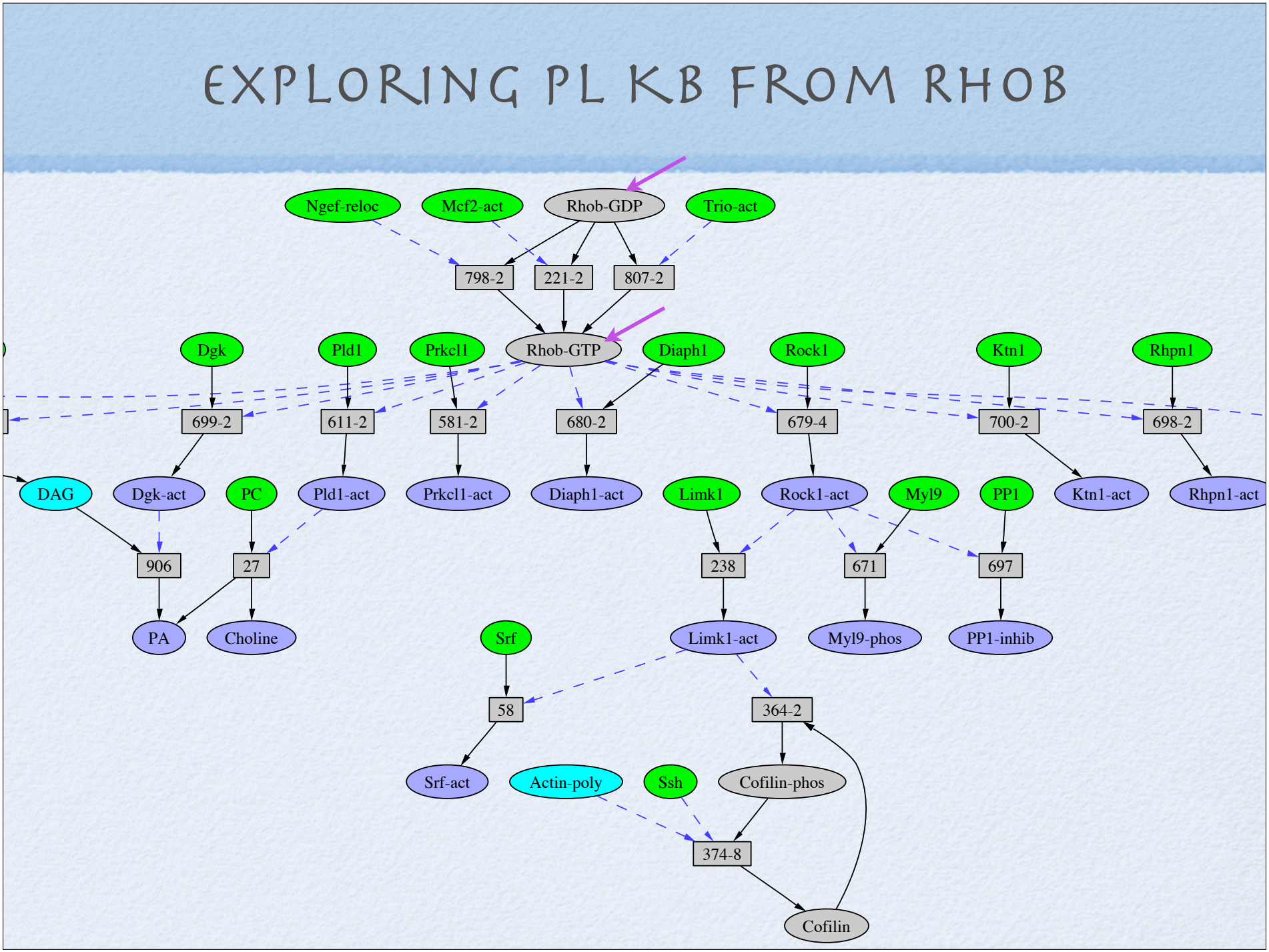
Those modeled in PL included Actin and Rhob

