Challenges in formal reasoning about signaling networks

Carolyn Talcott
Simmons
August 2015

Plan

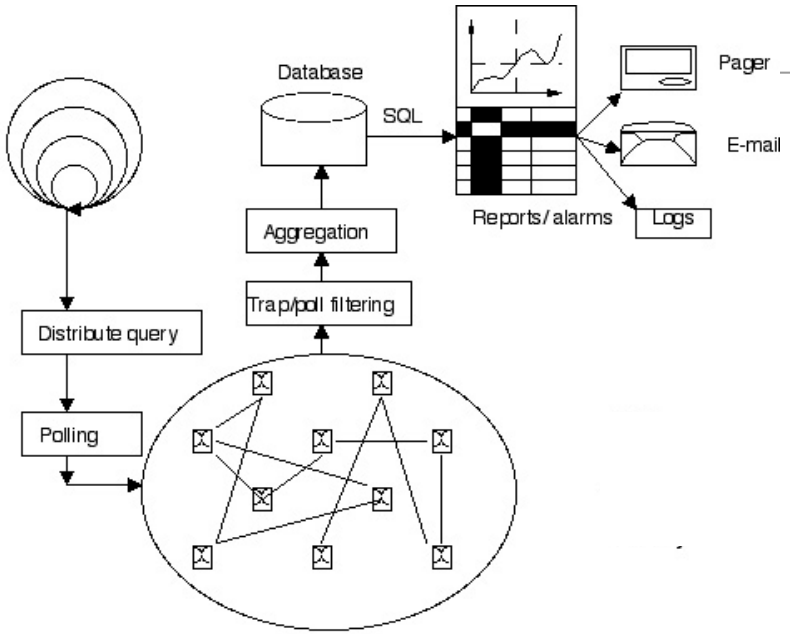
- What models are we talking about?
- Pathway Logic in a nutshell
- Challenges

Modeling 101

- What questions do you want the model answer?
- What can you observe/measure?
- What does that mean?
- Explain it to a computer!
    - Need a formal representation system

Formal Modeling Methodology

data



model

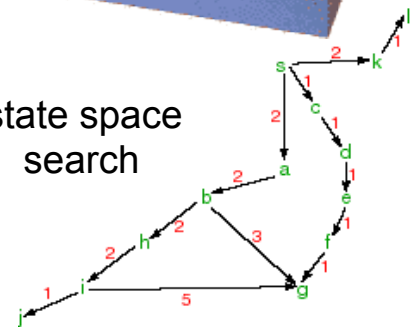
Curator/model builder

asking questions

$S \models \Phi$

model checking

rapid prototyping



state space search

Symbolic analysis -- answering questions

- Forward collection -- upper bound on possible states
- Backward collection -- initial states leading to states of interest
- Search -- for (symbolic) state of interest
- Model checking -- do all executions satisfy ϕ, find counter example
- Constraint solving -- steady state analysis

