PATHWAY LOGIC
FORMAL METHODS FOR BIOLOGY

Carolyn Talcott
SRI International
April 2009

PLAN

- Symbolic systems biology
- Executable Specification in RWL
- Biological Processes (What to model)
- Representation in PL
- Computing with PL models
  - Small KB
  - Egf Stimulation

SYMBOLIC SYSTEMS BIOLOGY

BIOLOGICAL SYSTEMS

- Biological processes are complex
  - genes, proteins, metabolites
  - cells, organs, organisms
- Dynamics that range over huge timescales
  - microseconds to years
- Spatial scales over 12 orders of magnitude
  - single protein to cell, cell to whole organism
- Oceans of experimental biological data generated
- Important intuitions captured in mental models that biologists build of biological processes

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
- Which biology? Causal networks of biomolecular interactions and reactions
- Goals:
  - Develop formal models that are as close as possible to domain expert's mental models
  - Compute with, analyze and reason about these complex networks
  - New insights into / understanding of biological mechanisms

LOGICAL FRAMEWORK

- Making description and reasoning precise
- Language
  - for describing things and/or properties
  - given by a signature and rules for generating expressions (terms, formulas)
- Semantic model -- mathematical structure (meaning)
  - interpretation of terms
  - satisfaction of formulas: M |= wff
- Reasoning -- rules for inferring valid formulae
- Symbolic model -- theory (axioms) used to answer questions

EXECUTABLE SYMBOLIC MODELS

- Describe system states and rules for change
- From an initial state, derive a transition graph
  - nodes -- reachable states
  - edges -- rules connecting states
- Path -- sequence of nodes and edges in transition graph (computation / derivation)
- Execution strategy -- picks a path

SYMBOLIC ANALYSIS 1

- Static Analysis
  - how are elements organized -- sort hierarchy
  - control flow / dependencies
  - detection of incompleteness
- Forward simulation from a given state (prototyping)
  - run model using a specific strategy
  - fast, first exploration of a model

SYMBOLIC ANALYSIS II

- Forward search from a given state
  - breadth first search of transition graph
  - find ALL possible outcomes
  - find only outcomes satisfying a given property
- Backward search from a given state S
  - run a model backwards from S
  - find initial states leading to S
  - find transitions that contribute to reaching S

SYMBOLIC ANALYSIS III

- Model checking
  - determines if all pathways from a given state satisfy a given property, if not a counter example is returned
  - example property:
    - molecule X is never produced before Y
  - counter example:
    - pathway in which Y is produced after X

SYMBOLIC ANALYSIS IV

- Constraint solving
  - Find values for a set of variables satisfying given constraints.
  - MaxSat deals with conflicts
    - weight constraints
    - find solutions that maximize the weight of satisfied constraints
  - Finding possible steady state flows of information or chemicals through a system can be formulated as a constraint problem.

SYMBOLIC ANALYSIS V

- Meta analysis -- reasoning about the model itself
  - find transitions producing / consuming X
  - find all phosphorylation reactions
  - check that transitions satisfy some property such as stoichiometry
  - transform a model and property to another logic (for access to tools)

A SAMPLING OF FORMALISMS

- Rule-based + Temporal logics
- Petri nets + Temporal logics
- Membrane calculi -- spatial process calculi / logics
- Statecharts + Live sequence charts
- Stochastic transitions systems and logics
- Hybrid Automata + Abstraction

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!)

BIOLOGICAL PROCESSES

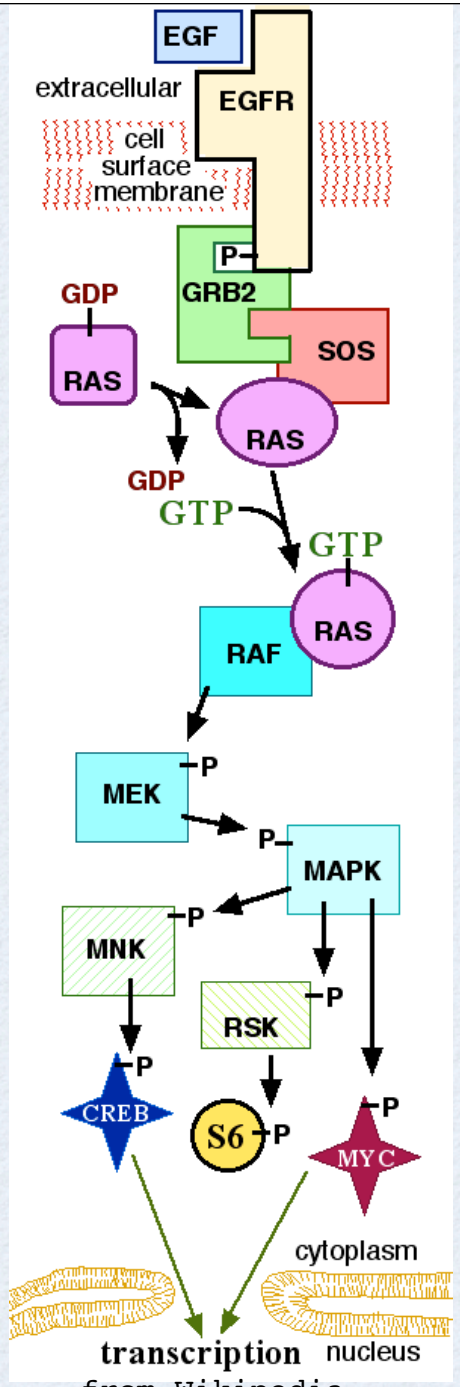
WHAT TO MODEL

CELLULAR SIGNALING

- Cells respond to changes in their environment through biochemical pathways that detect, transduce, and transmit information to effector molecules within different cellular compartments.
- Most signaling pathways involve hierarchical assembly in space and time of multi-protein complexes that regulate the flow of information according to logical rules.
- Biological subnetworks interact to produce high levels of physiological organization (e.g., circadian clock subnetworks are integrated with metabolic, survival, and growth subnetworks).