Use the PLA explorer to find signaling connections

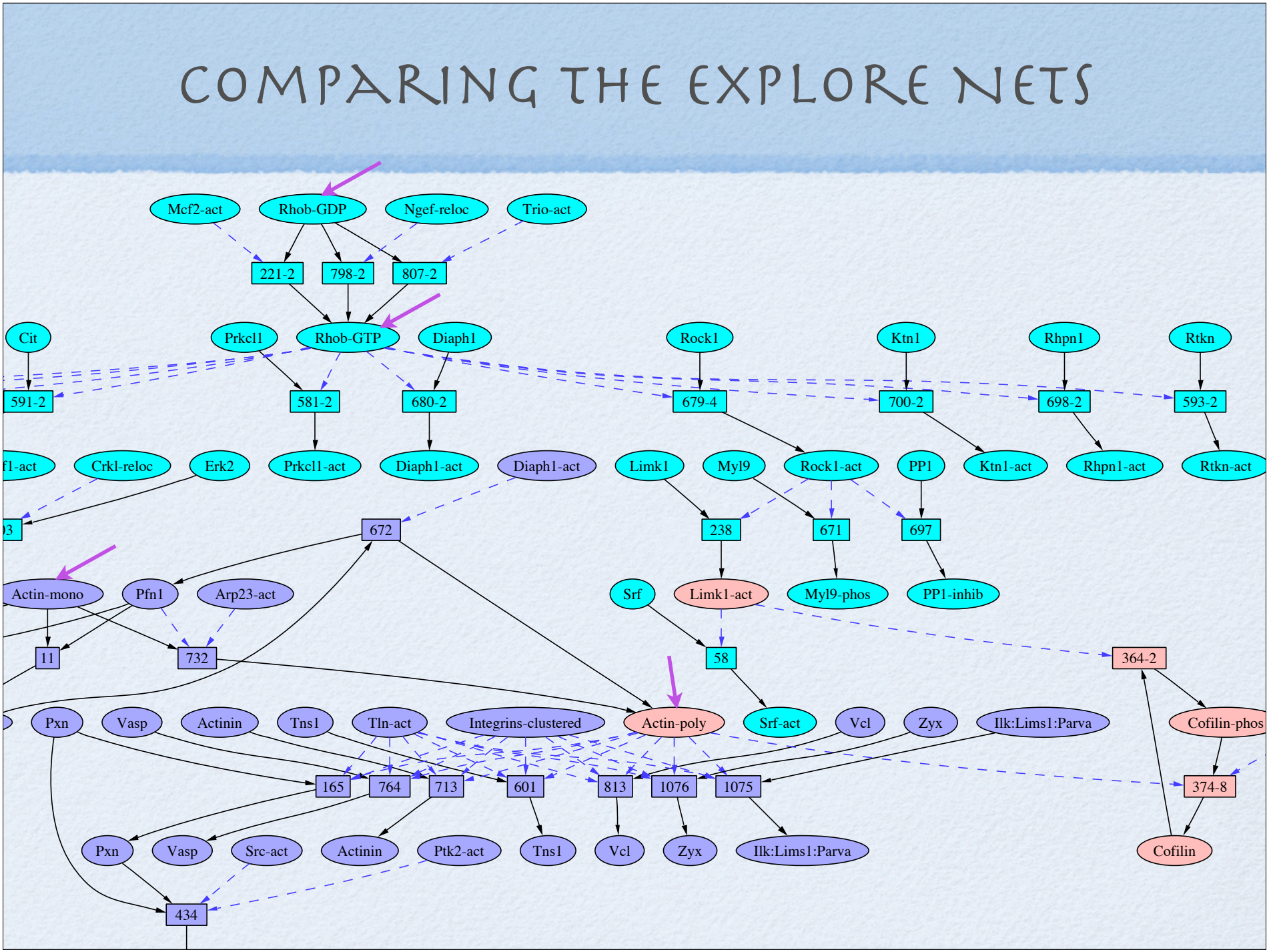
EXPLORING PL KB FROM ACTIN



EXPLORING PL KB FROM RHOB



COMPARING THE EXPLORE NETS

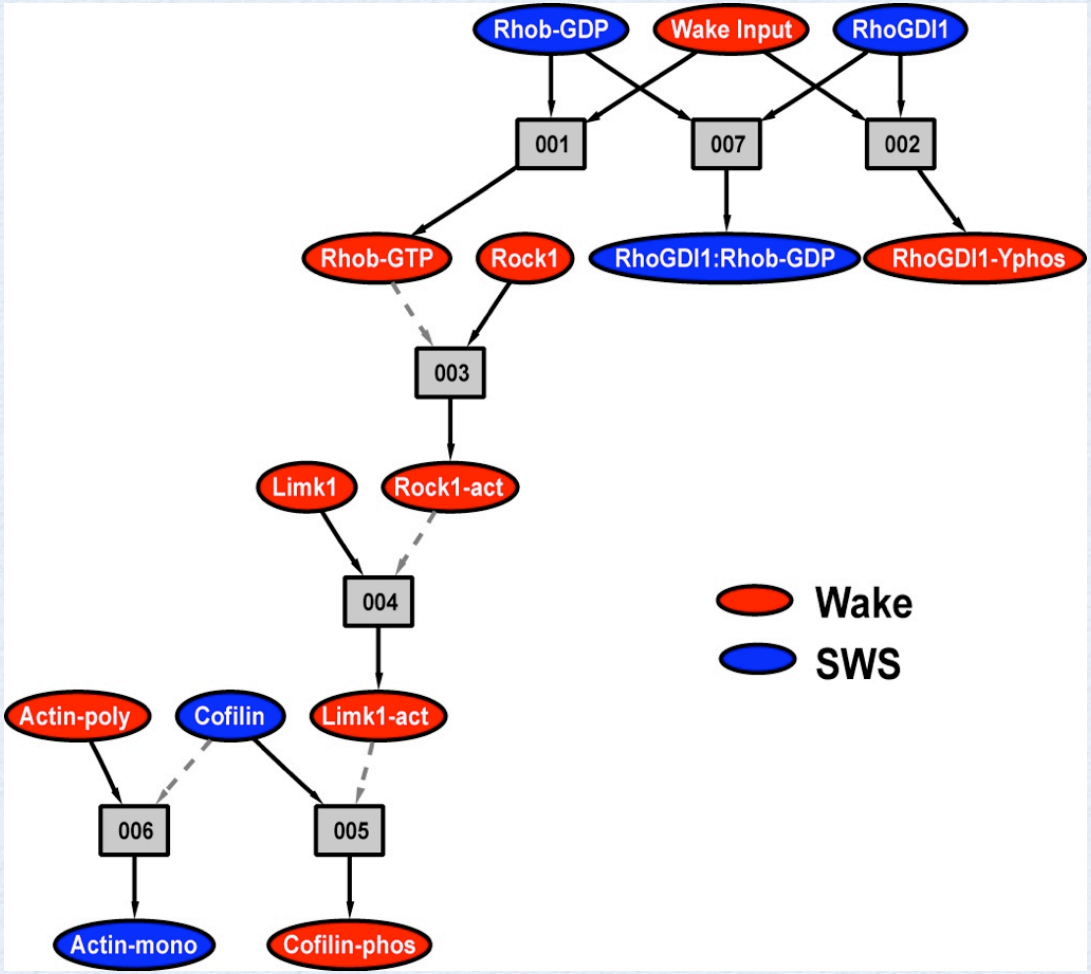


Picking out the interesting bits (with some additional curation)

from 2D-Gels:

State	Protein	Data	Hypothetical Modification
Wake	RhoGdi	unique spot	Yphos
	RhoB	unique spot	GTP
	Grb3{1}	unique spot	upreg (splice variant)
	Cofillin	common	phos
	Actin	phosphorylated	polymerized
SWS	RhoGdi	unique spot gone	unphosed
	Rhob	unique spot gone	GDP
	Grb3	unique spot gone	downreg(splice variant)
	Cofillin	common	unphos
	Actin-phos	less phosphorylated	depolymerized{2}

A HYPOTHETICAL MODEL PATHWAY RELATING STATE AND SYNAPTIC 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:
RhoDG11 binds Rhob-GDP
(is not phosphorylated)
=> Rock1, Limk1, and cofillin would not be phosphorylated and
=> actin depolymerization
=> decrease in synaptic weight

TODO -- test the hypothesis

MODELING METABOLISM

(work of Malabika Sarker)