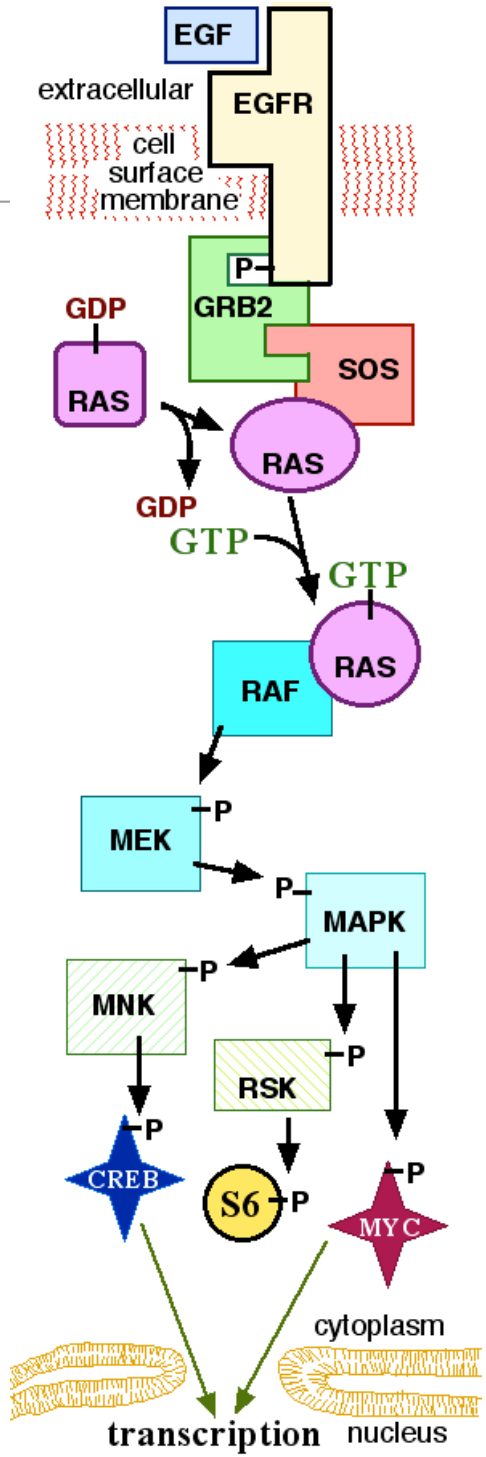
Pathway Logic

Executable models of cellular processes

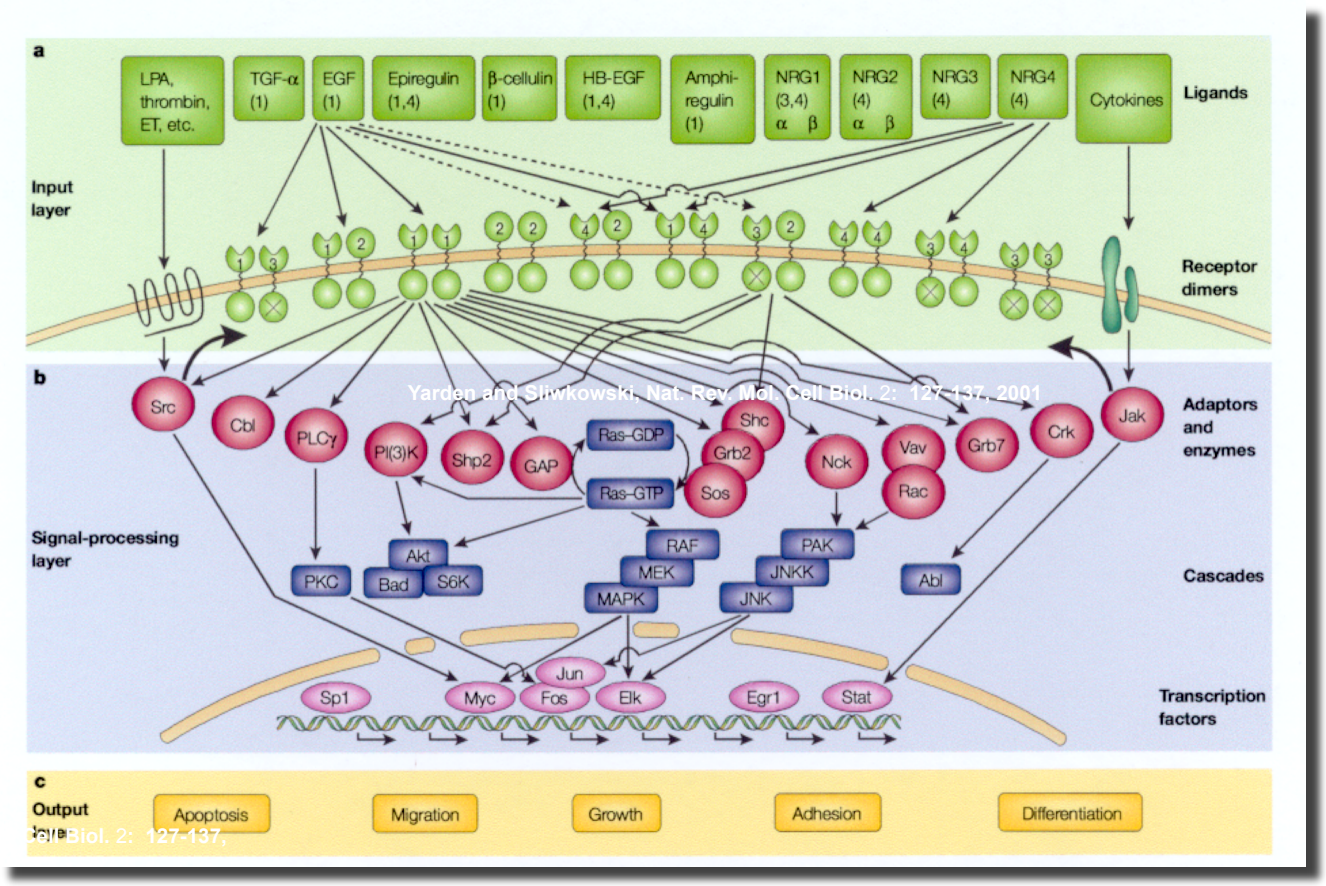
http://pl.csl.sri.com

Pathway Logic (PL) Goals

- Understanding how cells work
- Formal models of biomolecular processes that
  - capture biologist intuitions
  - can be executed
- Tools to
  - organize and analyze experimental findings
  - carry out gedanken experiments
  - discover/assemble execution pathways
- New insights into the inner workings of a cell.
- A new kind of review