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.



from Wikipedia

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)
- ...

FEATURES 1

- Naming
  - different biologists use different names for the same protein Egf vs EGF, Erk vs MAPK, EgfR vs ErbB1 vs HerbB1
  - link name to `standard' source: SwissProt, KEGG, HUGO ...
- Activity / state -- a protein may need to be is a specific state (active) to carry out its function
- Location -- what compartment, where in the compartment
  - media -- outside a cell
  - Cell compartment
    - Membrane -- integral, surface, interior
    - Cytoplasm

FEATURES II

- Roles / functions
  - Ligand -- Egf
  - Receptor -- EgfR -- binds ligand
  - Scaffold / adaptor -- complex formation
  - Kinase -- Raf1, Mek, Erk
- Processes
  - recruiting
  - postranslational modification
    - phosphorylation (by kinase)
    - ubiquitination

Pathway Logic (PL) Representation of Signaling

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

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

PATHWAY LOGIC ORGANIZATION

A Pathway Logic (PL) system has four parts

- Theops --- sorts and operations
- Components --- specific proteins, chemicals ...
- Rules --- signal transduction reactions
- Dishes --- candidate initial states

Knowledge base: Theops + Components + Rules

Equational part: Theops + Components

A PL cell signaling model is generated from

- a knowledge base
- a dish

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}]) .
```

RULE 5: RECRUITMENT

Activated EgfR recruits Grb2 to the inside of the cell membrane