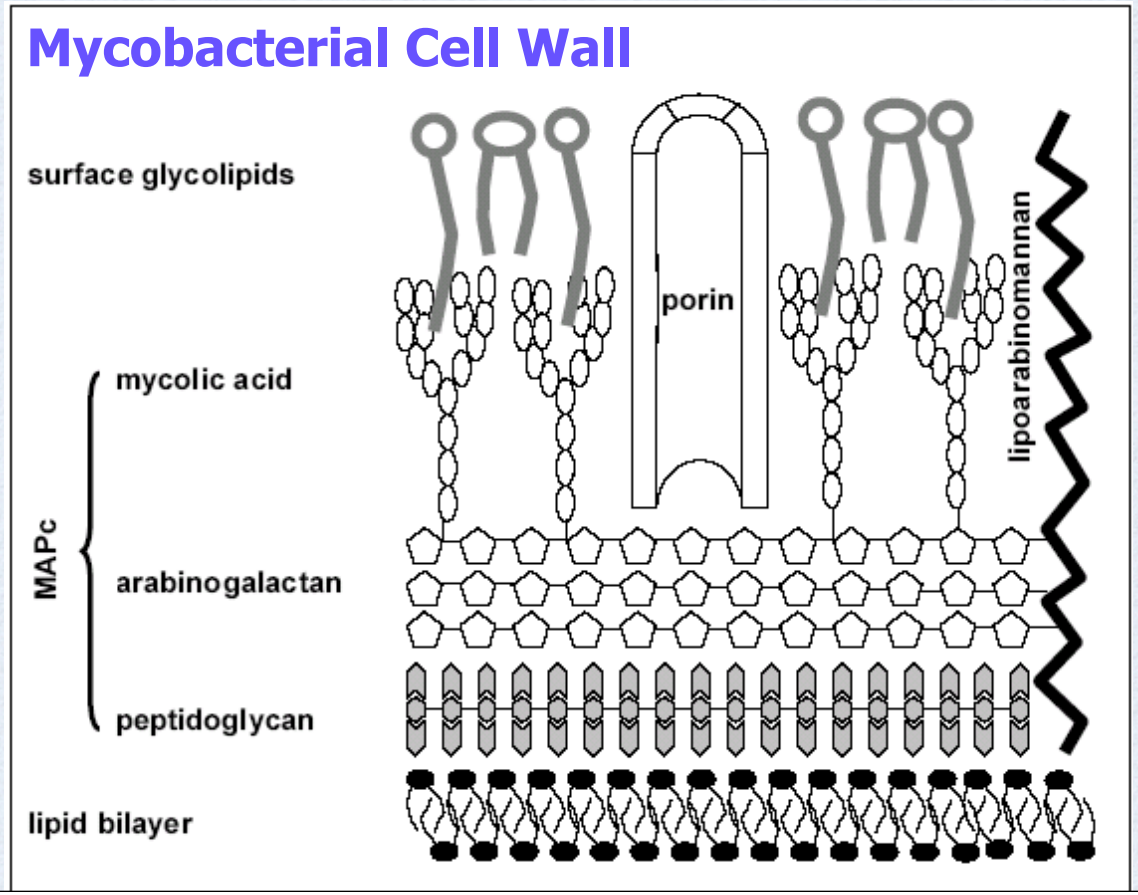
THE PROBLEM

- 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

