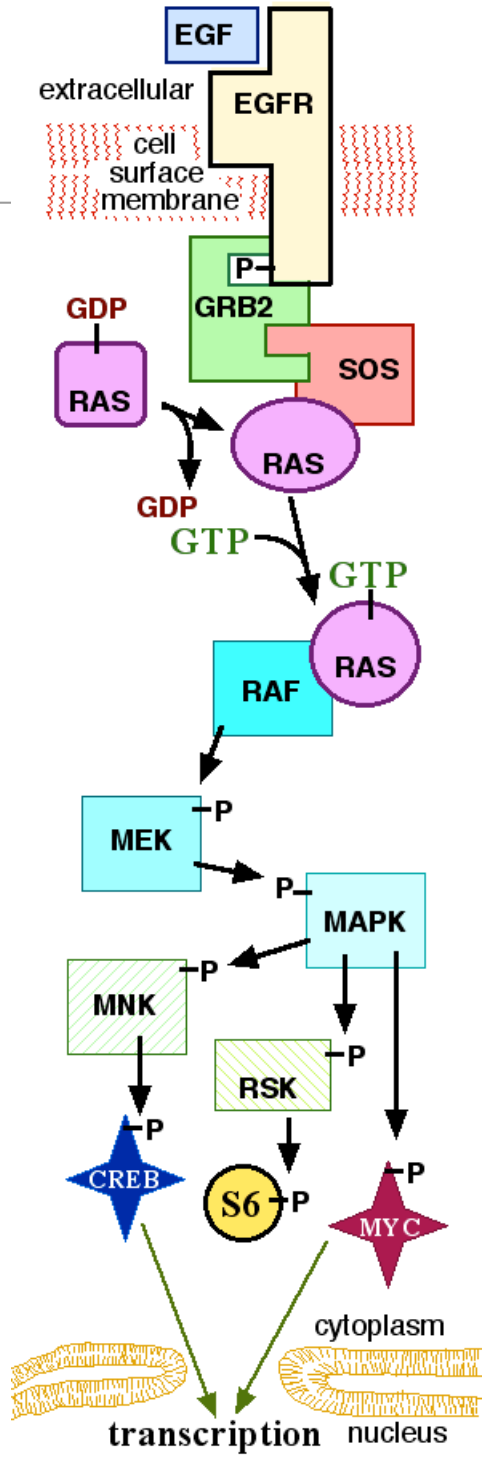
Pathway Logic: Symbolic Executable Models for Reasoning about Cellular Processes.