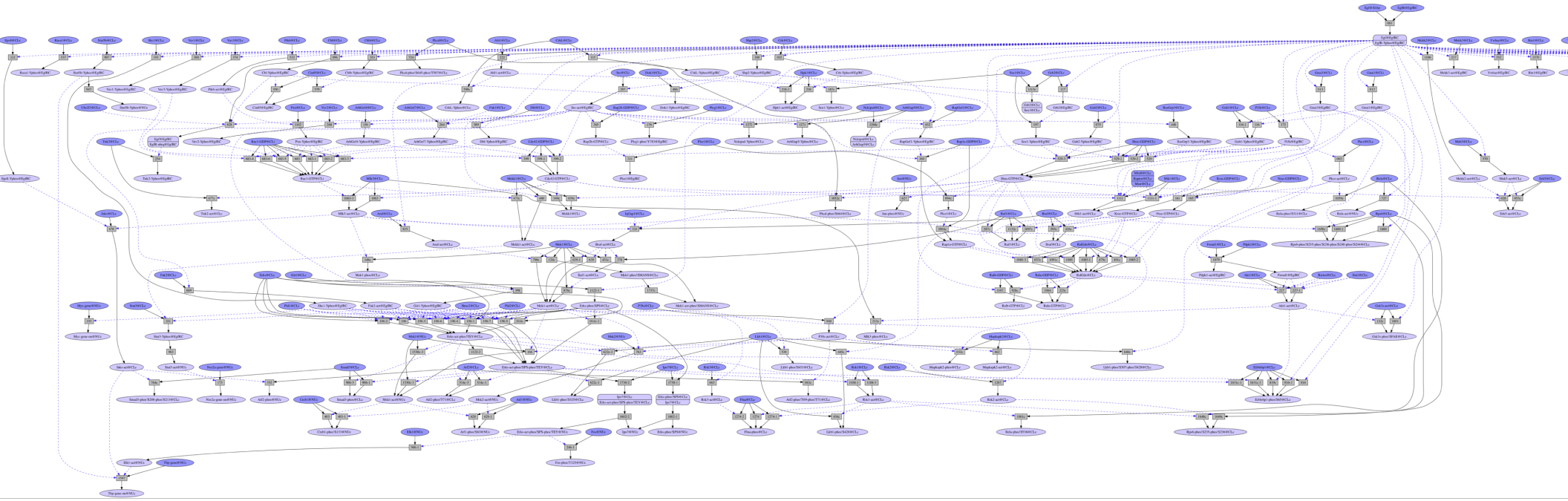


ErbB network cartoon—biologists review model

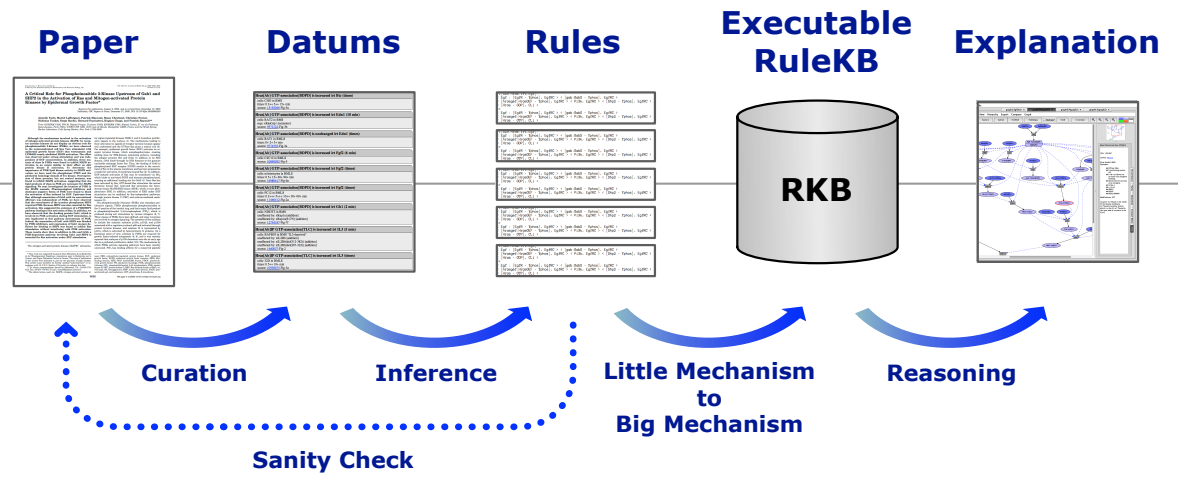


Yarden and Sliwkowski, Nat. Rev. Mol. Cell Biol. 2: 127-137, 2001

PL Egf dish - an executable review model



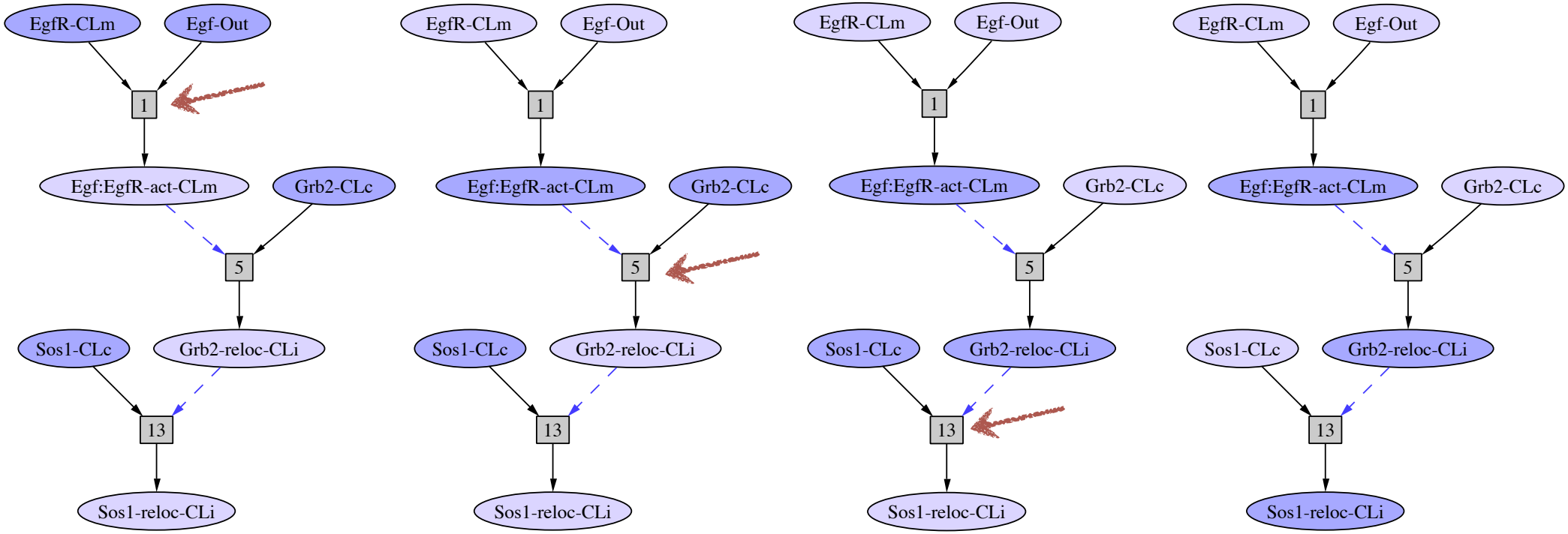
PL from 1k feet



Key components

- Representation system
  - controlled vocabulary
  - datums (formalized experimental results)
  - rules
- Curated datum knowledge base (KB) and search tool
- Evidence based rule networks
  - STM, Protease, Mycolate, GlycoSTM
- Executable models
  - generated by specifying initial conditions and constraints
  - query using formal reasoning techniques
- Visualize and browse subnets

Example signal propagation (using Petri Nets)



Sos1Dish =rule1=> Sos1Dish1 =rule5=> Sos1Dish2 =rule13=> Sos1Dish3

Ovals are occurrences -- biomolecules in locations (aka places).
Dark ovals are present in the current state (marked).
Squares are rules (aka transitions).
Dashed edges connect components that are not changed.

Rule Knowledge Base (RKB) — A Rewrite Theory

- Rewrite rules describe local change and specify required context

rl[529.Hras.irt.Egf]:

< Egf : [EgfR - Yphos], EgfRC > < [gab:GabS - Yphos], EgfRC >

< [hrasgef:HrasGEF - Yphos], EgfRC > < Pi3k, EgfRC > < [Shp2 - Yphos], EgfRC >

< [Hras - GDP], CLi >

=>

< Egf : [EgfR - Yphos], EgfRC > < [gab:GabS - Yphos], EgfRC >

< [hrasgef:HrasGEF - Yphos], EgfRC > < Pi3k, EgfRC > < [Shp2 - Yphos], EgfRC >

< [Hras - GTP], CLi >

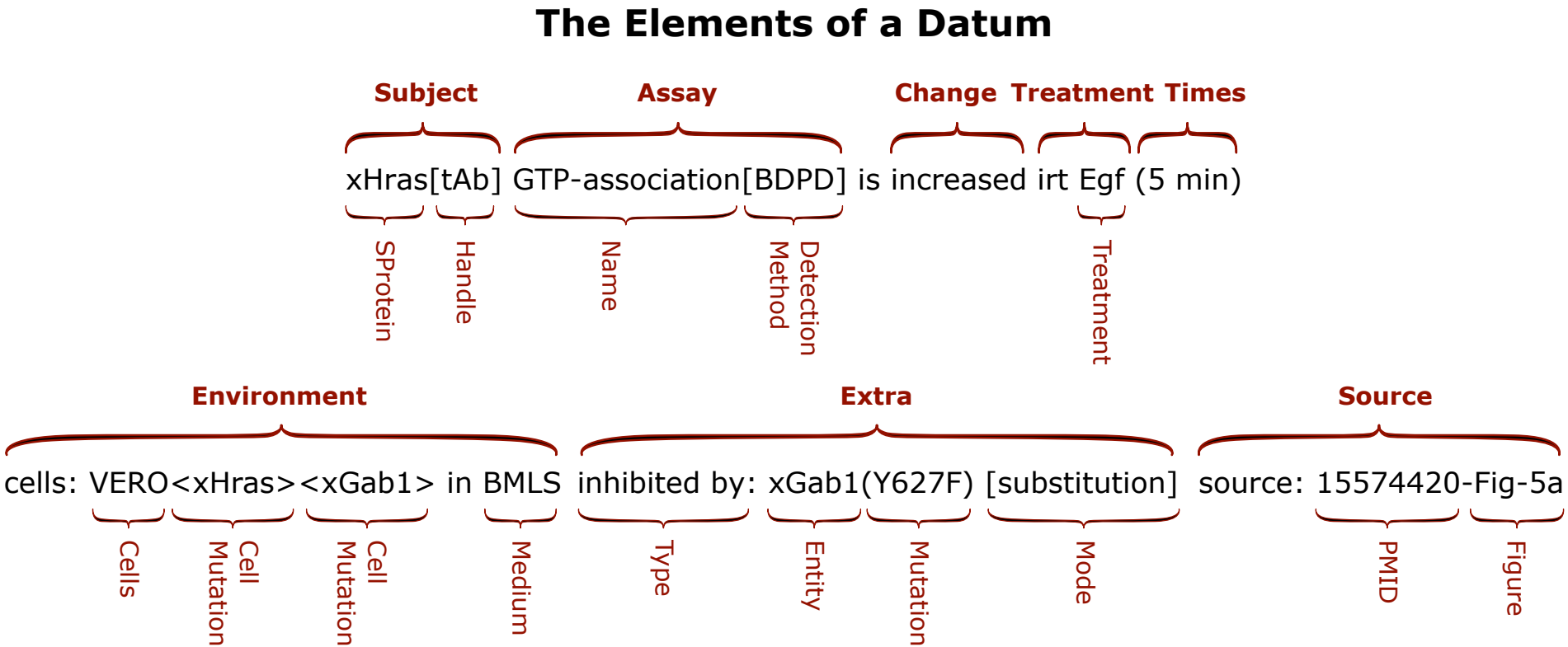
*** ~/evidence/Egf-Evidence/Hras.irt.Egf.529.txt

- Symbolic rules represent a family of rules using sorted variables
- EgfRC is the location of the Egf Receptor complex, it is populated in response to the Egf signal. CLi is the membrane interior
- gab:GabS is a variable standing for Gab1 or Gab2, hrasgef:HrasGEF is a variable for any of several HrasGEFs (enzymes to exchange GDP for GTP)

Where do rules come from?

They are inferred from experimental findings.

These are collected using a formal data structure call datums

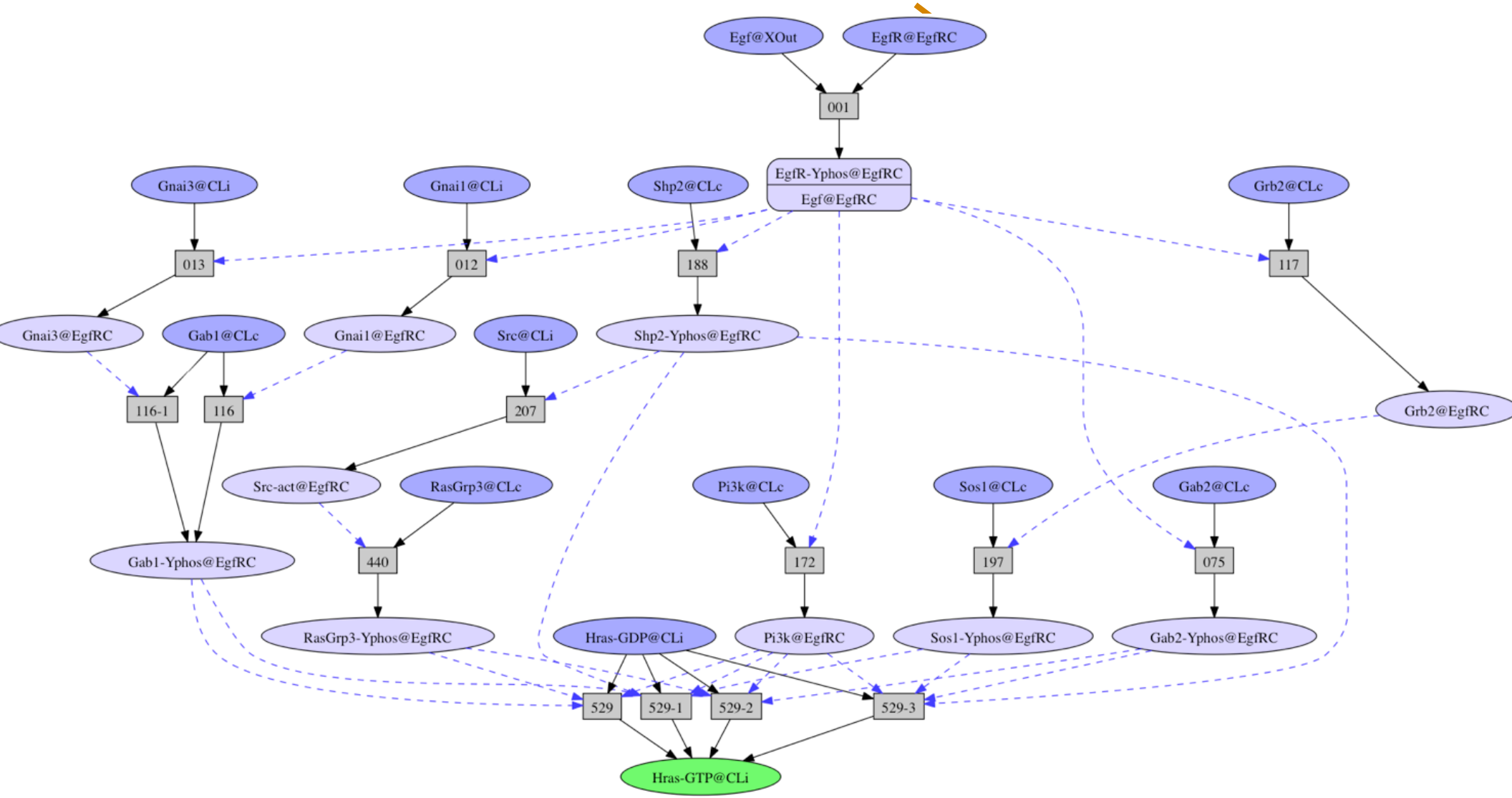


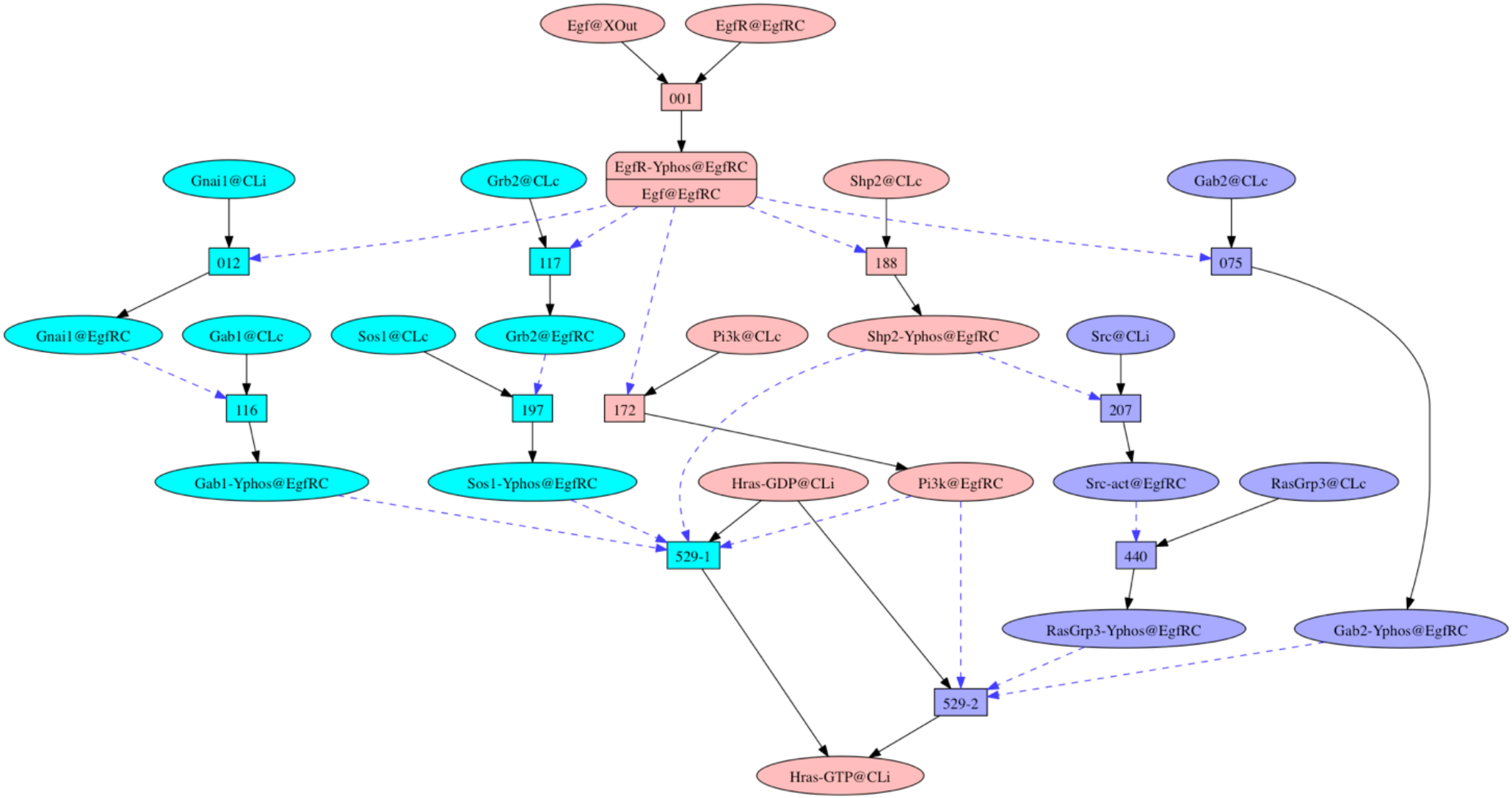
What can be done with an executable RKB?