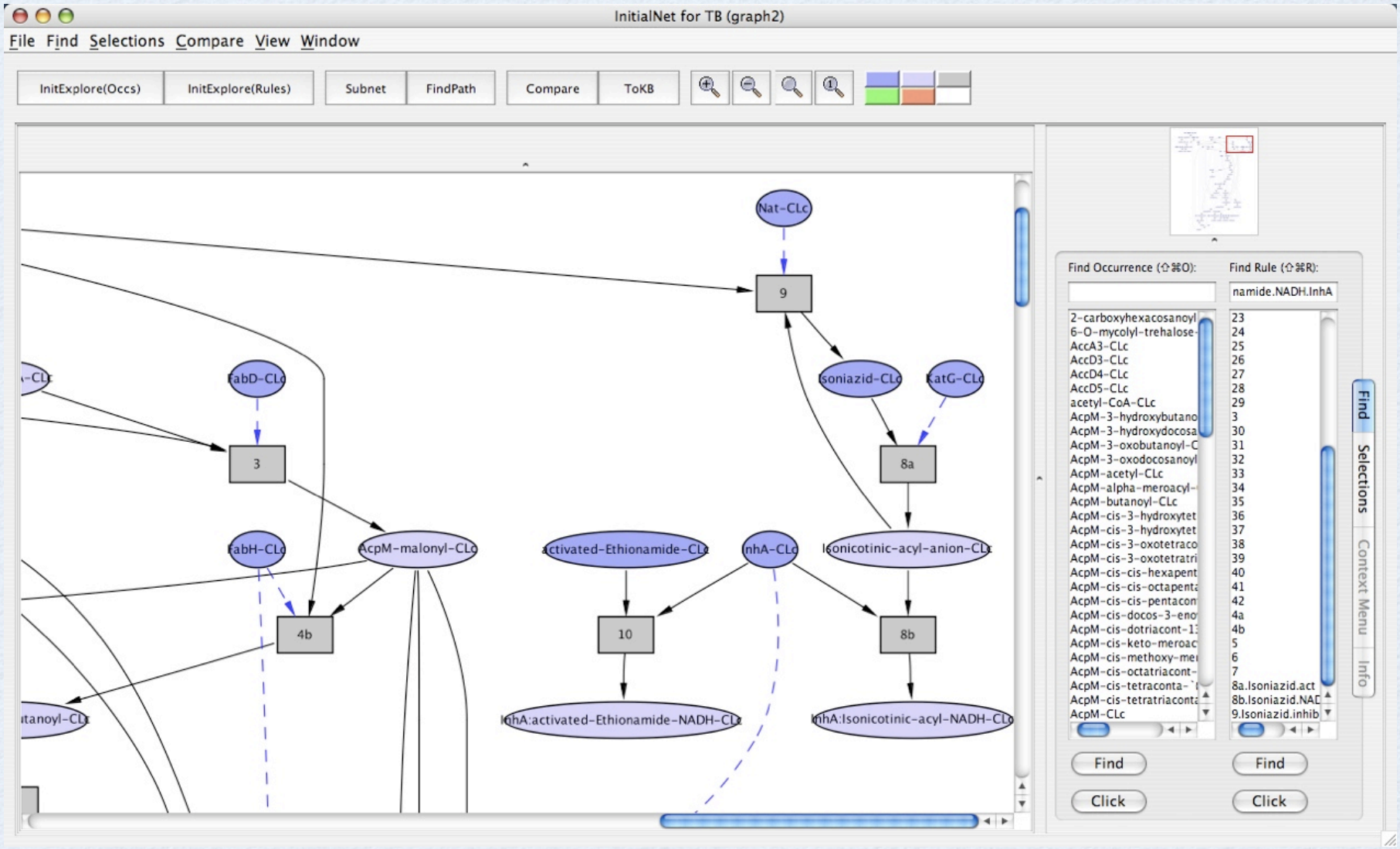
WHY MYCOLIC ACID

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