http://pl.csl.sri.com

Carolyn Talcott
LSB
July 13, 2018

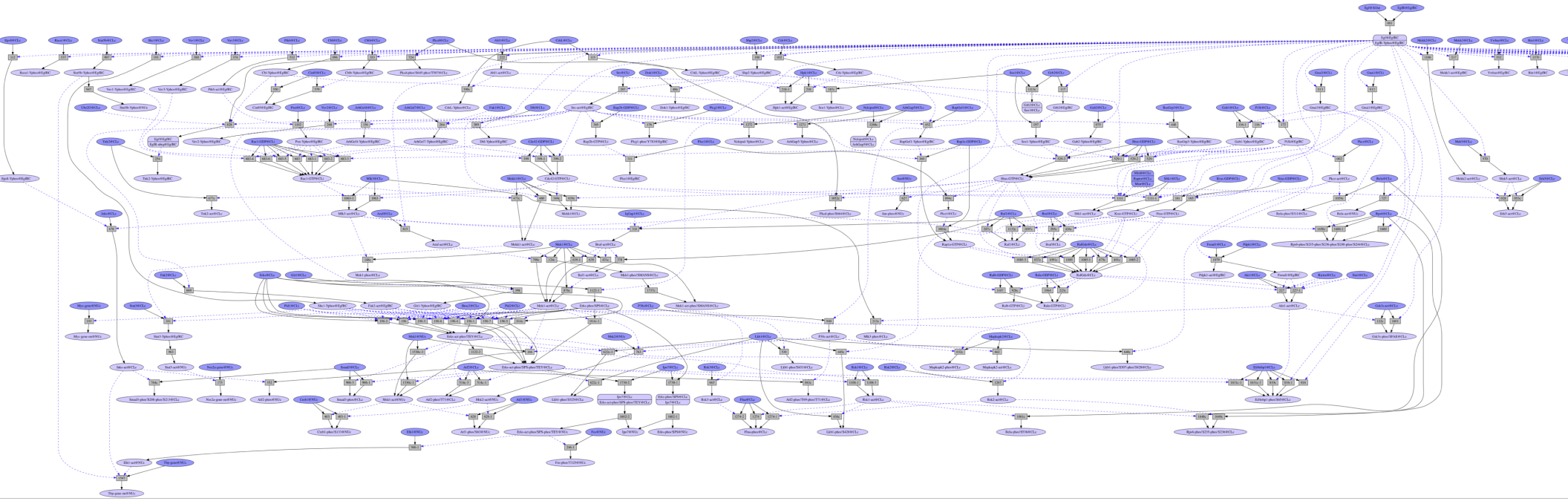
Pathway Logic (PL) Goals

- Understanding how cells work or not
- Response to stimuli or perturbations
- 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
- PL model as a new kind of review



Pathway Logic model of response to Egf

An interactive, executable review model

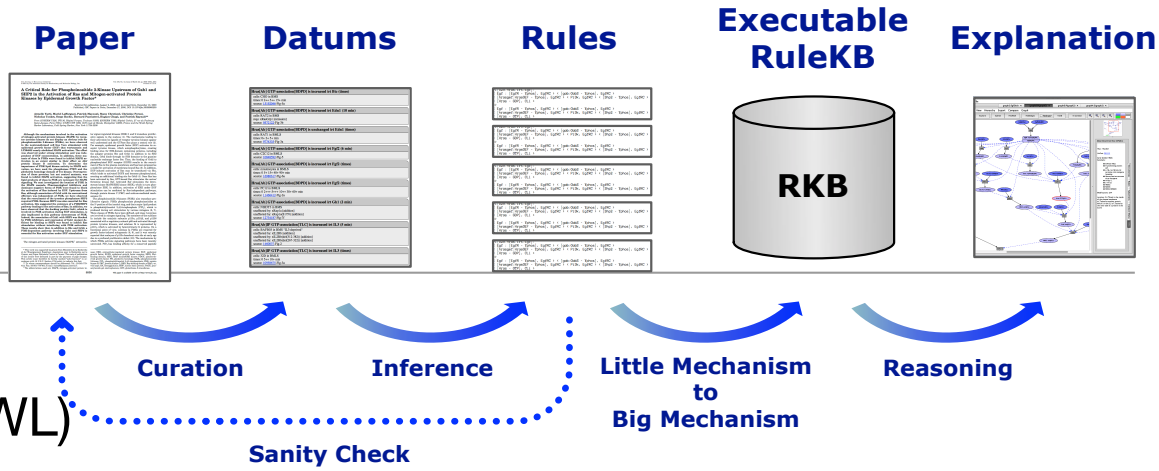


Plan

- Overview/Background
- PL representation basics
- The Pathway Logic Assistant
  - Visualization, interaction, reasoning
- Applications
  - Explaining drug data
  - Host pathogen interactions
- Conclusion

Pathway Logic (PL) Overview/Background

PL in a Nutshell



Key components

- Formal representation system (RWL)
  - controlled vocabulary, anchored to standards
  - rule knowledge base, founded on experimental evidence
- Executable models
  - generated by specifying initial conditions and constraints
  - queried using formal reasoning techniques
- PLA to visualize, browse, query rule networks
- Curated datum knowledge base (KB) and search tool
  - datums: computable representation of experimental results
- Curated rule networks
  - STM, Protease, Mycolate, GlycoSTM (pl.csl.sri.com/online.html)

Executable formal models

- Something to play with — ala model train, architectural model, Sim City…
- Computer Representations of
  - System state: collections of occurrences (name,state,location)
    - Initial state: an experimental setup/cell state -- what is expressed, where; what modifications to the cell; what treatment has been applied
  - State transition rules
    - metabolic reaction, signal transduction step, secretion, cell mobility
- Execution: set of rule applications -- possible behavior
  - How does a signal propagate, watch things light up as modified ...
  - Find collections of cellular components that function together
  - One possible notion of “Pathway”
- Do in silico experiments

