Pathway Logic:
Executable Models
of Biological Networks

Steven Eker, Merrill Knapp,

Keith Laderoute, Patrick Lincoln,

Carolyn Talcott

SRI International

Cellular Biology

● Good news:

vast amounts of experimental data
- genomic sequence information
- analysis of global gene expression

● Challenge: new theoretical models
- compute with and manipulate
- better understanding of cellular processes
- guide hypothesis creation and testing

Cells are complex machines

● Metabolic processes -- reactions using up some molecules and producing others ultimately resulting in the `expressed' proteins

● Signalling processes -- mechanisms by which cells detect, convert, and internally transmit information from their environment to intracellular targets

● Described in terms of networks of `rules' for
  – chemical reactions
  – signal transfer (physical / conformational changes)

● Regulatory processes -- form decision circuits

Traditional modelling

● Computational analysis of cellular metabolic and signaling networks based on rate and concentration data.

● Used for
  – predicting transcription and gene expression
  – understanding of regulation mechanisms, including
    * feedback mechanisms
    * multistable switches

● Involves complex comptations that require detailed quantitative data

Some Current Biological Models

● Numerical simulation using rates and concentrations

● Study of feedback and switch behavior
  - Hybrid logical/kinetic
  - Control theory
  - Hybrid system abstraction / model-checking

● Metabolic pathways using symbolic differential equations and flux-based analysis

● Pathway prediction by analogy from genome databases (EcoCyc, KEGG, WIT ...)

Problems

Modeling and analysis based on rate and concentration information is crucial for detailed understanding of cellular processes, but ...

● it is difficult to obtain needed experimental data

● the models are difficult not only to simulate, but also to build and to understand intuitively