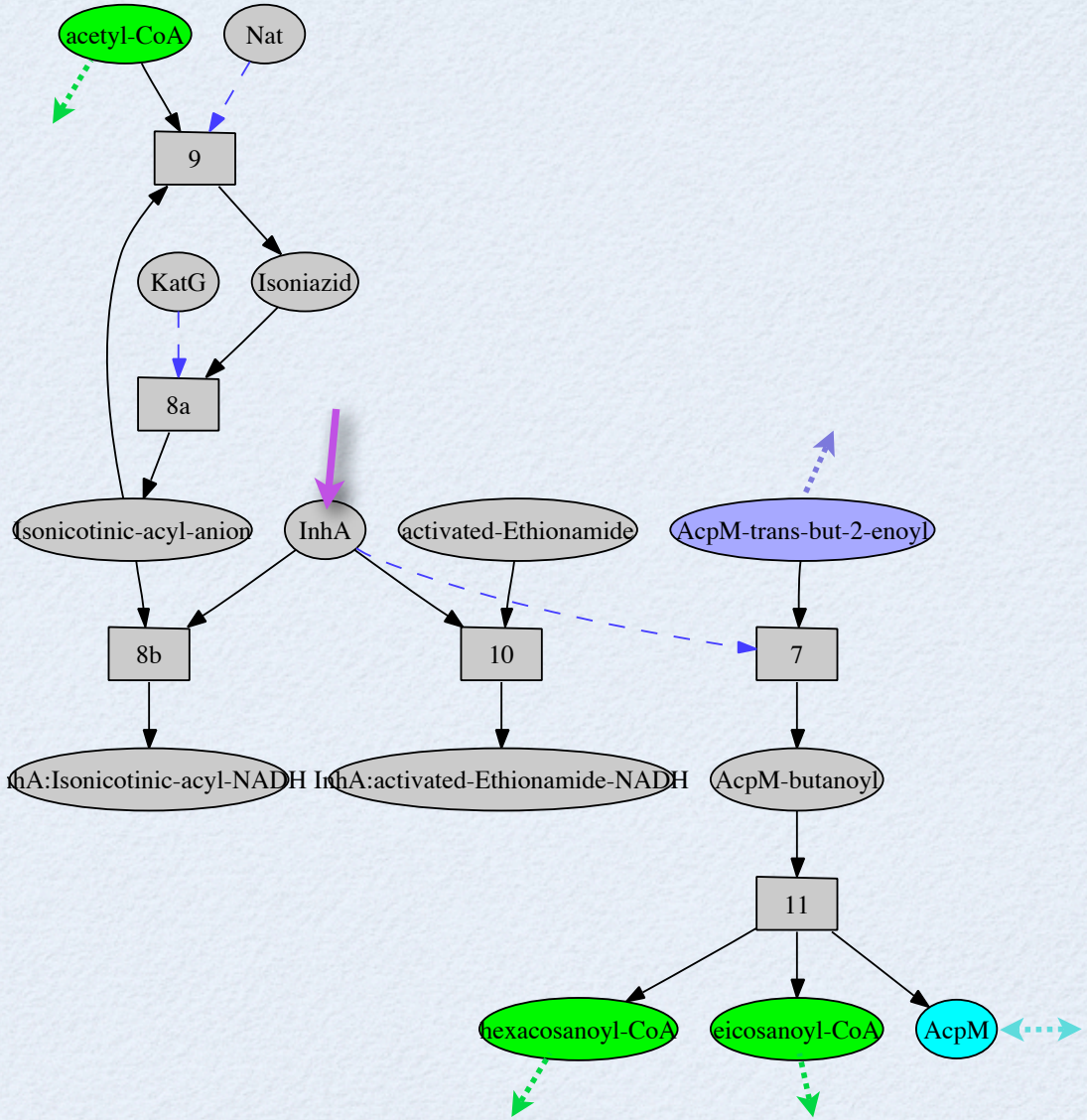
Isoniazid(INH)/Ethionammide(ETH)/Triclosan(TRC) --| InhA