```
rl[5.Grb2.reloc]:
  {CLm | clm [EgfR - act]      }
  {CLi | cli                   }
  {CLc | clc Grb2              }
  =>
  {CLm | clm [EgfR - act]      }
  {CLi | cli [Grb2 - reloc]    }
  {CLc | clc                   } .
```

Rewriting rasDish1 with rule 5 results in

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

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)

rasNet

Rule instances relevant to Hras activation

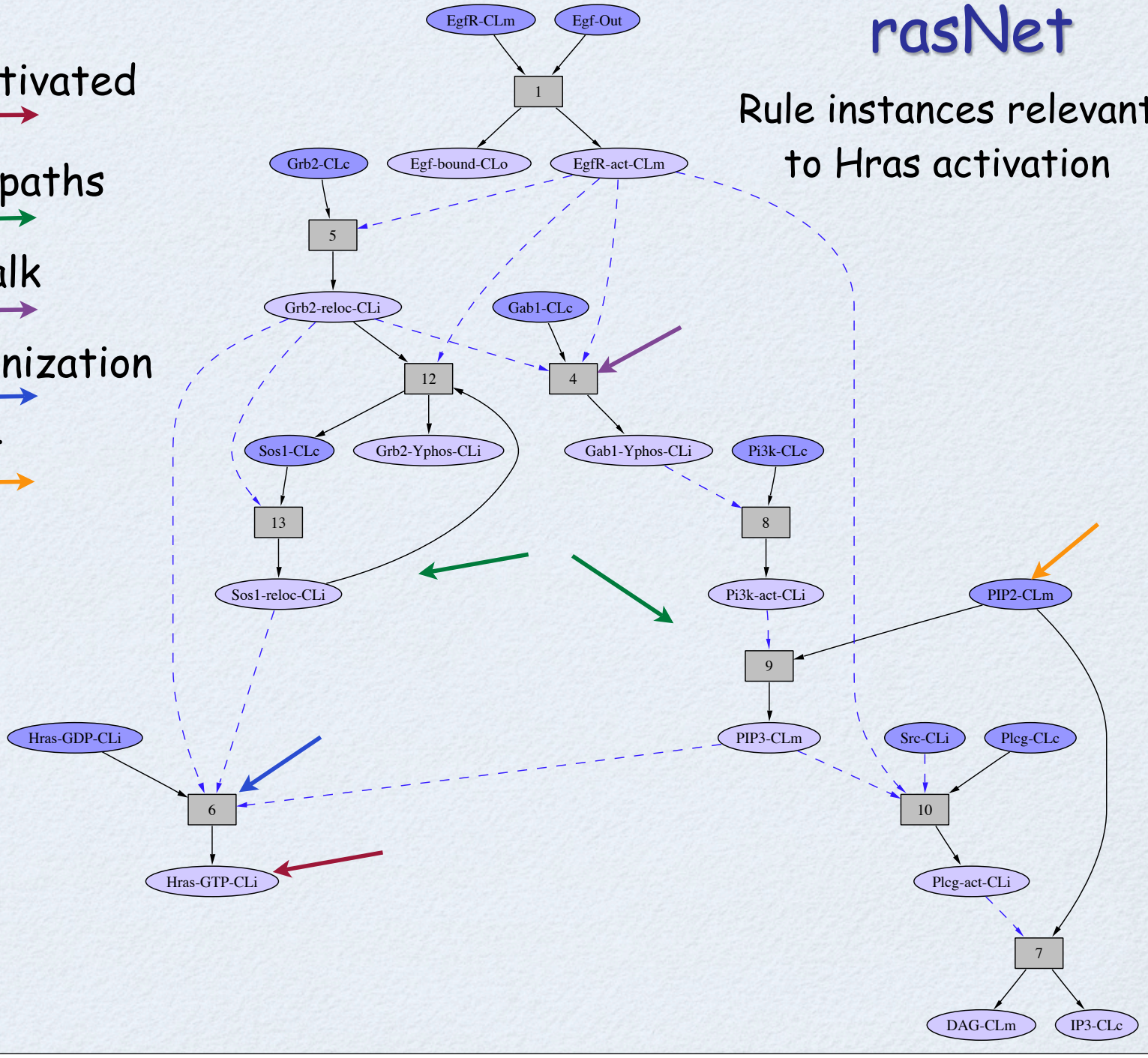
Hras activated

Parallel paths

Cross talk

Synchronization

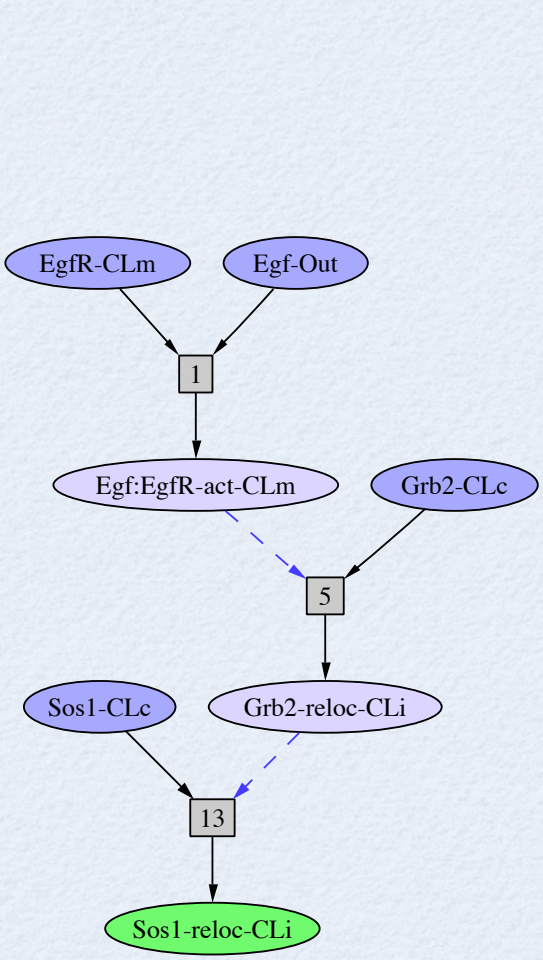
Conflict

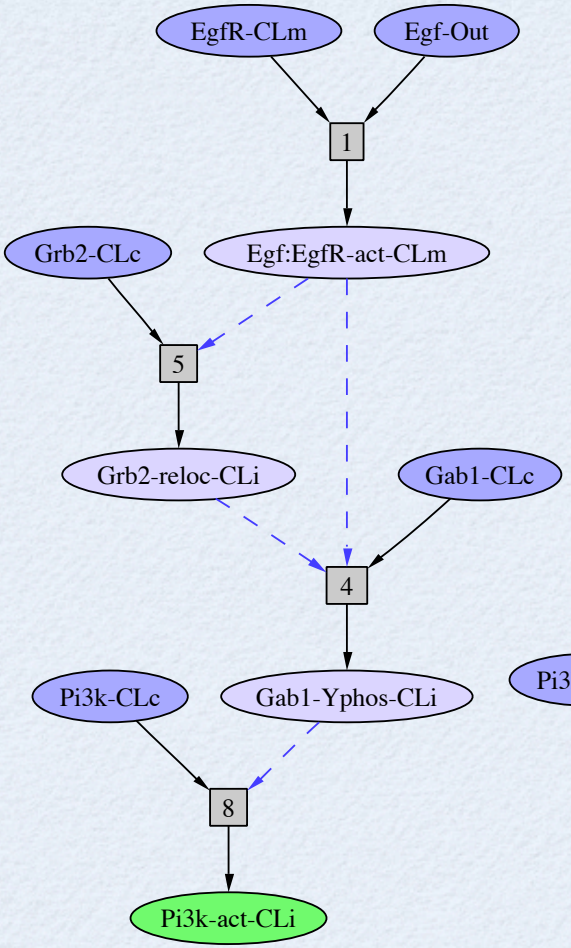


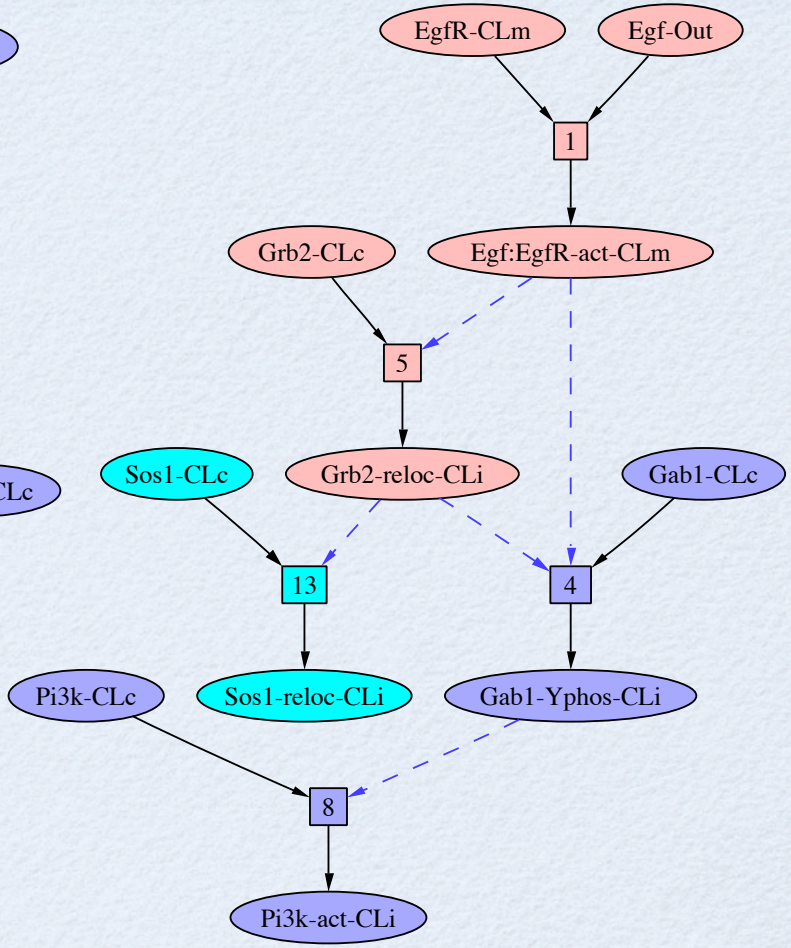
A SIMPLE QUERY LANGUAGE

- 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 requiremments (using model checking on the negation).

