PATHWAY LOGIC

APPLICATION OF FORMAL MODELING TECHNIQUES TO UNDERSTANDING BIOLOGICAL SIGNALING PROCESSES

Carolyn Talcott, SRI International

May 2005

PLAN

- **Symbolic** systems biology -- setting context
- Pathway Logic
  - Pathway Logic Models
  - Pathway Logic Assistant
- Other formally based approaches
- Future challenges

SYMBOLIC SYSTEMS BIOLOGY

BIOLOGICAL SYSTEMS

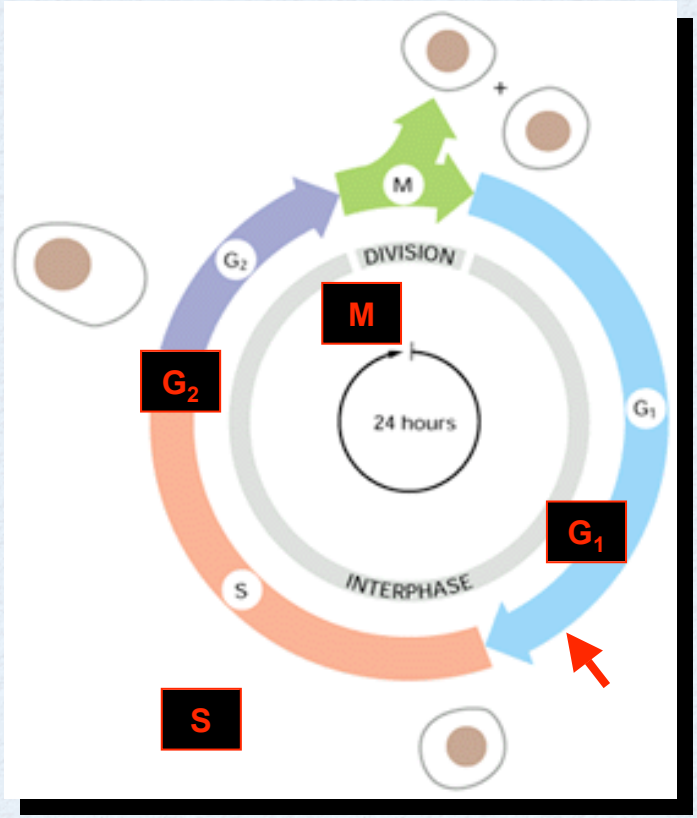
- Biological processes are complex
- Dynamics that range over huge timescales
  - microseconds to years
- Spatial scales over 12 orders of magnitude
  - single protein to cell, cell to whole organism
- Oceans of experimental biological data generated
- Important intuitions captured in mental models that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

UNDERSTANDING HOW CELLS WORK

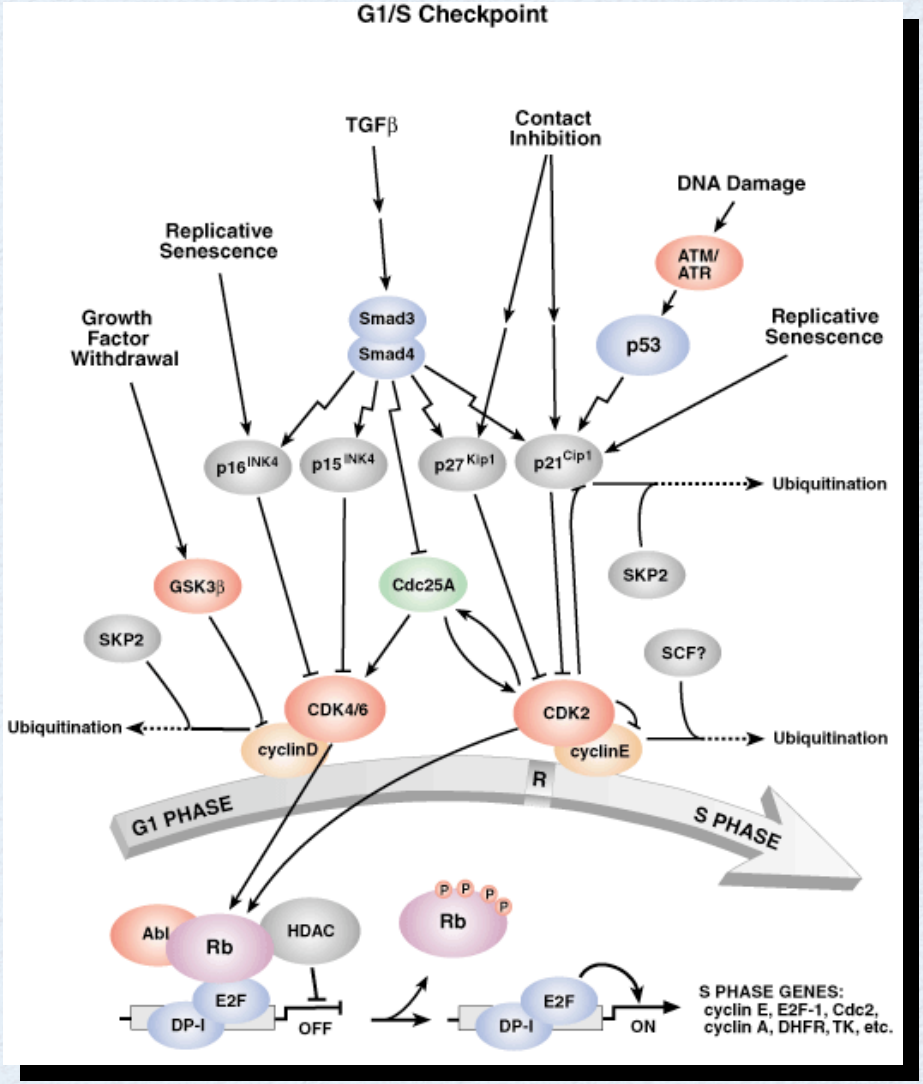
Challenges

- Choosing the right abstractions
  - protein and regulatory networks are large and diverse
  - balance complexity and level of detail
  - move between levels and combine them consistently
- Composing different views or models of different components
  - biological networks combine to produce high levels of physiological organization (e.g., circadian clock subnetworks are integrated with metabolic, survival, and growth subnetworks)