- Generate a model, for example, of response to some stimulus
  - Define initial state -- cell components and additions -- experimental setup
  - Forward collection gives a network of all the possibly reachable rules
- The Signal Transduction Model (STM) RKB comes with > 30 models of response (of a resting cell) to different treatments (Egf,Insulin,Tnfa,Tgfb, Lps (bug bit), Serum, ...)
- From a model you can
  - Backwards collection generates a subnet relevant to a specific outcome (Erk activated in the nucleus)
  - find an execution path—model-checking the assertion that no path exists
  - carryout in silico knockouts
  - compare nets
  - explore connections up/down stream

The subnet of the Egf model for activating (GTPing) **Hras**.

(Represented as a Petri net.)





Comparing two pathways

Symbolic analysis -- answering global questions

- RMP -- the set of all reaction minimal paths from an initial state to a given set of goals
- From this set we can compute
  - Essential transitions – reactions that are in all pathways to an output.
  - Used places – biochemical species that are in at least one pathway.
  - Knockouts – biochemical species that are in all pathways to an output.
  - Multisignal cellular responses – at least one pathway to an output has more than one stimulus.
- In the Hras subnet
  - 6 execution pathways (3 using Sos1, 3 using RasGrp3 — GEFs)
  - 20 double knockouts (from 4 protein pairs)

Challenges

- Collecting data
- Data to Knowledge
- Knowledge to Model
- Dynamics

Automating all of this!

Collecting Data

- Finding sources
  - abstracts are NOT sufficient sources for indexing / search
  - inspiring Biologists to do the missing experiment
- Extracting datums (semi) automatically
  - Naming things
  - What was measured, under what conditions
  - Resolving ambiguities

Data to Kapta:

Assembling diverse formal factletts into signaling rules

- Each datum is a partial view — constraint solving can integrate views
- Resolving apparent inconsistencies
- Resolving different experimental conditions
- Interpreting variants
  - mutations, deletions, fragments
  - Why was the perturbation chosen? How does that change the meaning?
- Combining experiments
  - activity in a test-tube + activity in a cell

KB to Model

- Rules may not connect
    - Different levels of detail
    - Location location — experiments rarely give location information :-(
    - Activity vs modification state
- What does a give cell type/state express — the initial state
    - It is rare that cells are well characterized

Background knowledge!!!

- Notions such as kinase, Yphos more specific than phos ….
- Needed in all steps
- The is a lot of it
  - some in databases, some in books
  - can it be relied upon — computational vs experimental
- Needs to be curated into computable form.
- Need a well-defined formal representation that can be used by

Dynamics

- PL models capture before/after but not how much or how fast.
- Data for real kinetics is hard to find
- Cells aren't well stirred solutions
- What really matters depends on the question being asked
- What about symbolic quantitative reasoning?
  - using variables and constraints among them
  - what are the right abstractions
  - what are relative rates of processes
  - characterizing effects of competition between processes
- We need biological reasoning principles (which can then be formalized and helpful tools built)

Questions ???

Controlled vocabulary

```
sort HrasSort .
subsort HrasSort < RasS < BProtein .

op Hras : -> HrasSort [ctor metadata (  (spnumber P01112)  (hugosym HRAS)
  (synonyms "GTPase HRas"
            "Transforming protein p21"
            "v-Ha-ras Harvey rat sarcoma viral oncogene homolog"
            "Harvey murine sarcoma virus oncogene"
            "H-Ras-1"
            "c-H-ras"
            "HRAS1"
            "RASH1"
            "RASH_HUMAN"))] .

op Rass : -> RasS [ctor metadata (  (category Family)  (members Hras Kras Nras))] .

op Pi3k : -> Composite [ctor metadata "(
  (subunits Pik3cs Pik3rs)
  (comment  "PI3 Kinase is a heterodimer of:"
            "a p110 catalytic subunit: Pik3ca, Pik3cb, Pik3cd or Pik3cg"
            "a p85 regulatory subunit: Pik3r1, Pik3r2, or Pik3r3"))] .
```