● cellular scale populations of key molecules leads to stochastic behavior

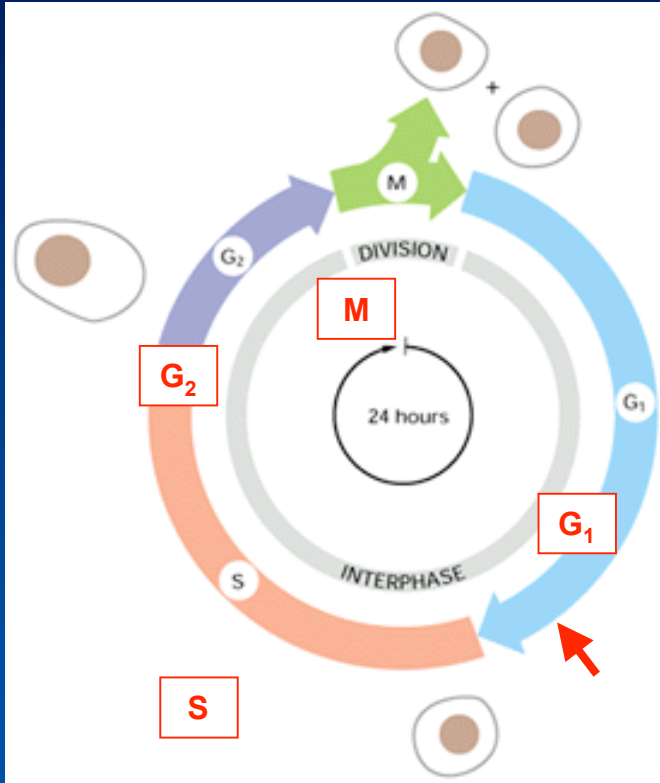# Symbolic Systems Biology

- **What**

  - Abstract qualitative models of metabolic and signaling processes

  - Formal methods tools to study a wide range of questions
    - ▽ Static analysis