Mycolic Acid Synthesis (TB) Curated



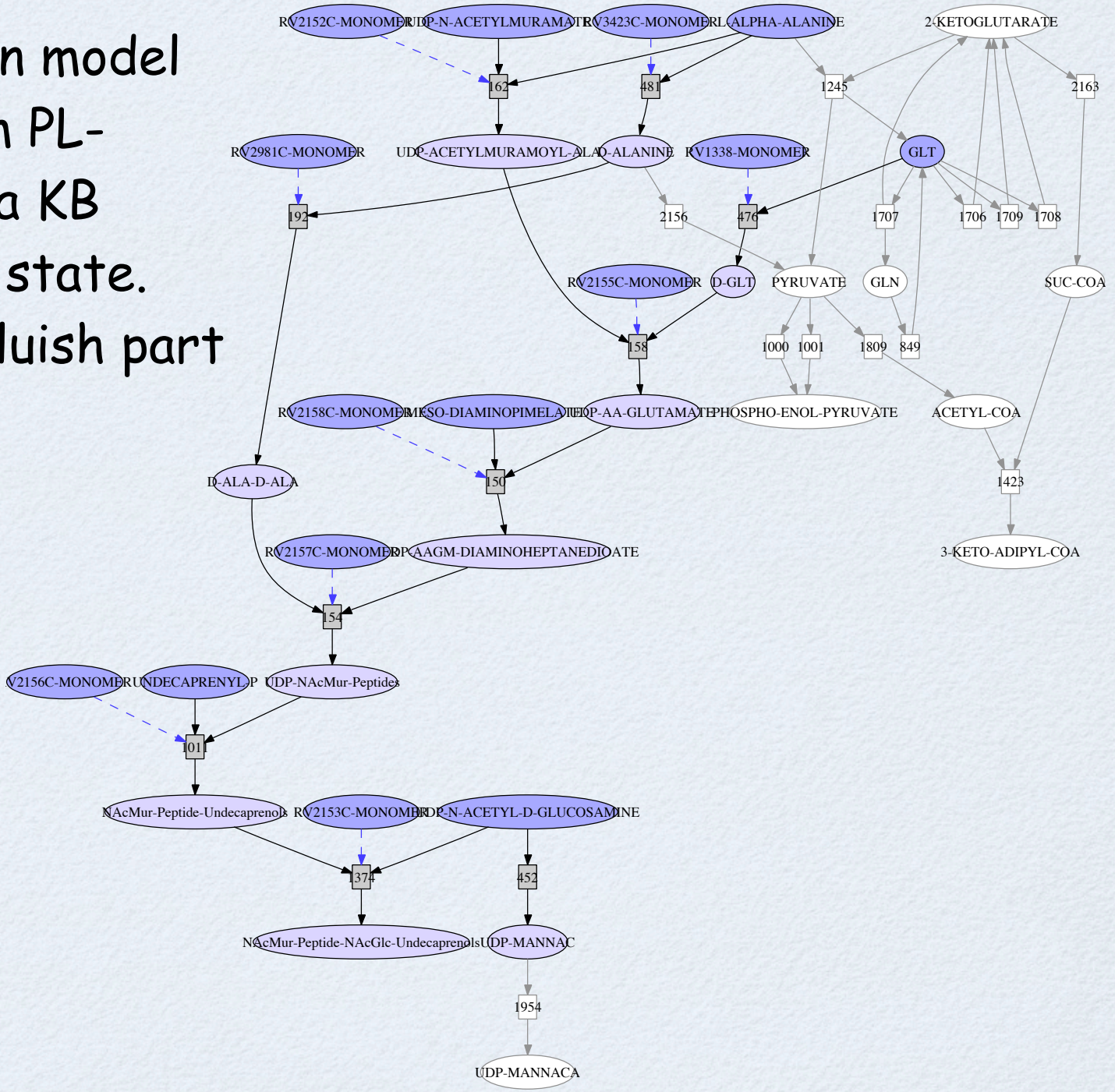
Mycolic acid fragment showing inhibition of InhA



IMPORTING PGDBS INTO PL

- Map compounds to PL components
- Start with reaction and enzrxn files
- Extract information for PL rules
  - lhs, rhs, enzyme
  - (determine direction)
- Convert to PL syntax
- Apply to M. tuberculosis H37Rv PGDB

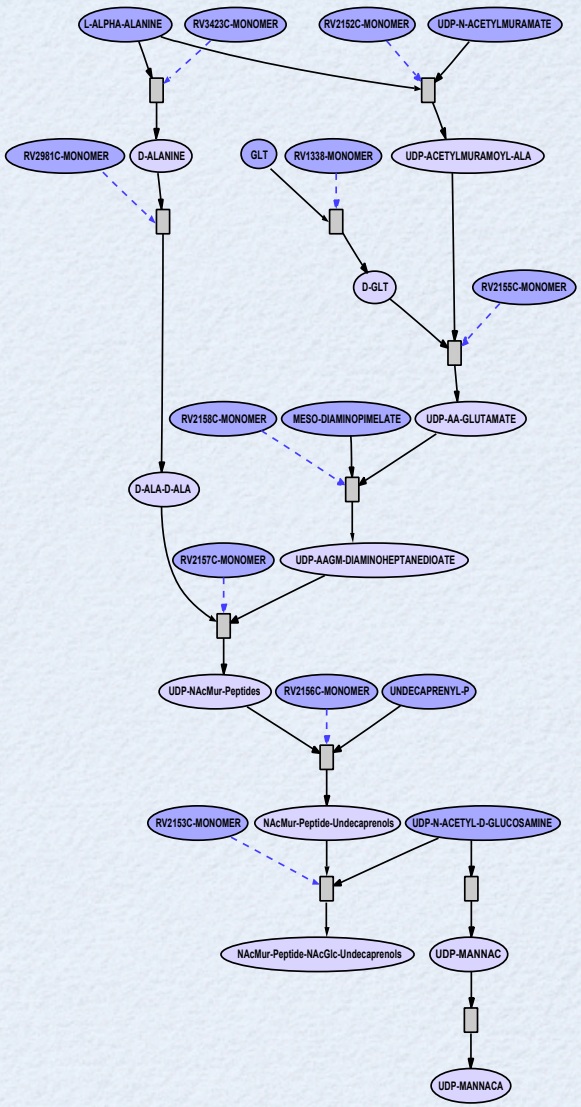
Peptidoglycan model derived from PL-mycobacteria KB and starting state. Pathway is bluish part