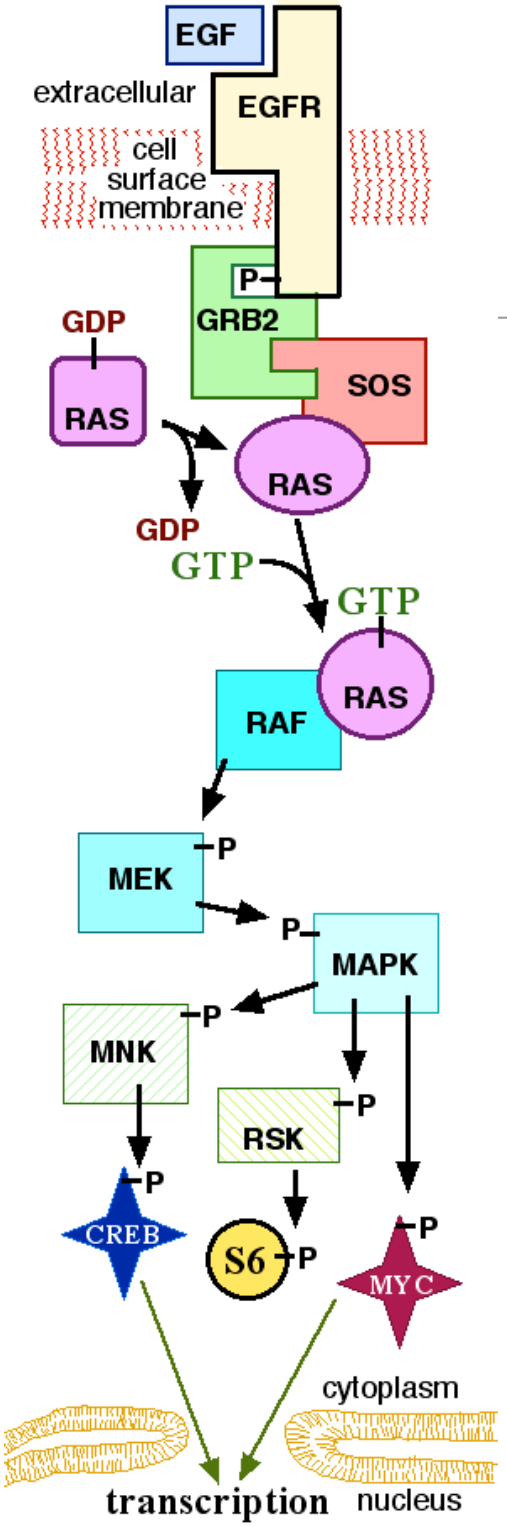
Rewriting Logic (RWL)

- Rewriting Logic is a logical formalism that is based on two simple ideas
  - states of a system are represented as elements of an equationally specified 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!)

Maude

- Maude is a language and tool based on rewriting logic
- Available at: http://maude.cs.uiuc.edu
- Features:
  - High performance engine
    - {ACI} matching
    - position /rule/object fair rewriting
  - Modularity and parameterization
  - Builtins -- booleans, number hierarchy, strings, SMT solving
  - Reasoning: search and model-checking
  - Reflection -- using descent and ascent functions

Pathway Logic (PL) Basics



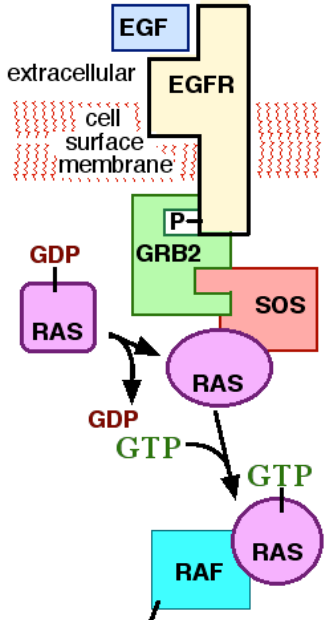
Classic example: Egf stimulation of the Mitogen Activated Protein Kinase (MAPK)