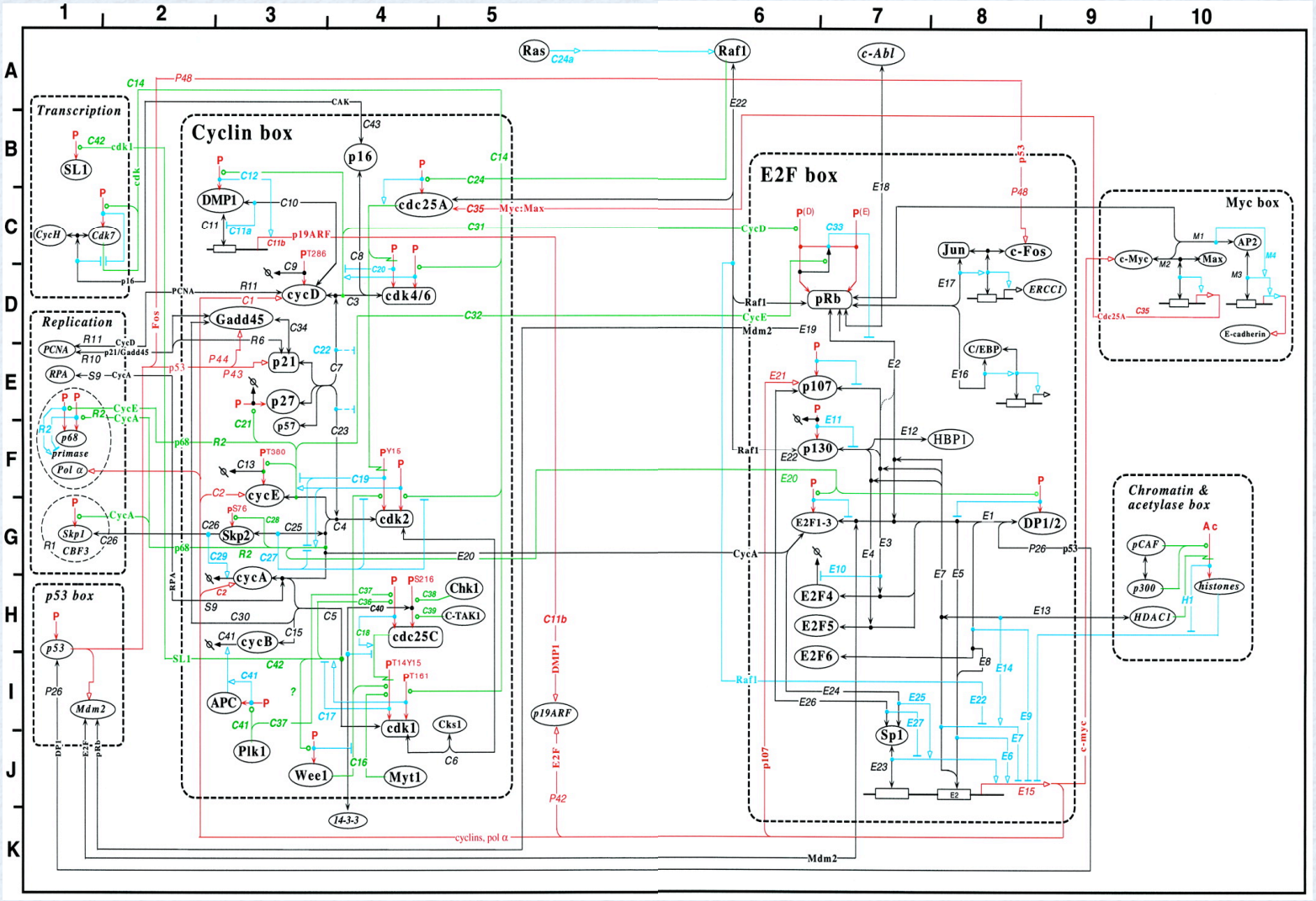
ENTRY TO CELL CYCLE



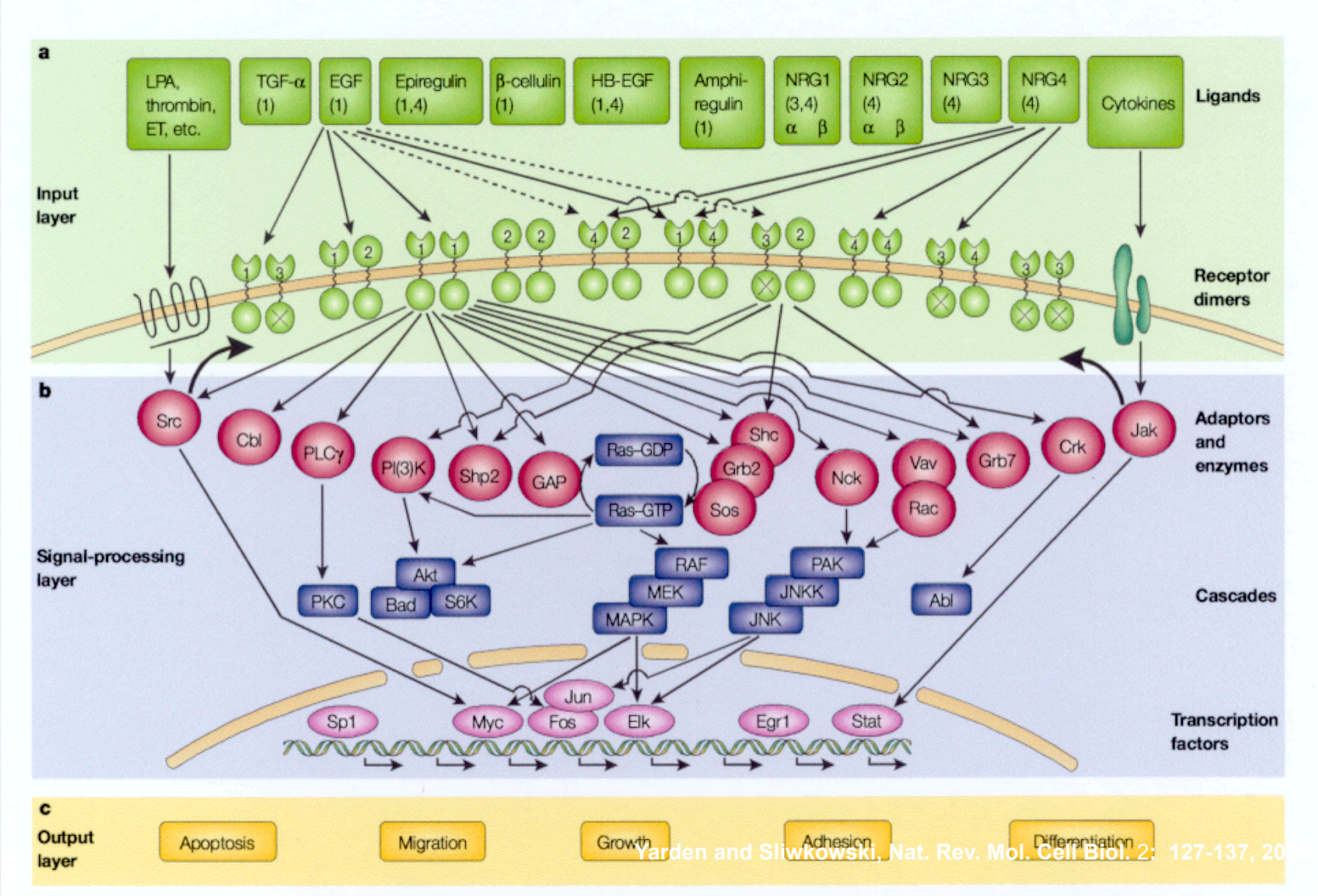
Somatic Cell Cycles Consist of Alternating DNA Synthesis (S) and mitotic (M) Phases, Separated by Gap Phases (G_1 and G_2)