PATHWAY EXAMPLES

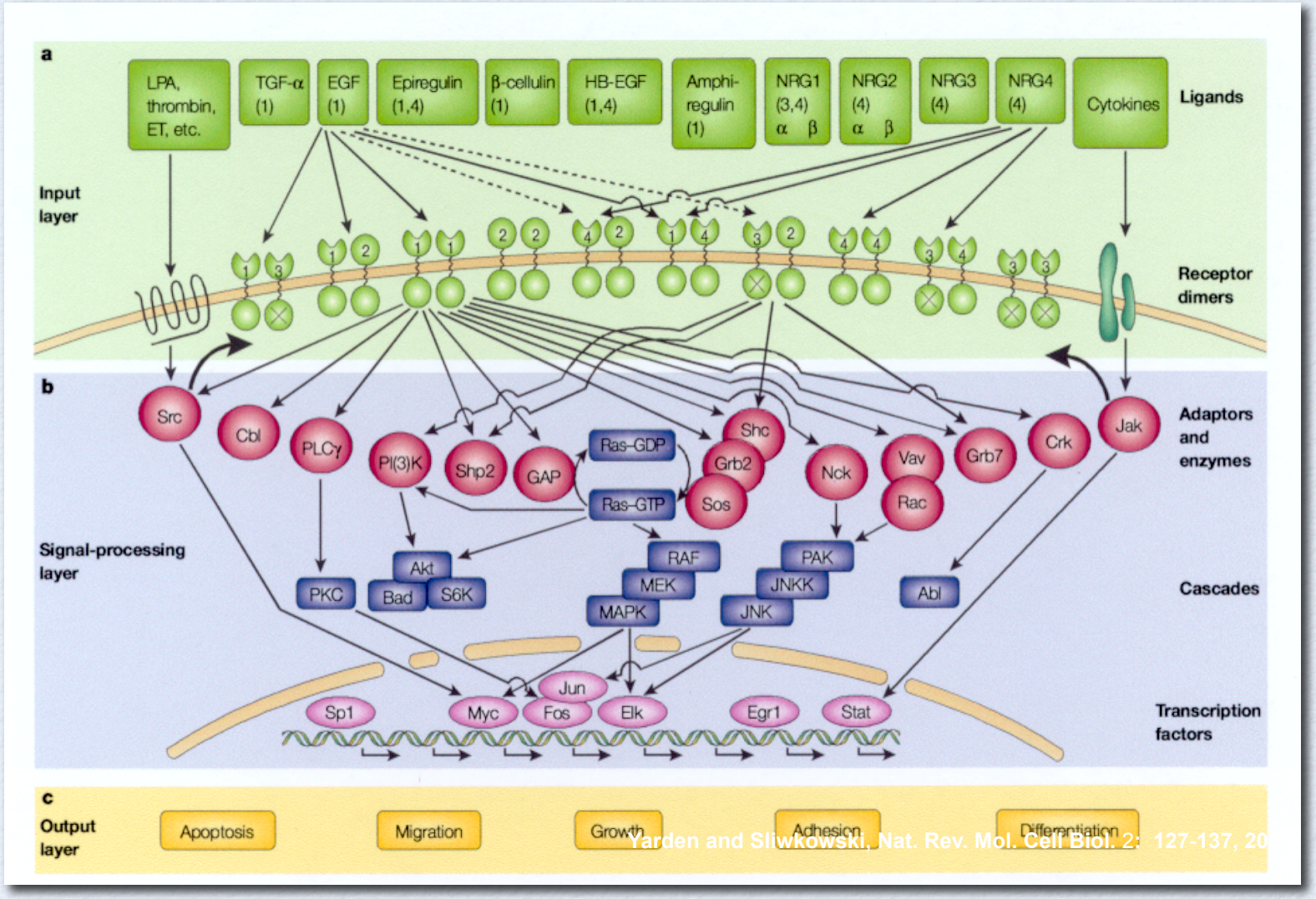






FULL MODEL OF EGF STIMULATION

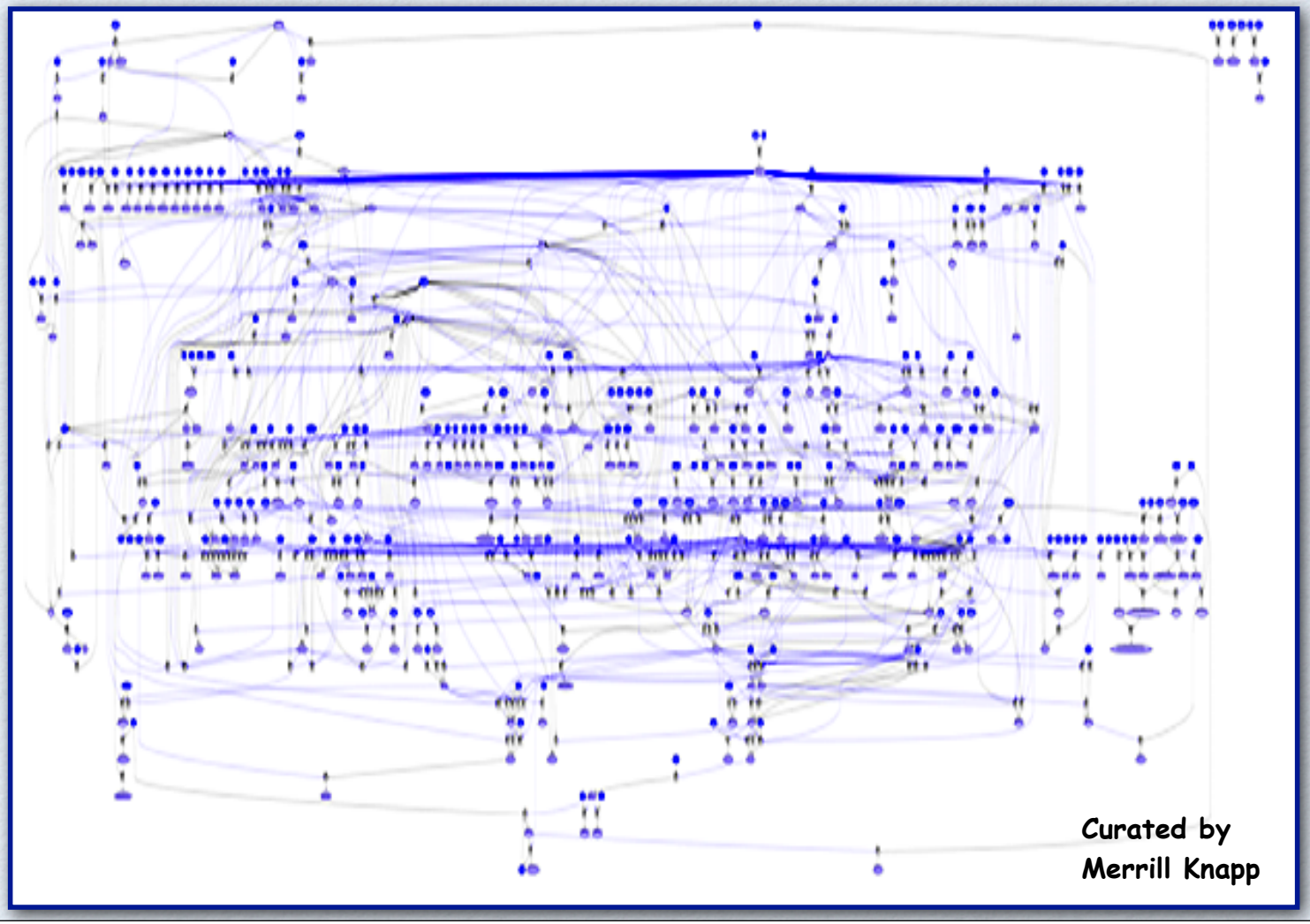
THE ERBB NETWORK (CARTOON FORM)



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

PL EGF MODEL

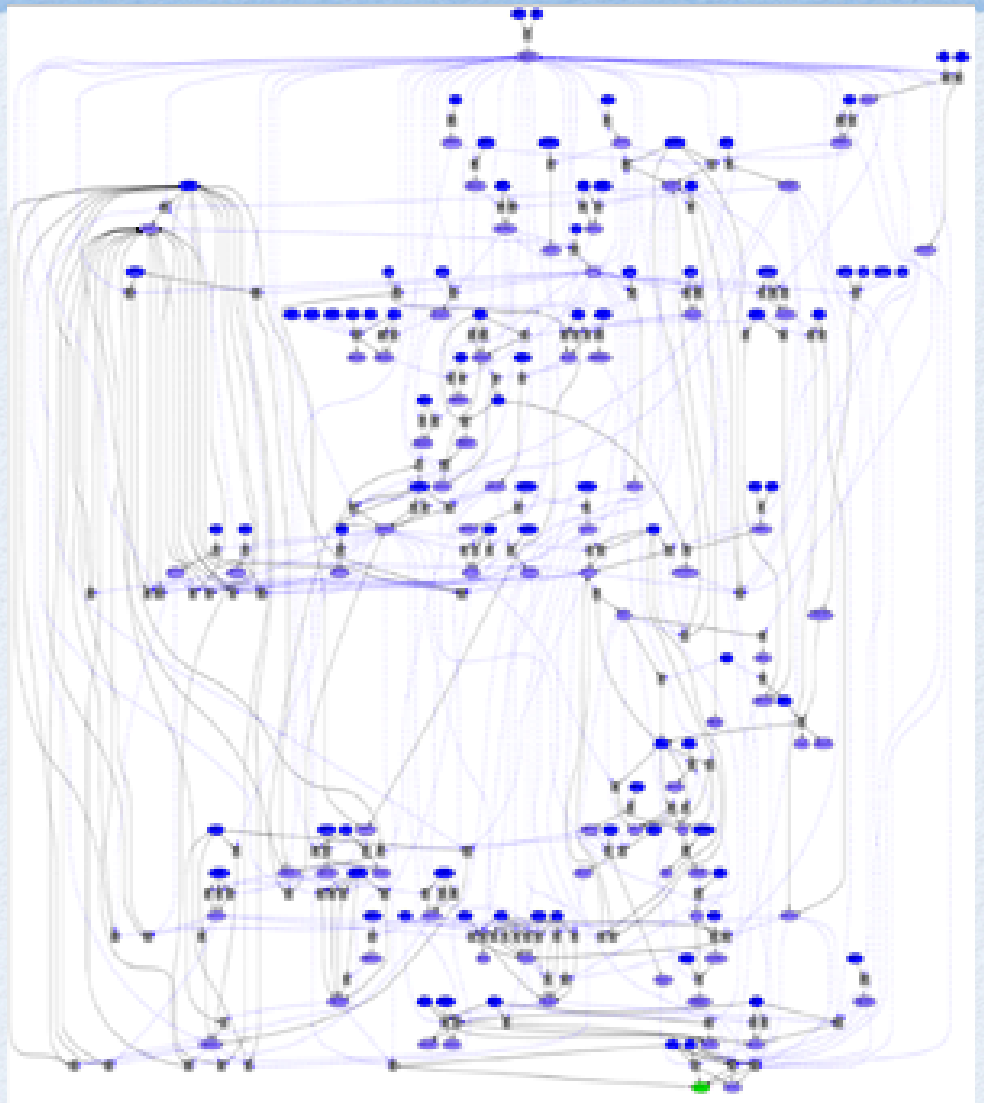
Events that could occur in response to Egf



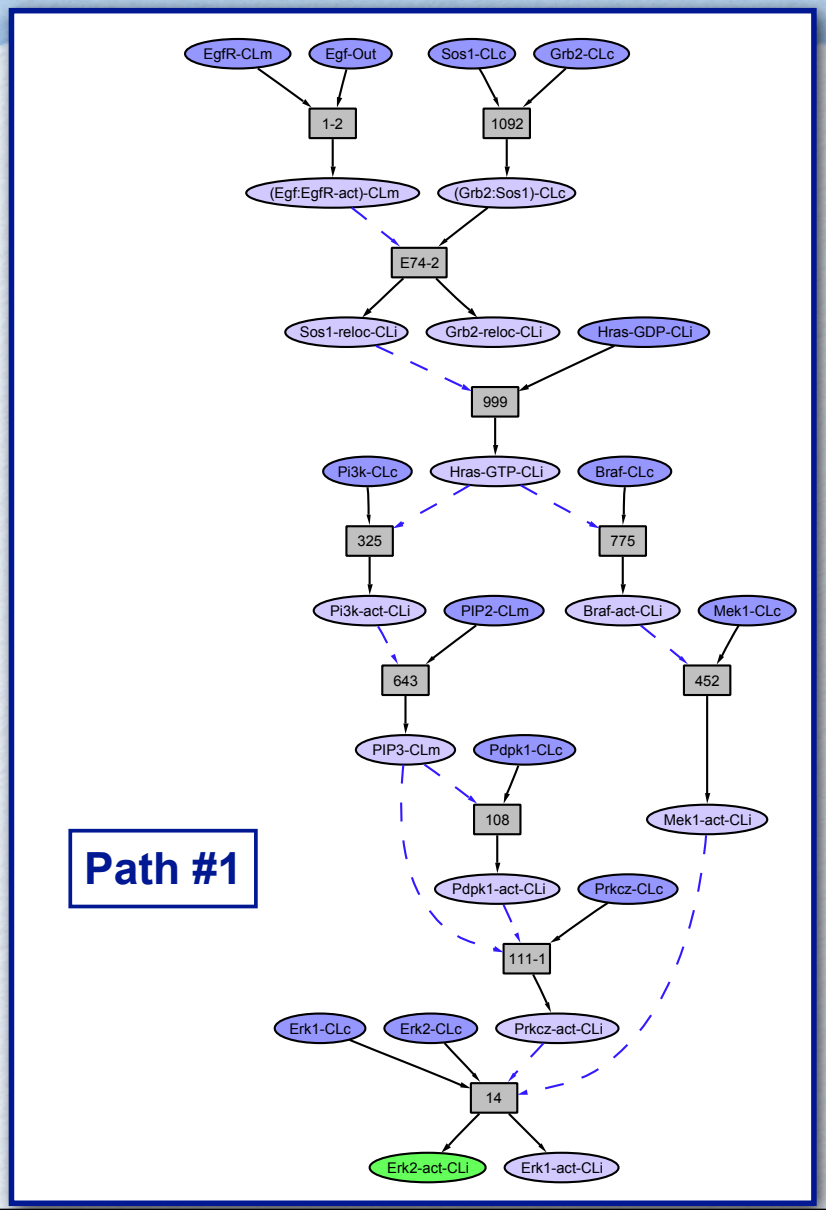
SUBNET RELEVANT TO ERK ACTIVATION

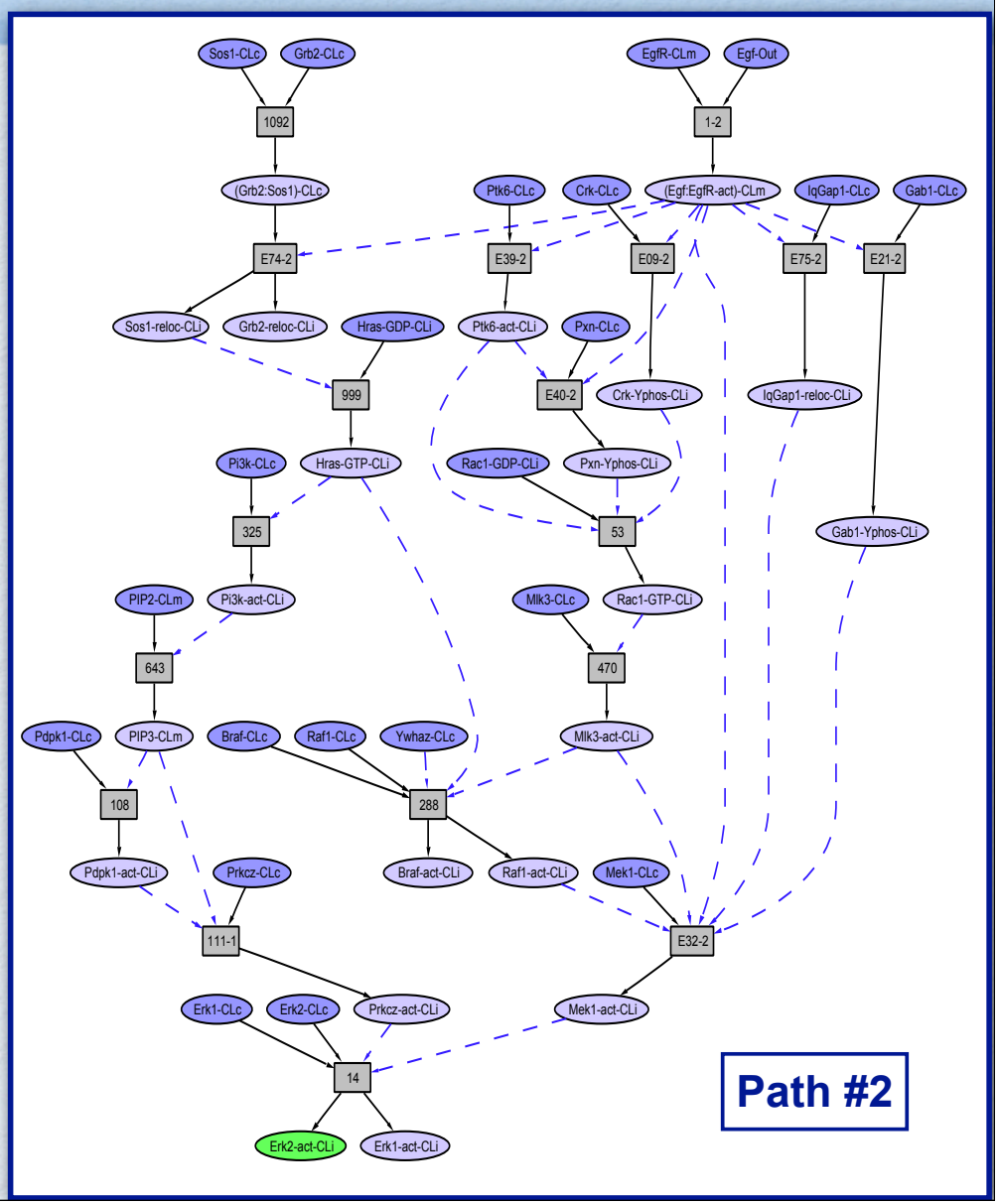
Subnet containing all pathways leading to activation of Erk.

Obtained by backwards followed by forwards collection



POSSIBLE PATHWAYS TO ERK



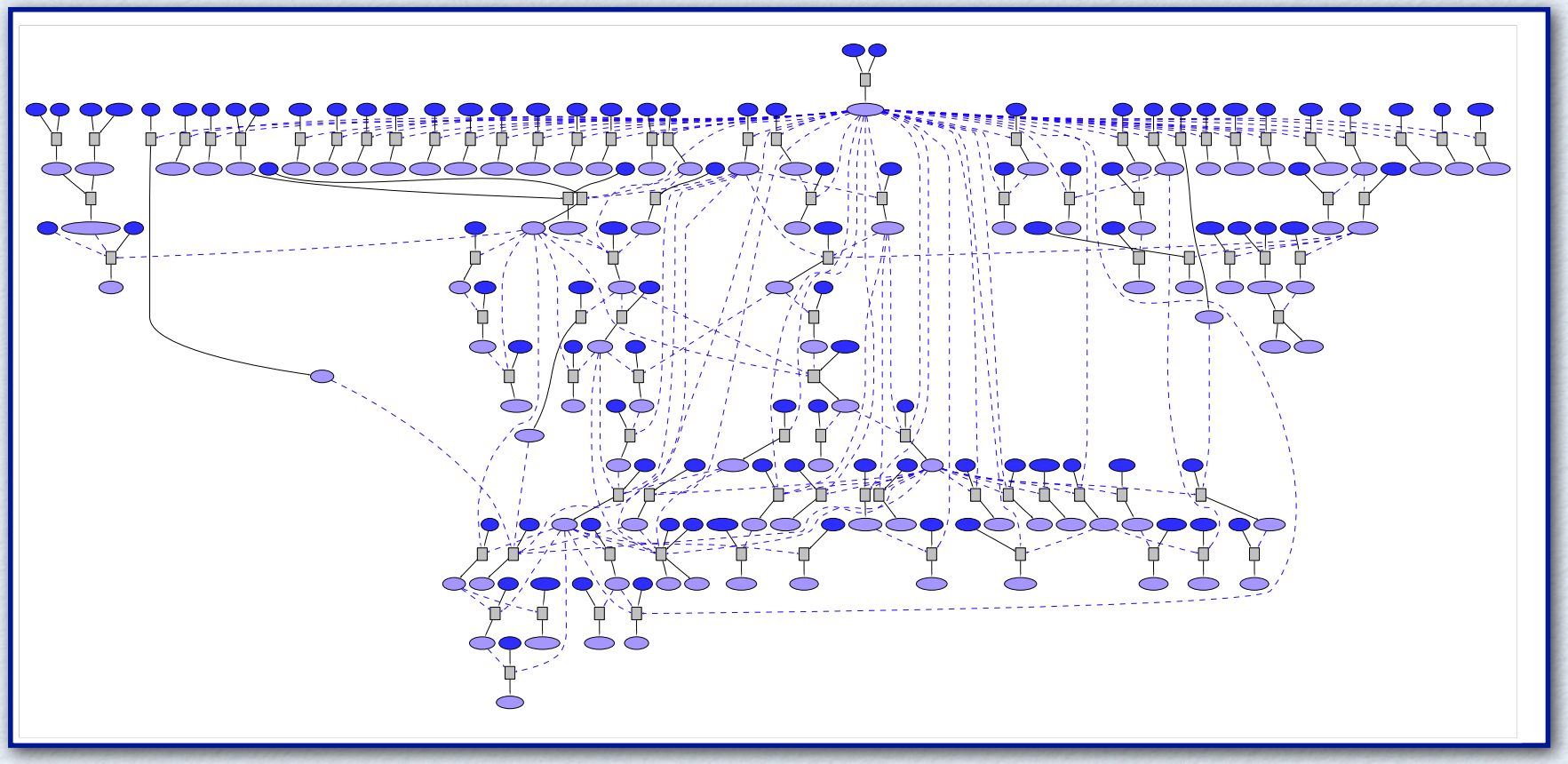


CONSTRAINING THE EGF MAP

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

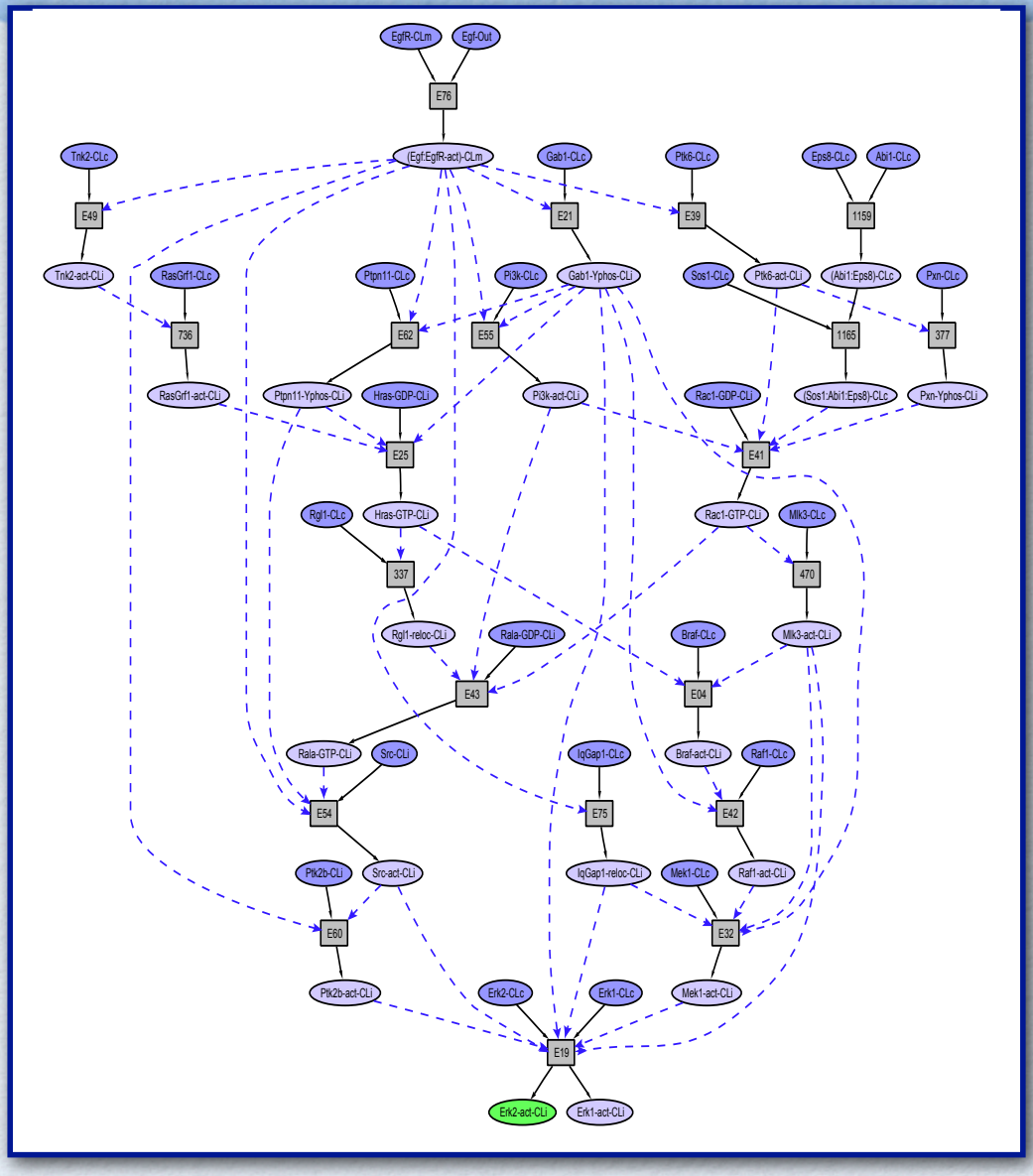
THE CONSTRAINED EGF MAP



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,



SUMMARY

- Pathway logic is a symbolic systems biology framework for modeling networks of reactions / processes
- The pathway logic assistant provides for interactive visualization, navigation, and query of complex networks
- Using the same basic approach we can model metabolic networks
- PL networks can be exported as constraints for alternative analyses such as flux and conflict detection

FUTURE CHALLENGES

- Integration of signaling and metabolic networks
- Modeling action of transcription factors
- Modeling domain-domain interactions
- Adding semiquantitative information
- Algorithms to discover meaningful subnets