    - ▽ Forward/backward simulation/search
    - ▽ Abstraction and model checking
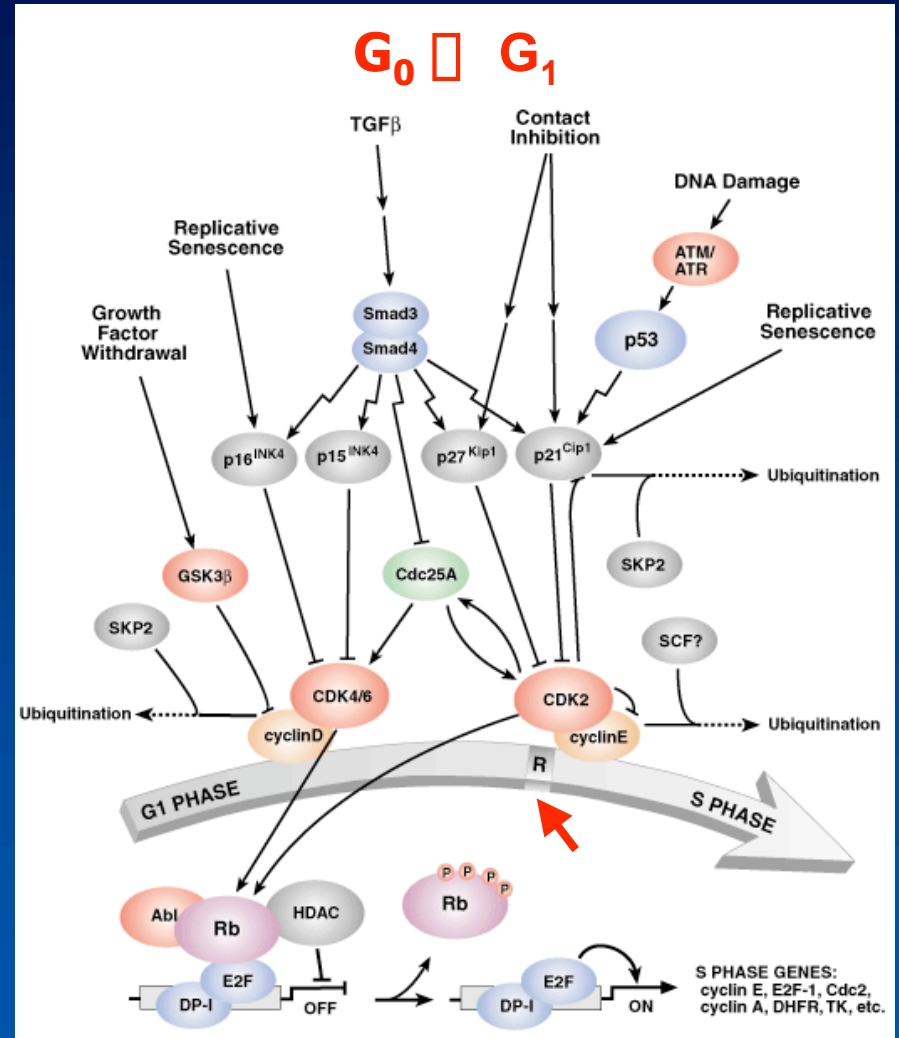   - ▽ Meta analysis

- **Vision**

  - New understanding of complex biological systems

  - New ways to generate hypotheses for testing experimentally
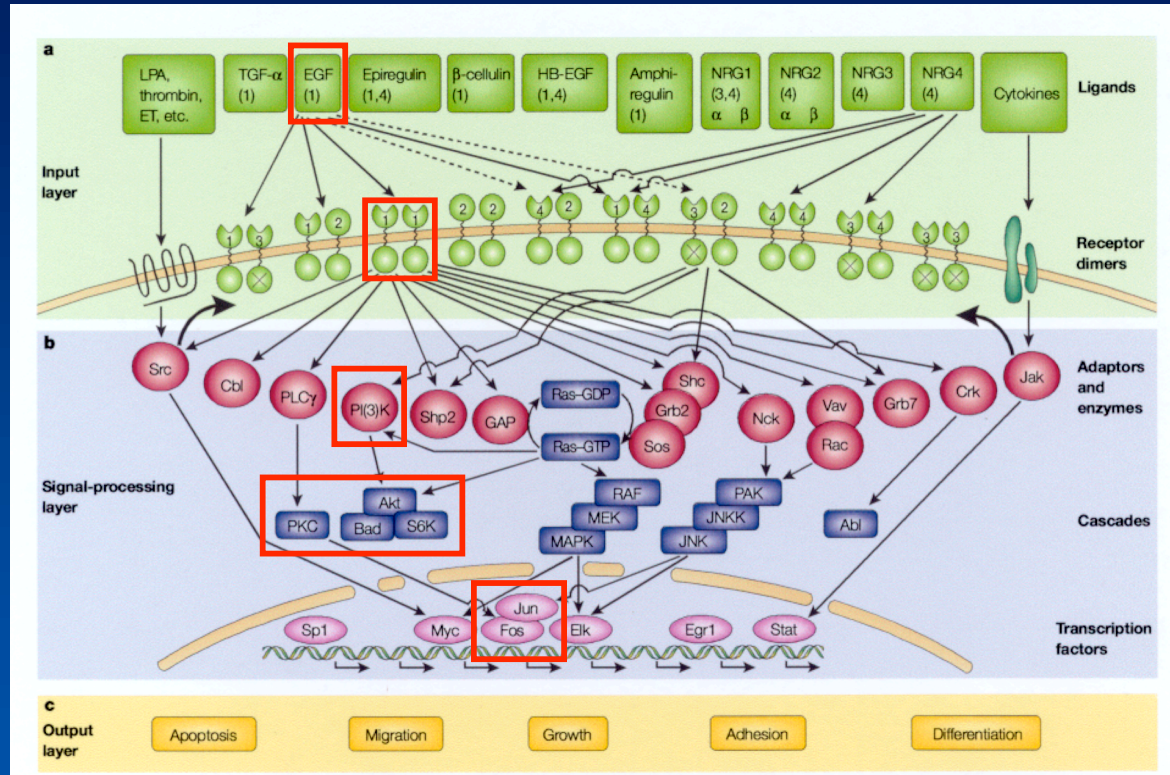
Mamallian cell

Top level: Entry Into the Cell Cycle

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

The ErbB Network

EPL in Maude signature: basics

sorts Protein Chemical Thing .

subsorts Protein Chemical < Thing .

ops EGFR EGF Pdk1 PKCe : -> Protein .

ops Ca++ PIP3 : -> Chemical .

sorts Modification ModSet .

subsort Modification < ModSet .

ops GDP GTP act deact : -> Modification .

op none : -> ModSet .

op _ _ : ModSet ModSet -> ModSet [assoc comm id: none] .

op [_-_] : Protein ModSet -> Protein [right id: none] .

EPL in Maude signature: structures

sort Complex .

subsort Complex < Thing .

op _:_ : Thing Thing -> Complex [comm] .

sorts Soup Enclosure MemType.

subsort Thing < Soup .

op empty : Soup .
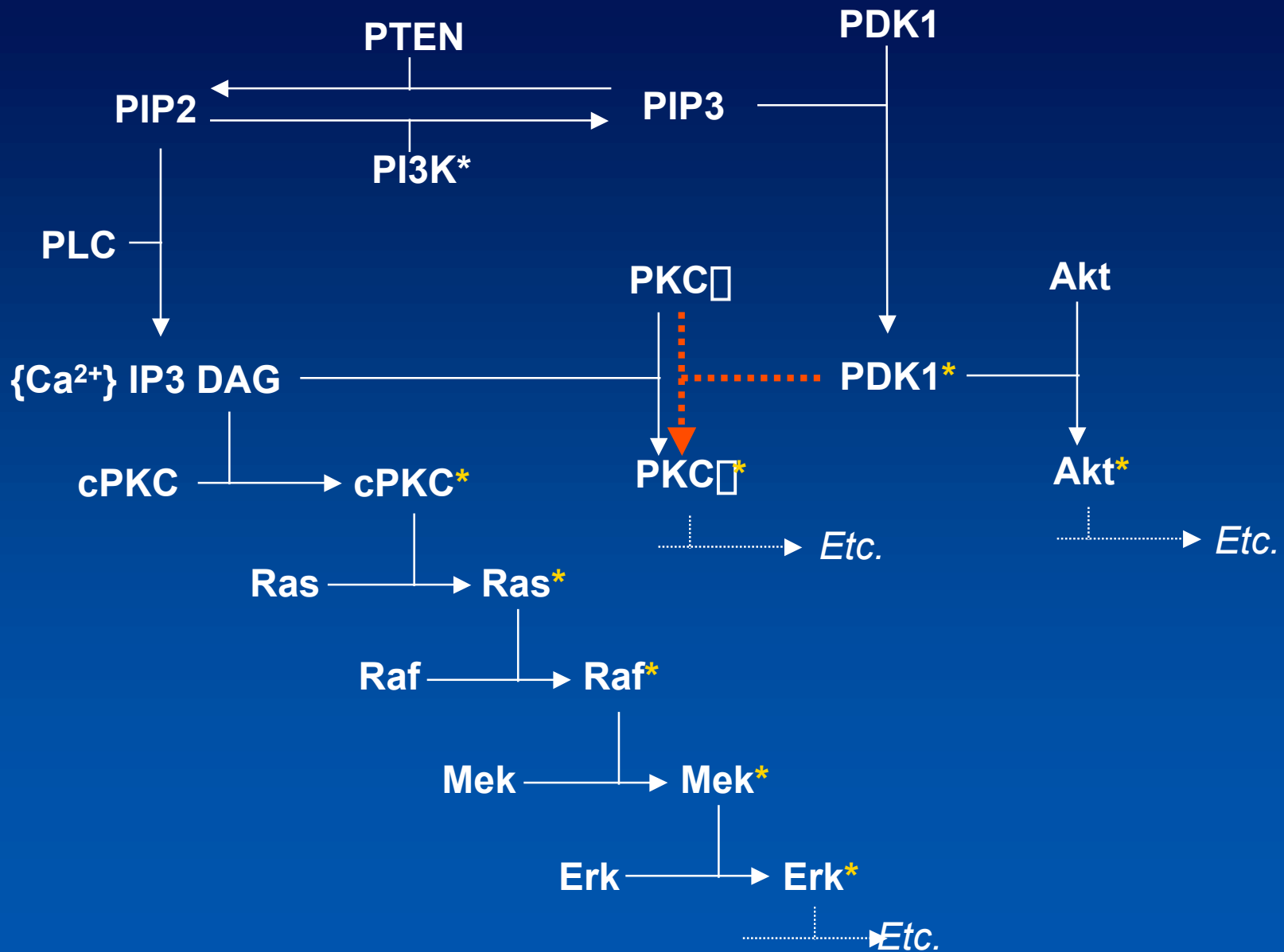
op _ _ : Soup Soup -> Soup [assoc comm id: empty] .

ops CM NM : -> MemType .

op {_|_{_}} : MemType Soup Soup -> Enclosure .

PKC Regulatory Network

PKC Rules in Maude

```
rl [643.PI3KI.act.PIP3]:
   PIP2 [?PI3KI:PI3KI - act]
   =>
   PIP3 [?PI3KI:PI3KI - act]


rl [84.PLC.act.DAG]:
   Ca++ {CM| cm:Soup PIP2 [?PLC:PLC - act]
               {cyto:Soup}}
   =>
   {CM|cm:Soup (Ca++ DAG IP3) [?PLC:PLC - act]
       {cyto:Soup}} .
```

Execution

```
eq q14 =
    PD(Ca++ {CM | PIP2 [PI3Ka - act]
                    [PLCb1 - act] [Pten - act]
                    {Erk1 Pdk1 PKCa   PKCe
                     {NM | empty {empty }}}}) .
rewrite q14 .
result Dish:
    PD({CM | Ca++ DAG IP3 [PI3Ka - act]
                [PLCb1 - act] [Pten - act]
                {Erk1 Pdk1 [PKCa - act] [PKCe - act]
                 {NM | empty {empty }}}})
frewrite q14 .
result Dish:
    PD({CM | Ca++ DAG IP3 [PI3Ka - act]
                [PLCb1 - act] [Pten - act] [Pdk1 - act]
                {Erk1 [PKCa - act] [PKCe - act]
                 {NM | empty {empty }}}})
```

Search

```
search q14 =>! D:Dish .
Finds 3 States:

Solution 1 (State 6) is the rewrite result

Solution 2 (State 20) is the frewrite result.

Solution 3 (state 23) is
  PD({CM | Ca++ DAG IP3 [PI3Ka - act]
          [PLCb1 - act] [Pten - act]
          [Pdk1 - act][PKCe - act]
          {[PKCa - act]
           {NM | empty {[Erk1 - act]}}}})
```

PKC (q14) Search graph

EGF - EGFR Network

EGF/EGFR experiments

Activation of a transcription factor (cJun cFos) following binding of extracellular epidermal growth factor (EGF) to its receptor (EGFR)

```
ops q1 q1x : -> Dish .
eq q1 = PD(EGF {CM | EGFR Pak1 PIP2 nWasp [H-Ras - GDP]
                    {Akt1  Gab1 Grb2 Gsk3 Eps8 Erk1
                     Mek1 Mekk1 Mekk4 Mkk4 Mkk3 Mlk2
                     Jnk1 p38a p70s6k Pdk1 PI3Ka PKCb1
                     Raf1 Rsk1 Shc Sos [Cdc42 - GDP]
                     {NM | empty {cJun cFos }}}}) .


eq q1x = q1 - < PI3Ka , cyto >
```

Model Checking

```
subsort  Dish < State .
eq PD(out:Soup
        {CM | cm:Soup
              {cyto:Soup
                {NM | nm:Soup
                        {nuc:Soup
                          [cJun - act] [cFos - act]}}}})
      |= prop1 = true .

eq findPath(S:State,P:Prop) = getPath(P:Prop,
   S:State |= ~ <> P:Prop) .

red  findPath(q1,prop1) .
red  findPath(q1x,prop1) .
```

Model Checking q1.m -- simple paths

- Maude> red findPath(q1,prop1) .

- result SimplePath: spath('1.EGFLike.binds.EGFR '353M.EGFR.act.Gab1 '418B.EGFR.act.Cdc42 '349.Cdc42.act.Mekk1 '471.Cdc42.act.Mekk4 '470.Cdc42.act.Mlk '3.EGFR.act.Shc '429.Gab1.act.PI3K '498.PI3K.act.Eps8 '411.Shc.act.Grb2 '324.Shc.Grb2.act.Sos '618.Cdc42.act.Wasp '6.Sos.act.Ras '280.Ras.act.Raf '274.Raf1.act.Mek1/2 '643.PI3KI.mets.PIP2.to.PIP3 '108.PIP3.act.Pdk1 '109.PIP3.Pdk1.act.Akt '122.Akt.deact.Gsk3   '121.Akt.to.nuc '21.Mek1.act.Erk1/2 '437.Erk.act.Rsk '410.Erk1/2.to.nuc '163.Rsk.to.nuc '36.Mekk1.act.Mkk4/7 '457.Mekk4.act.Mkk3/6 '61.Mkk3/6.act.p38 '450.p38.to.nuc. '43.Mkk4/7.act.Jnk '49.Jnk.to.nuc '50B.Jnk.Gsk3.act.cJun '162.Erk.Rsk.act.cFos,

- PD(EGF {CM | Grb2 Pak1 PIP3 PKCb1 [Cdc42 - gtp] [EGFR - act] [Eps8 - act] [Gab1 - act] [Pdk1 - act] [PI3Ka - act] [Raf1 - act] [H-Ras - gtp] [Shc - act] [Sos - act] [nWasp -   act]{p70s6k [Gsk3 - deact] [Mek1 - act] [Mekk1 - act] [Mekk4 - act] [Mkk3 - act] [Mkk4 - act] [Mlk2 - act] {NM | empty{[Akt1 - act] [cFos - act] [cJun - act] [Erk1 - act] [Jnk1 - act] [p38a - act] [Rsk1 - act]}}}}))

- Maude> red findPath(q1x,prop1) .

- result SimplePath: spath('1.EGFLike.act.EGFR '353M.EGFR.act.Gab1 '418B.EGFR.act.Cdc42 '349.Cdc42.act.Mekk1 '471.Cdc42.act.Mekk4 '470.Cdc42.act.Mlk '3.EGFR.act.Shc '411.Shc.act.Grb2 '324.Shc.Grb2.act.Sos '618.Cdc42.act.Wasp '6.Sos.act.Ras '280.Ras.act.Raf '274.Raf1.act.Mek1/2 '21.Mek1/2.act.Erk1/2 '437.Erk.act.Rsk '410.Erk1/2.to.nuc '36.Mekk1.act.Mkk4/7 '457.Mekk4.act.Mkk3/6 '61.Mkk3/6.act.p38 '450.p38.to.nuc. '43.Mkk4/7.act.Jnk '49.Jnk.to.nuc '149.Rsk.iphos.Gsk3 '50B.Jnk.Gsk3.act.cJun '163.Rsk.to.nuc '162.Erk.Rsk.act.cFos,

- PD(EGF {CM | Grb2 Pak1 PIP2 [Cdc42 - GTP] [EGFR - act] [Gab1 - act] [Raf1 - act] [H-Ras- GTP] [Shc - act] [Sos - act] [nWasp - act] {Akt1 Eps8 Pdk1 p70s6k PKCb1 [Gsk3 - deact] [Mek1 - act] [Mekk1 - act] [Mekk4 - act] [Mkk3 - act] [Mkk4 - act] [Mlk2 - act] {NM | empty {[cFos - act] [cJun - act] [Erk1 - act] [Jnk1 - act] [p38a - act] [Rsk1 - act]}}}}))

Roadmaps for q1 and q1x

Meta Analysis

Metadata associated with the EPL model

- **rule justification**
  - paper citation
  - database entry
  - medline citation
- **ordering/strategy information**
  - activation rules before translocation rules
- **constraints on rule structure**
  - biological type checking
- **interpretation of rule labels**
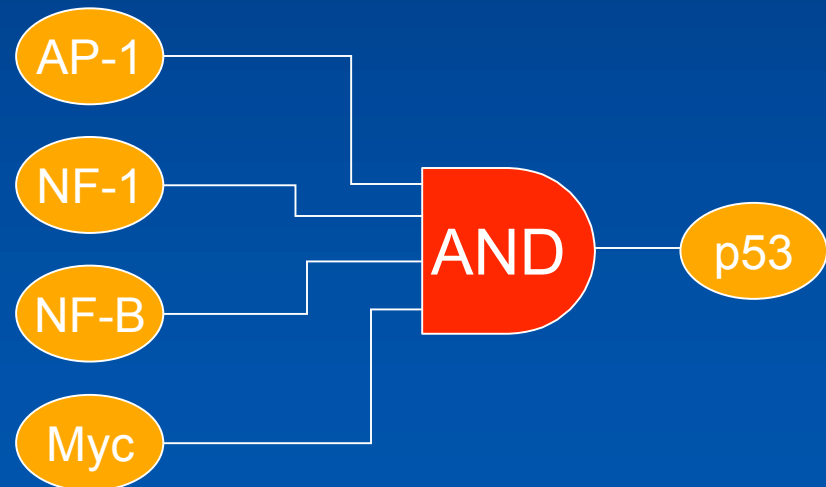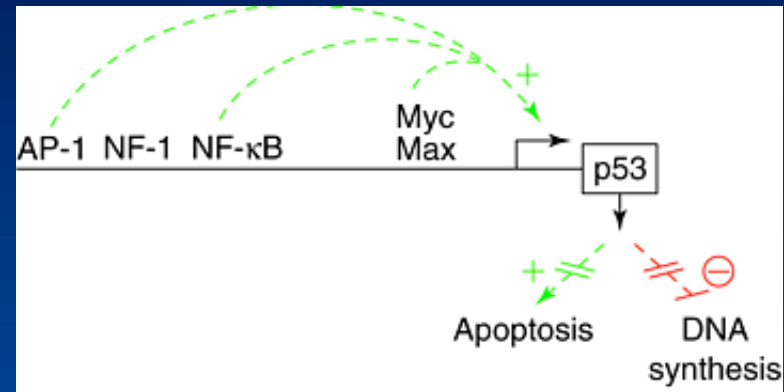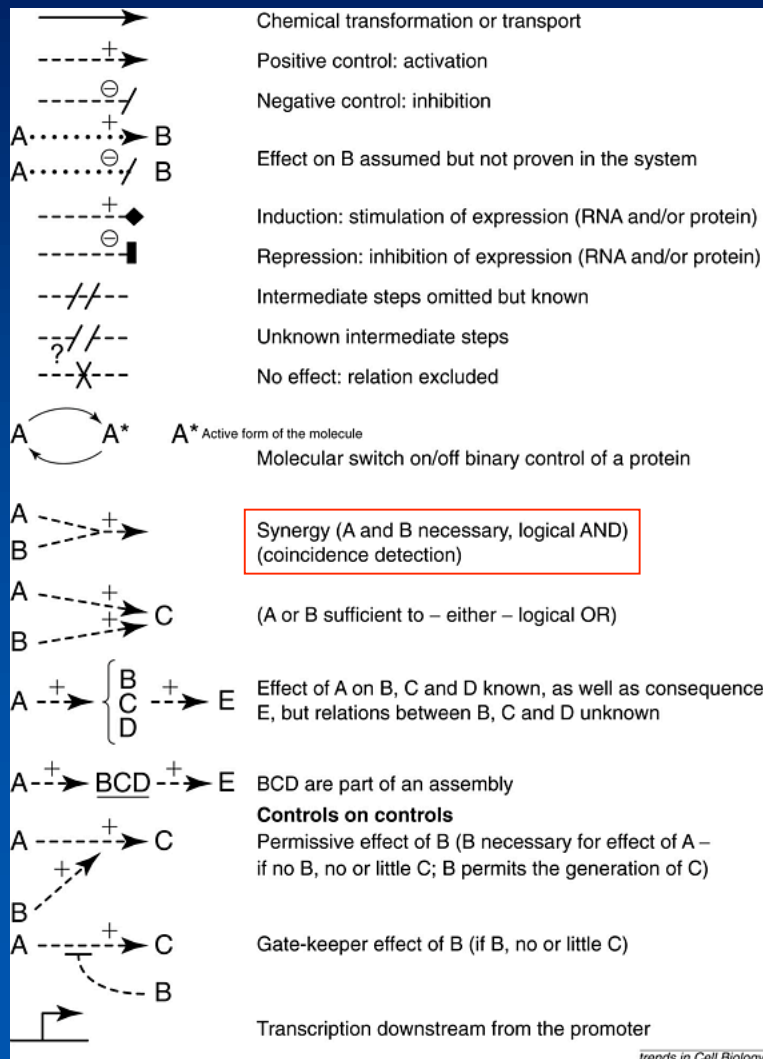
Model Analysis

● What are all the rule labels?

● What constants of sort Protein have been declared?

● What rules activate Erk1?

Model Transformation

● findPath with modified

- ruleset

- dish

- property

● Petri net representation

Visualizing
Model Elements
and
Analysis Results

Symbolic Representation of Signal Transduction: p53 Induction

Chemical transformation or transport

Positive control: activation

Negative control: inhibition

Effect on B assumed but not proven in the system

Induction: stimulation of expression (RNA and/or protein)

Repression: inhibition of expression (RNA and/or protein)

Intermediate steps omitted but known

Unknown intermediate steps

No effect: relation excluded

A* Active form of the molecule
Molecular switch on/off binary control of a protein

Synergy (A and B necessary, logical AND) (coincidence detection)

(A or B sufficient to – either – logical OR)

Effect of A on B, C and D known, as well as consequence E, but relations between B, C and D unknown

BCD are part of an assembly

Controls on controls
Permissive effect of B (B necessary for effect of A – if no B, no or little C; B permits the generation of C)

Gate-keeper effect of B (if B, no or little C)

Transcription downstream from the promoter

trends in Cell Biology

rl: {CM| cm {cyto AP1 NF1 NFB Myc {N}}
=> {CM| cm {cyto p53 {N}} .

Rule Cartoons

- Biologist cartoons
  - conceptually useful
  - ambiguous -- missing information
  - multiple proposals for notation
  - no underlying formal model

- Formal cartoons
  - contain all information of a rule
  - suppress place holder information
  - focus on activation and location

Mapping rules to formal cartoons
Metalevel function rs2cts
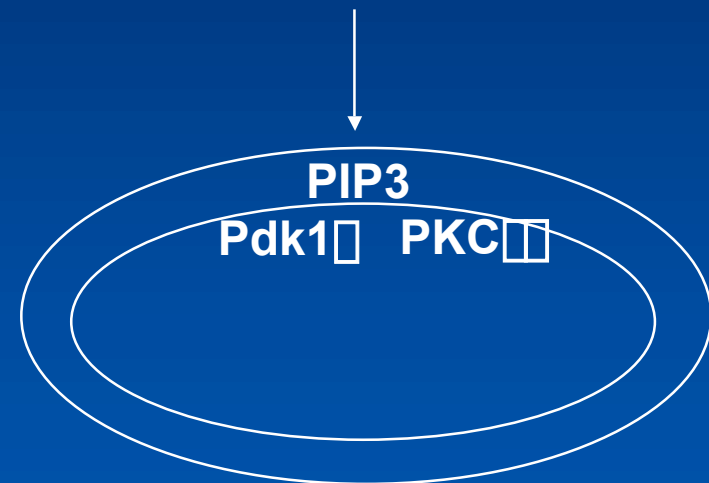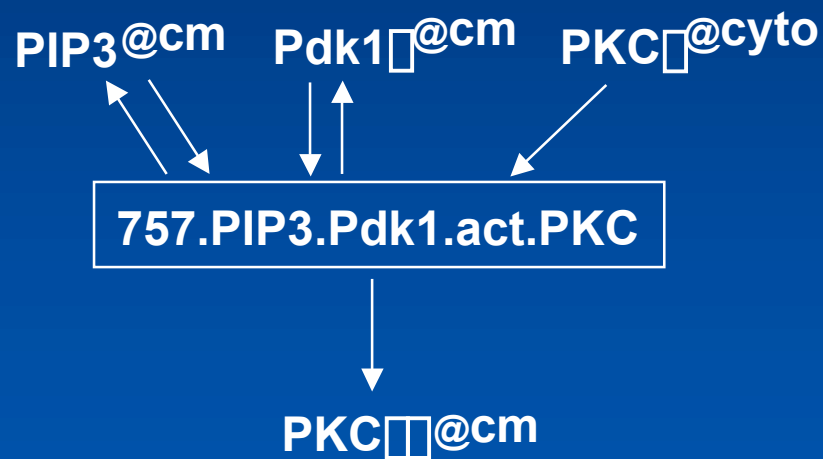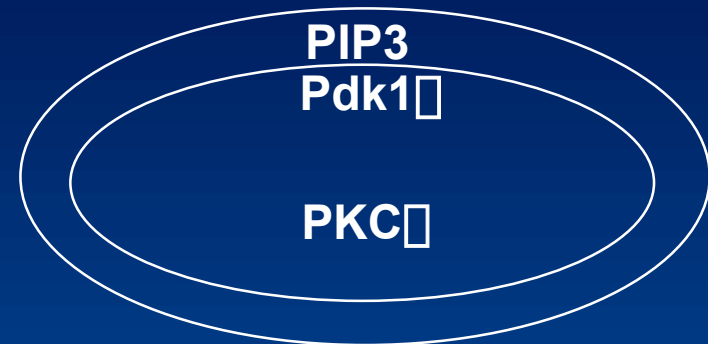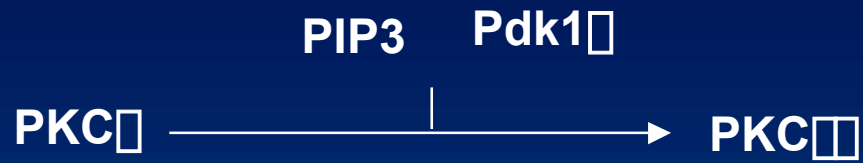
rl[757.PIP3.Pdk1.act.PKCe]:

    {CM | cm PIP3 [Pdk1 – act] {cyto PKCe }} =>

    {CM | cm PIP3 [Pdk1 – act] [PKCe – act] {cyto }} .

red rs2cts(757.PIP3.Pdk1.act.PKCe) .

result:

    ctoon('757.PIP3.Pdk1.act.PKCe,

       (item("PIP3", Protein, membrane('CM))

        item("PKCe", Protein, interior('CM))

        item("Pdk1-act", Protein, membrane('CM))),

       (item("PIP3", Protein, membrane('CM))

        item("PKCe-act", Protein, membrane('CM))

        item("Pdk1-act", Protein, membrane('CM))))

PKC Rule Cartoons

Recapping

● Put new kinds of computational capabilities in the hands of biologists

● Predictive power: generate hypotheses to be tested computaional and biologocial means

● Encouraging results -- biologists already are able to use EPL in designing and analyzing experiments

● A starting point for a wide range of exciting applications of formal modeling.

Future Directions

- Add more data
    - (partially) automate curation
    - interoperation with existing database

- Develop abstraction mechanisms to allow analysis of
    - more complex systems
    - more complex questions

- Develop strategies for more informed state space exploration

- Biologist friendly interface