CELL CYCLE CONTROL

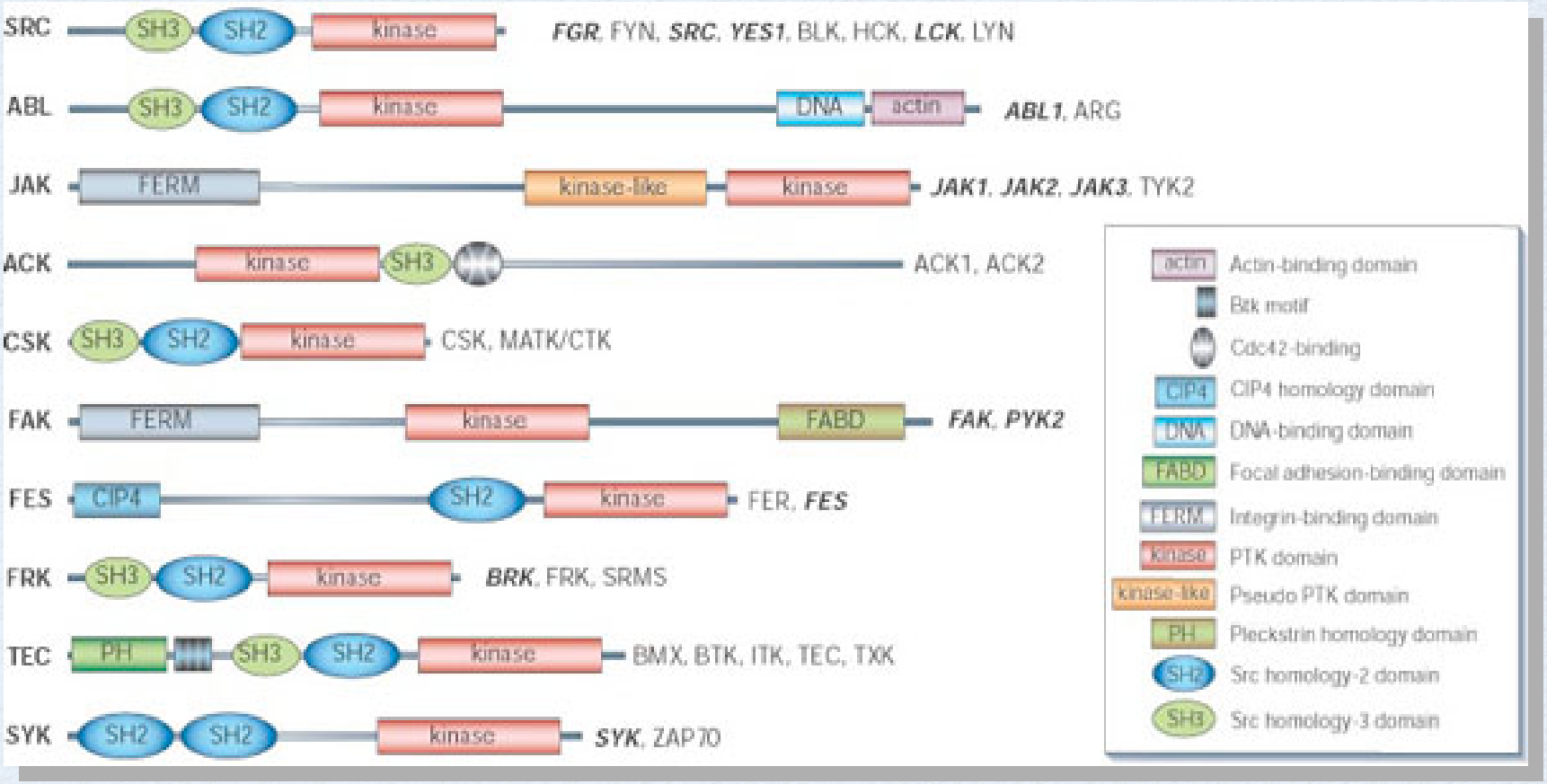


THE ERBB NETWORK



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

PROTEINS IN MORE DETAIL



Signaling Proteins Are Collections of Domains or Modules (PFDs)

(Blume-Jensen and Hunter, Nature 411: 355-65, 2001)

How does the ErbB network work?