PEPTIDO-GLYCAN PATHWAY

From Biocyc

Assembled in PL



WHITHER

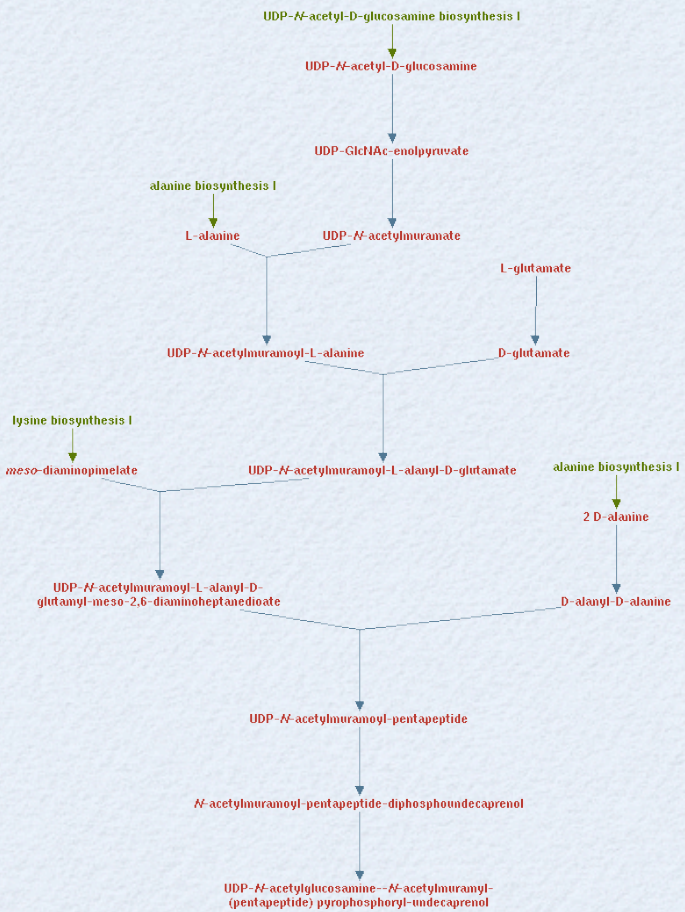
- Further refining of PGDB to PL (and back)
- Integrate probablistic / stochastic reasoning
- Integration of signaling and metabolic networks
  - What can be learned by qualitative reasoning?
- Host response -- integrating host and pathogen models

PATHWAY LOGIC TEAM

- Keith Laderoute
- Patrick Lincoln
- Carolyn Talcott

- Linda Briesemeister
- Steven Eker
- Merrill Knapp
- Ian Mason
- Andy Poggio
- Malabika Sarker
- Ashish Tiwari

- Biology Computer Science


UDP-*N*-acetyl-D-glucosamine biosynthesis I
UDP-*N*-acetyl-D-glucosamine
UDP-GlcNAc-enolpyruvate
alanine biosynthesis I
L-alanine
UDP-*N*-acetylmuramate
L-glutamate
UDP-*N*-acetylmuramoyl-L-alanine
D-glutamate
lysine biosynthesis I
meso-diaminopimelate
UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate
alanine biosynthesis I
2 D-alanine
UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminoheptanedioate
D-alanyl-D-alanine
UDP-*N*-acetylmuramoyl-pentapeptide
N-acetylmuramoyl-pentapeptide-diphosphoundecaprenol
UDP-*N*-acetylglucosamine--*N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol