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 computations 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
Mammalian cell
Somatic Cell Cycles Consist of Alternating DNA Synthesis (S) and mitotic (M) Phases, Separated by Gap Phases (G₁ and G₂)
The ErbB Network

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] .
sort Complex .
subsort Complex < Thing .
op `_:` : Thing Thing -> Complex [comm] .

sorts Soup Enclosure MemType.
subsort Thing < Soup .
op empty : Soup .
ops CM NM : -> MemType .
op `{|{|}}` : MemType Soup Soup -> Enclosure .
PKC Regulation Network
PKC Rules in Maude

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

rl [84.PL.C.act.DAG]:
  Ca++ {CM| cm:Soup PIP2 [?PLC:PLC - act]
       {cyto:Soup}}
  =>
  {CM|cm:Soup (Ca++ DAG IP3) [?PLC:PLC - act]
    {cyto:Soup}} .
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 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]}}}})
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 qlx : -> 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 qlx = q1 - < PI3Ka , cyto >
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(qlx,prop1).
Model Checking q1.m -- simple paths

Maude> red findPath(q1,prop1) .

•PD{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) .

•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]}}}}
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

Symbolic representation:

\[ rl: \{CM\| cm \{cyto \text{AP1 NF1 NFB Myc} \{N\}\} \Rightarrow \{CM\| cm \{cyto p53 \{N\}\}\} . \]
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

\[rl[757.PIP3.Pdk1.act.PKCe]: \]
\[
\{CM | \text{cm PIP3 [Pdk1 – act]} \{\text{cyto PKCe }\}\} \Rightarrow \\
\{CM | \text{cm PIP3 [Pdk1 – act]} [PKCe – act] \{cyto\}\} .
\]

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

PKC → PKC

PIP3 → Pdk1

PKC

757.PIP3.Pdk1.act.PKC

PKC@cyto

PiP3@cm

Pdk1@cm

PKC@cm
Recapping

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

● Predictive power: generate hypotheses to be tested computational and biological 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