- What is the information flow?
- What are the controls?
- Can Jun and Fos be activated?
- What if Raf is blocked?
- How do subnets interact?
- ... a host of more detailed questions ...

Modeling Landscape

- Statistical/probablistic analysis of LARGE data sets.
  - Correlations, dependencies, patterns
- Mathematical models of processes
  - Solving equations (linear, polynomial, differential …)
  - Numerical simulation of individual reactions
- Formal (symbolic/logical) models
  - Aspects of system represented as logical formulas expressing both structure and process
  - Logical inference used to answer queries/make predictions
  - Executable models allow to explore system behavior

Symbolic Systems Biology

The ***qualitative and*** quantitative study of biological processes as ***integrated*** systems rather than as isolated parts

Goals:

- Model causal networks of biomolecular interactions in a logical framework
- Develop formal models that are as close as possible to domain expert's mental models
- Compute with and analyze these complex networks
  - Abstract and refine logical models
  - Simulate or use deduction to check properties
  - Make predictions about possible outcomes, experiment, update model

FORMAL METHODS

- Formal methods = logic based modeling
- Logic = language + models + inference
  - for example arithmetic expressions + Nat + algorithms and princples for proving theorems
- Why inference? Consider the claim

$$(x + y) * (x - y) = x^2 - y^2$$

Check by example:

(5 + 3) * (5 - 3) = 5 * 5 - 3 * 3
8 * 2 = 25 - 9
16 = 16

Symbolic calculation (deduction):

(x + y) * (x - y)
= x * (x - y) + y * (x - y) dist
= x*x - x*y + y*x - y*y dist
= x^2 - y^2 cancel

PATHWAY LOGIC (PL)

ABOUT PATHWAY LOGIC

Pathway Logic (PL) is an approach to modeling biological processes based on **rewriting logic**.

- Using PL signal transduction processes can be modeled at different levels of abstraction involving:
  - the overall state of proteins, or
  - protein functional domains and their interactions
- The resulting signaling networks can be queried using formal methods tools: given an initial state
  - execute --- find some pathway
  - search --- find all pathways
  - model-check --- find a pathway satisfying a temporal formula

PATHWAY LOGIC GOALS

- A formal framework for developing network models that naturally express biologists intuitions.
- Integrate formal methods tools to allow working biologists interact with, query, and modify network models.
- Enable bench researchers to generate informed hypotheses about complex biological networks. For example to investigate questions such as:

"How is the network perturbed when I knockout/in gene X".

"How does the signaling pathway activated by X interact with that activated by Y?"

REWRITING LOGIC

- Designed to specify and reason about concurrent systems
- A specification has two parts
  - A description of the structure of possible system states (as terms in a formal language)
  - Rewrite rules describing how a system might change
    - rules have the form (t => t' if C)
    - rules apply locally and concurrently, modulo equations
- Deduction = computation = rule application (rewriting)

PL REPRESENTATION 1

- Cell components and state -- terms
  - Proteins/modifications -- EGFR, [EGFR - act], [Rac1 - GDP]
  - Cell structure -- membrane, nucleus, endosomes..
- Network of reactions: set of rewrite rules
  - phosphorylation, complex formation, translocation
- System state: PD(soup of cells and ligands)
  - PD for petri dish

PL REPRESENTION II

- Simple reaction rule:

  [protein$_0$ - act] protein$_1$ => [protein$_0$ - act] [protein$_1$ - act]

- Located reaction rule:

  {CLi | ... [protein$_0$ - act] } {CLc | protein$_1$ ... }
  =>
  {CLi | ... [protein$_0$ - act] [protein$_1$ - act] } {CLc | ... }

- Initial state: what proteins are present where
- Pathways: assembled from initial state using rules (by execution, search, model-checking)

SIMPLE RULES

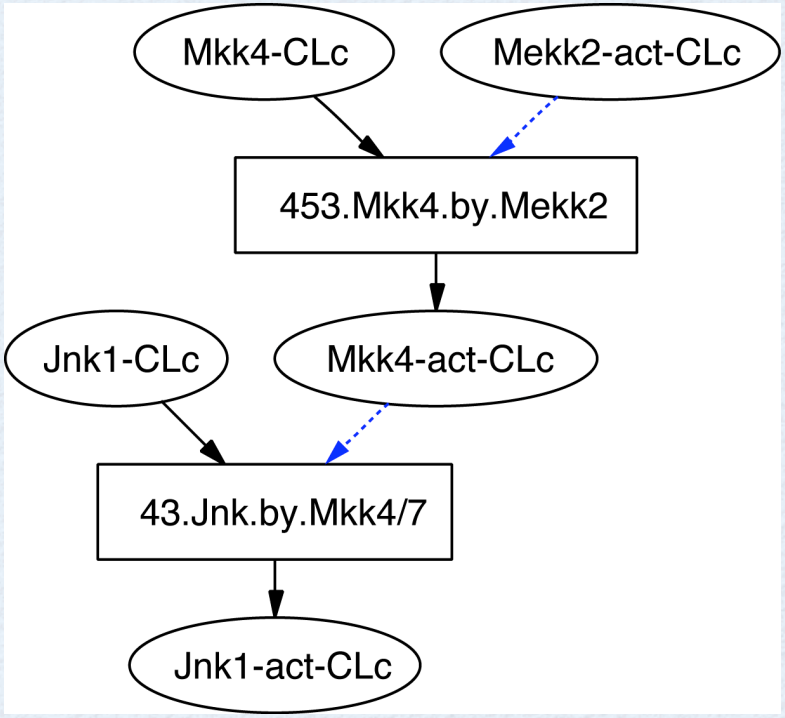
General Rule

```
rl[43.Mkk4/7.->.Jnk]:
   [?Mkk4/7 - act]  ?Jnk
  =>
   [?Mkk4/7 - act] [?Jnk - act] .
```

Specific rule

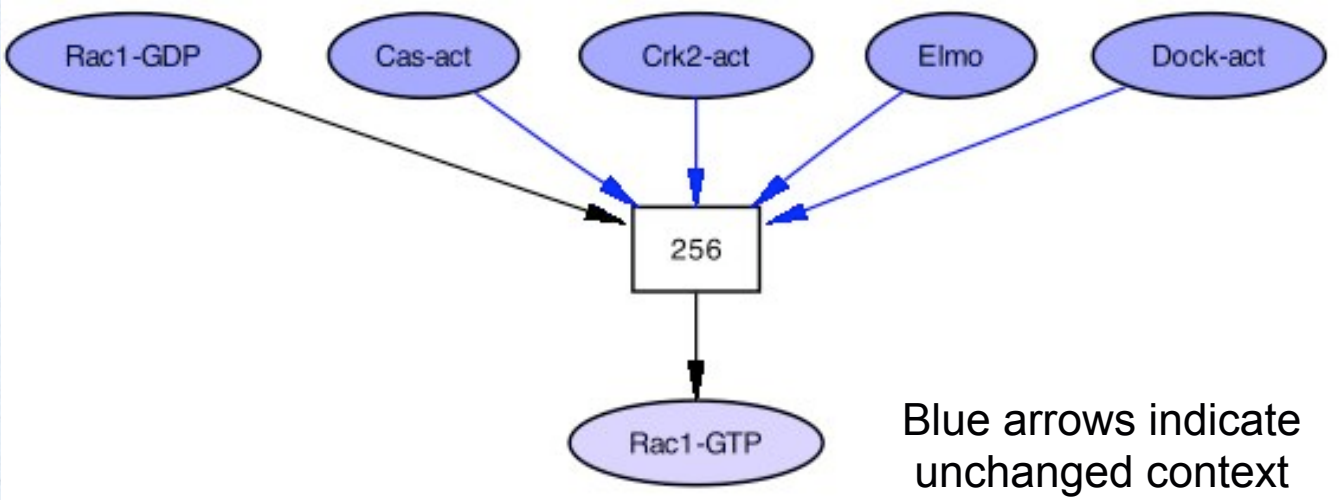
```
?Mkk4/7 := Mkk4,   ?Jnk := Jnk1

rl[43.Mkk4.->.Jnk1]:
  [Mkk4 - act]  Jnk1
 =>
 [Mkk4- act] [Jnk1 - act] .
```



LOCATED RULE

rl[256.Rac1.is.act-3]:
{CLi| cli [Cas - act][Crk2 - act][Dock - act] Elmo [Rac1 - GDP]}
=>
{CLi| cli [Cas - act][Crk2 - act][Dock - act] Elmo [Rac1 - GTP]}



The Pathway Logic Assistant (PLA)

PLA 1

- Provides a means to interact with a PL model
  - Inspect, Modify, Query
- Manages multiple representations
  - Maude module (logical representation)
  - PetriNet (process representation for efficient query)
  - Graph (for interactive visualization)

PLA II

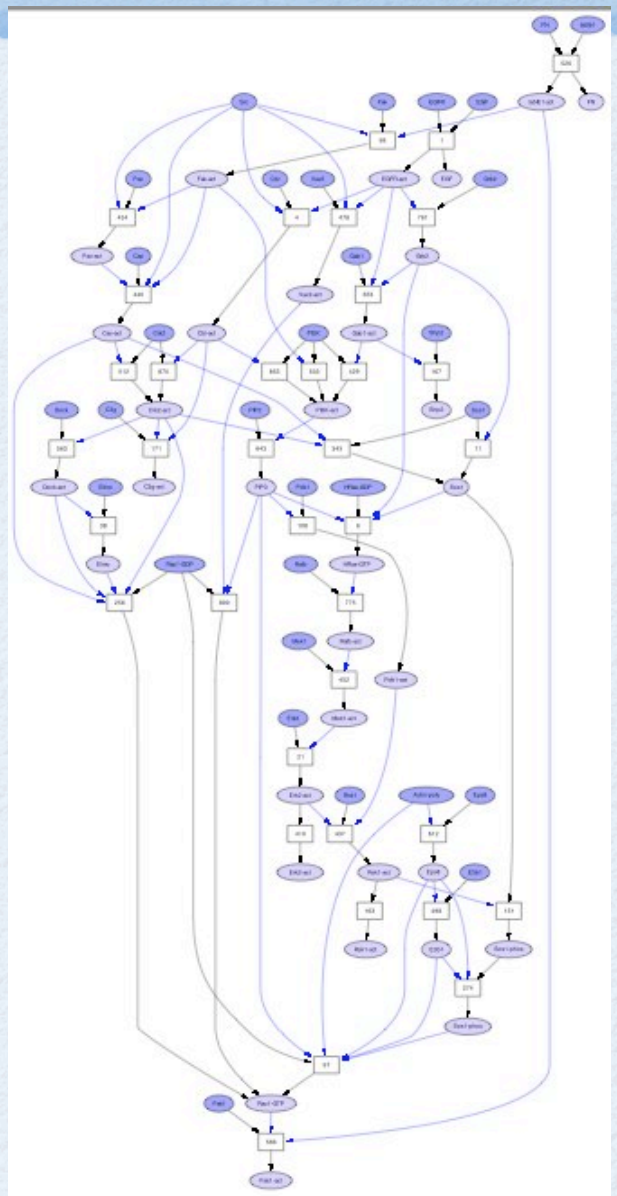
- Exports Representations to other tools
  - Pathalyzer
  - Lola
  - Graphics2d --- interactive visualization
- Imports (some) SBML based models

VISUALIZING A PL MODEL

- Initial state:

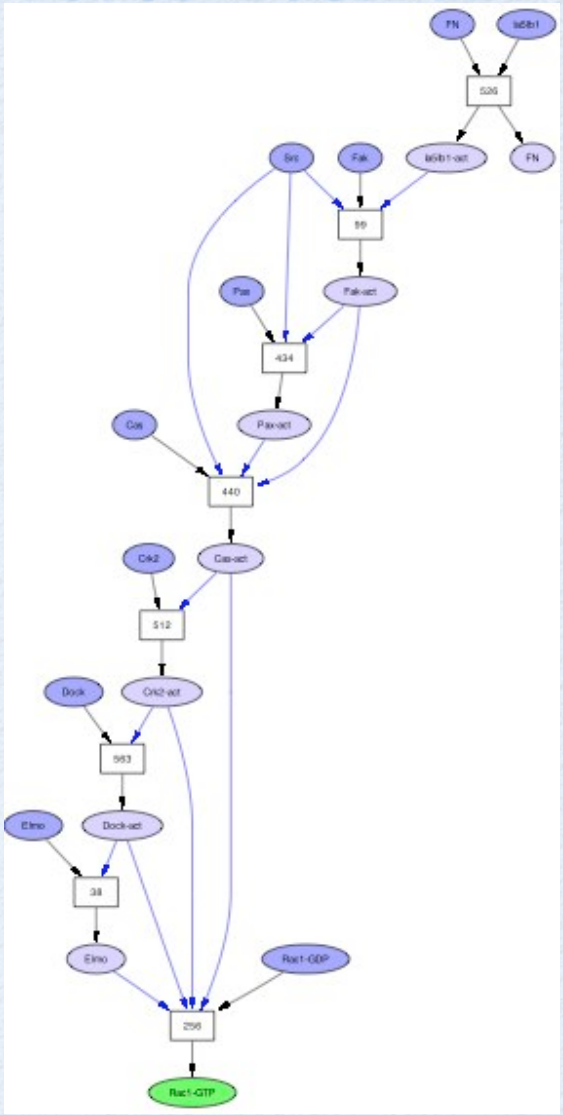
PD(EGF FN [Fibroblast |
{CLm | EgfR Ia5Ib1 PIP2 }
{CLi | [Hras - GTP][Rac1 - GDP] Src}
{CLc | 1433b Cas Cbl Crk2 [Ctak1 - act]
Dock Elmo Erk2 Fak Gab1 Grb2
Ksr1 Mek1 Pak3 Pax PI3K
PP2A Raf1 Sos1 Vav2}])

- Special general rules to obtain relevant reaction net (a Petri net)



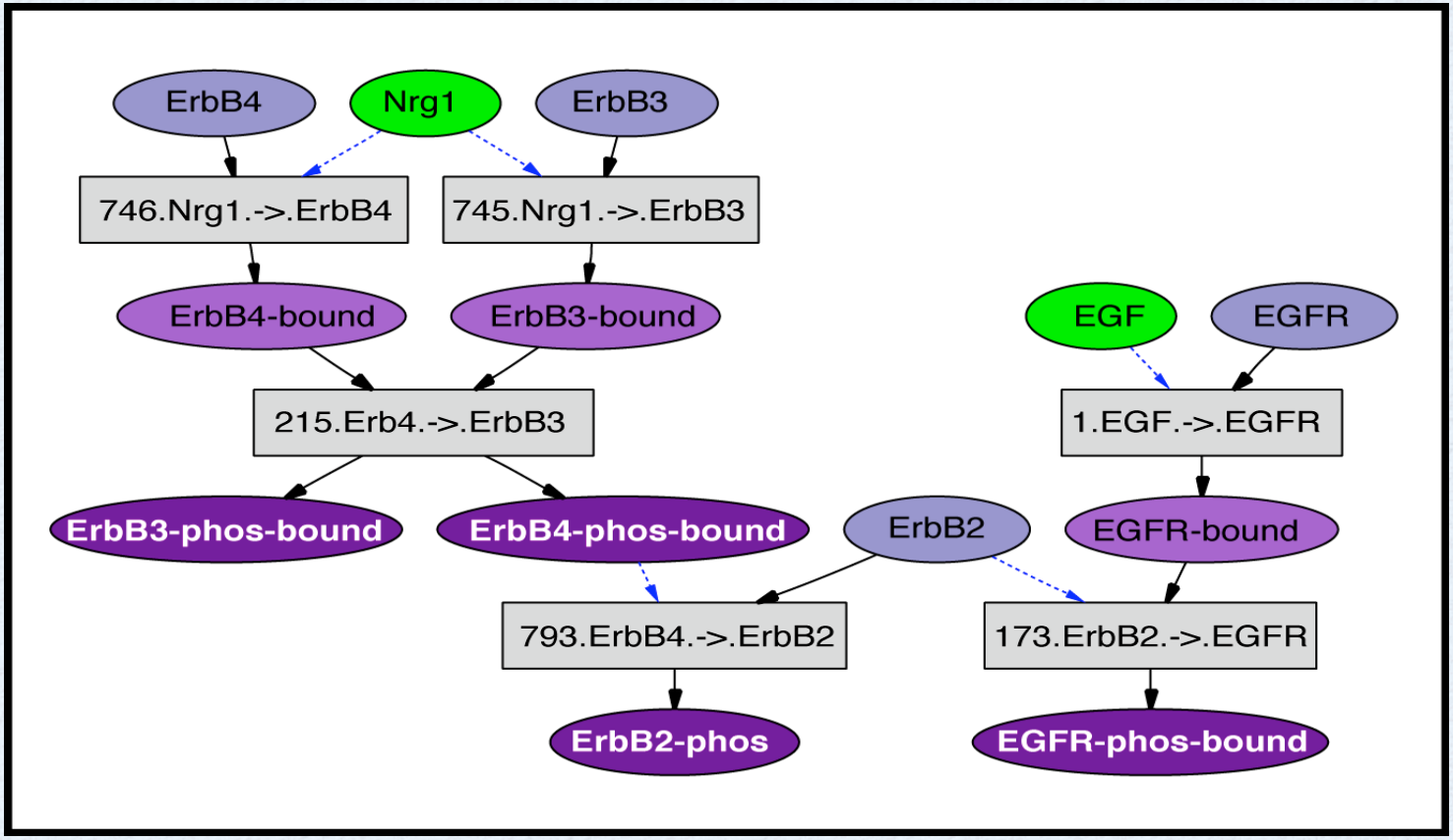
QUERYING A PL MODEL

- Query: can Rac1 be activated without activating EGFR?
  - Goal: Rac1
  - Avoid: EGFR
- Answer, a computation (path) obtained by model-checking
- visualized as a network of rule instances



AN ERBB NETWORK QUERY

Starting with NRG1 and EGF stimuli find a pathway leading to phosphorylation of all four ErbB receptors.



What about expression levels?

OTHER FORMALLY BASED SYSTEMS

EXAMPLE FORMALISMS

- BIOCHAM
  - Chemical abstract machine, constraint logic programs
  - Computation tree logic (CTL)
- Membrane calculi -- spatial process calculi / logics
  - Brane calculus -- mobility of membranes
  - P Systems -- mobility of processes
- Hybrid SAL -- hybrid (discrete + continuous) systems

BIOCHAM

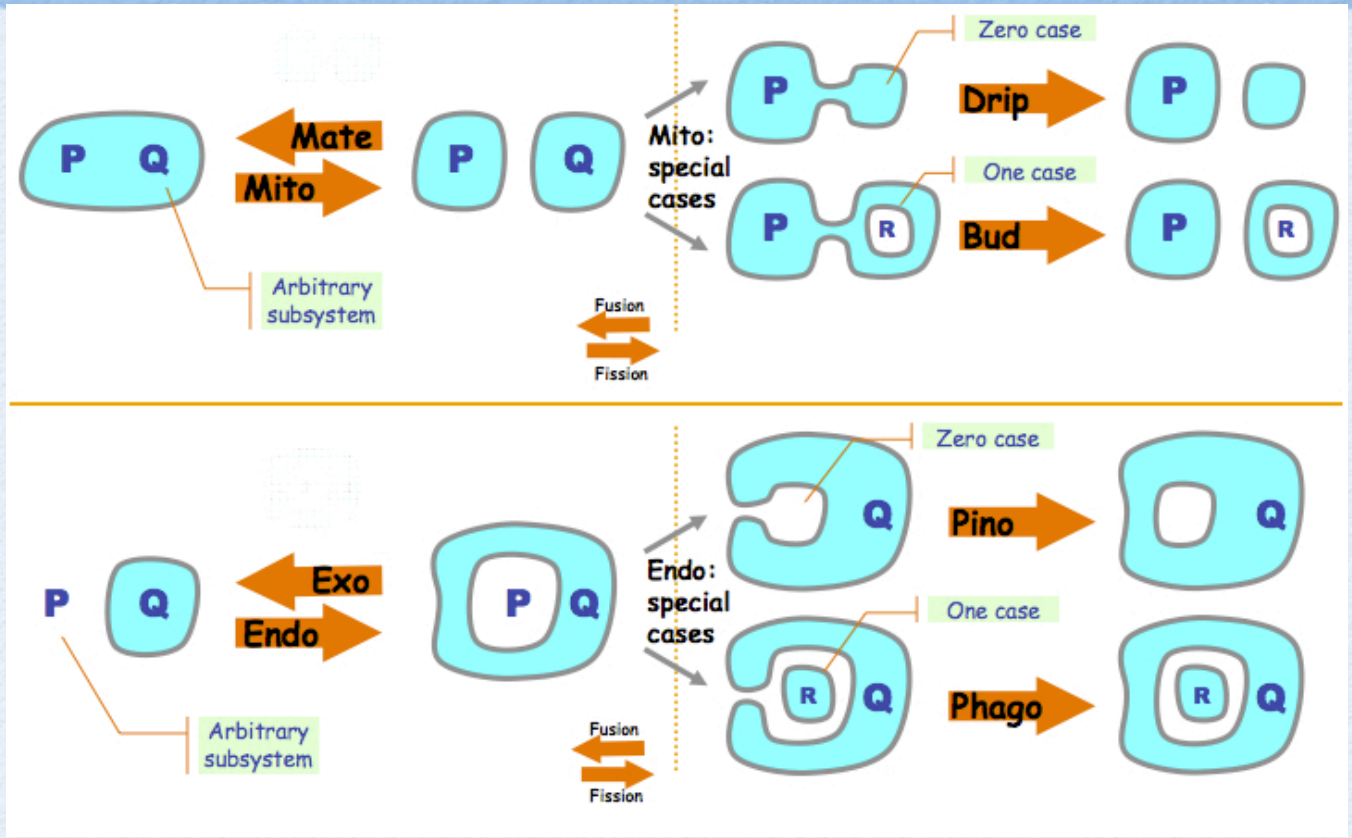
- Modeling language similar to PL
- Multiple interpretations
  - boolean (ala PL)
  - quantitative
  - stochastic
- Richer Query Language

2004 JBPC. Fages, Soliman, Chabrier-Rivier. Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM.

PL VS BIOCHAM QUERIES

- PL Query Language LTL
- Example queries
  - Is there a pathway leading to state s? (Goal s)
  - Is state s2 a necessary checkpoint for reaching state s1? (Goal s1 and avoid s2)
  - Can state s2 be reached after passing through s1
  - Is a certain partial s of the cell a steady state?
- BIOCHAM Query Language CTL can also express queries about cyclic behavior.

MEMBRANE MACHINE



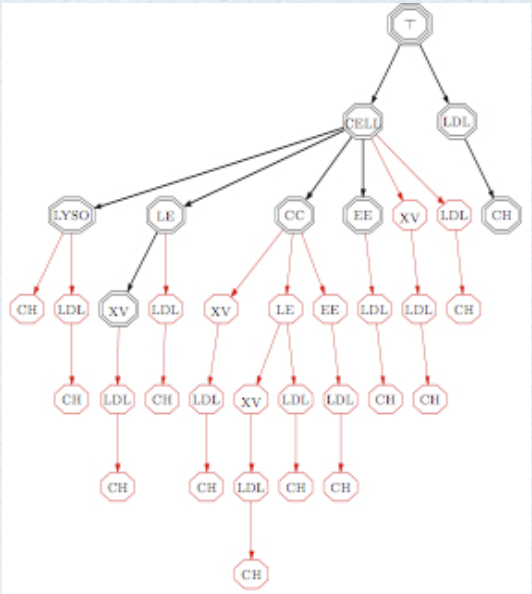
Membrane algorithms

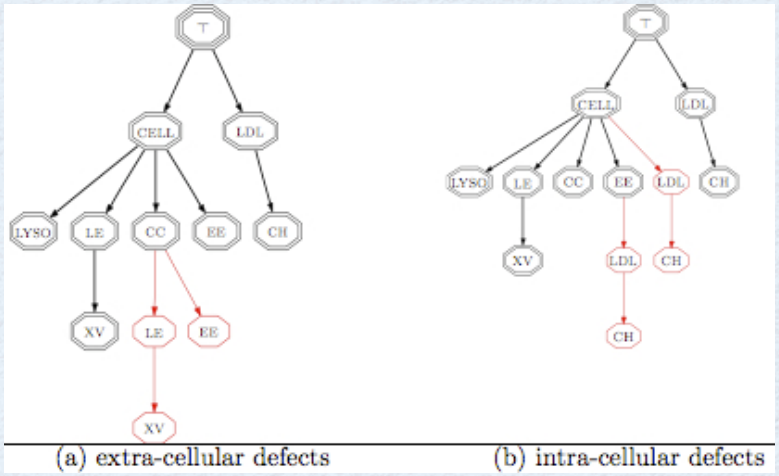
- protein production
- viral reproduction
- LDL cholesterol degration

2005 TCSB. Cardelli. Abstract machines of systems biology

Membrane Analysis

- Control flow analysis of LDL
  - abstract states and rules
  - compute possible nestings
    - finds all possible nestings,
    - may find false nestings
    - precision can be tuned
- Example: find where cholesterol may be released in presence of defects

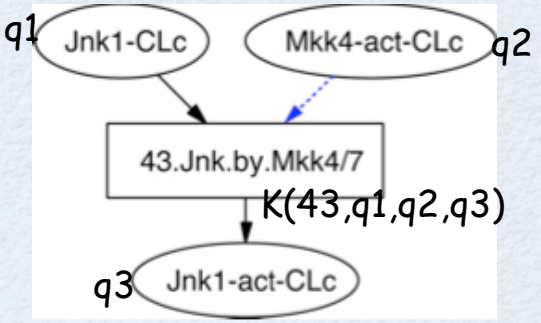




2005 CMSB. Pilegaard, Nielson, & Nielson. Static Analysis of a Model of the LDL Degradation Pathway

PSYSTEMS

- Elaborates PL models with quantity and rate information
- Can reason about feedback loops, effects of concentration change.
- Simulation using CLIPS
- Applied to study of EGF-ERK signalling
- Problem: numbers come from multiple cell-types and conditions



2005 CMSB. Perez-Jimenez, Romero-Campero
Modelling EGFR signalling cascade using continuous membrane systems.

HYBRID SAL

- Abstract given quantitative model into discrete regions, based on
  - Input model
  - Property of interest
  - First and higher derivatives
- Apply variety of model checking tools
- Concretize output back to biological quantitative domain

HYBRID SAL APPLICATION

- Advantage: handles
  - uncertainty
  - multiple scales
  - quantitative/qualitative

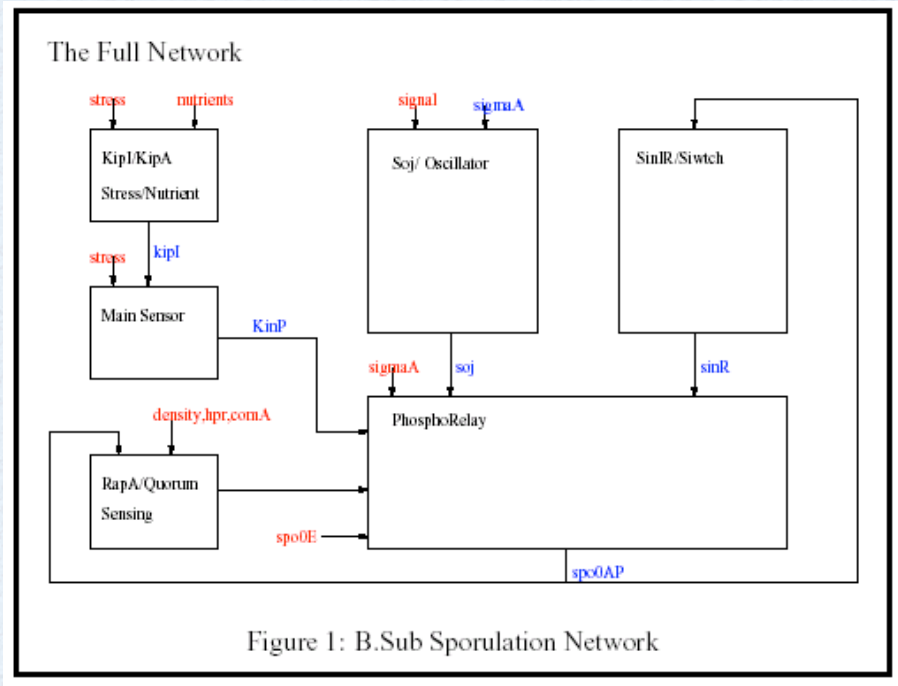


Figure 1: B.Sub Sporulation Network

- Case studies
  - Delta Notch
  - B Subtilis Sporulation
  - Insulin/glucose metabolism (whole organism)

FUTURE CHALLENGES I

- Scale to bigger models
  - optimize Petri net generation
  - property preserving abstractions
  - hierarchical networks
- Richer model
  - express semi-quantitative information
    - n-fold up/down regulation
  - integrate kinetic information
  - more detailed representation of interactions

FUTURE CHALLENGES II

- Integration of models
  - quantitative and qualitative
  - time scales
  - spatial scales
- Understanding data from diverse sources
  - genomic/proteomic
- Integration of skills
  - computational/logical models informed by biological intuitions
  - biological intuitions informed by computational/logical models

PATHWAY LOGIC TEAM

- Keith Laderoute
- Patrick Lincoln
- Carolyn Talcott

- Linda Briesemeister
- Steven Eker
- Merrill Knapp
- Andy Poggio

- Biology Computer Science