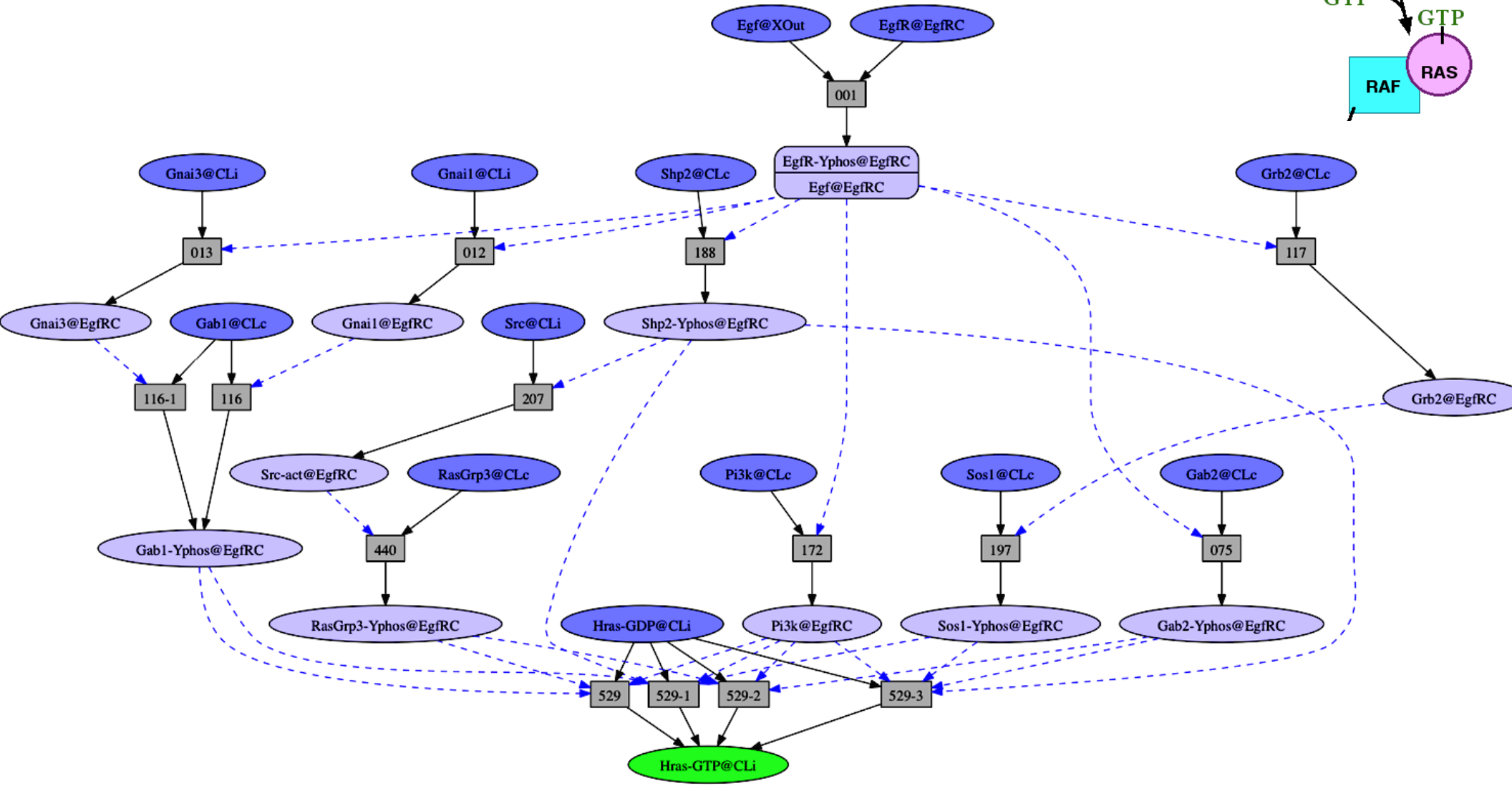
Egf → EgfR → Grb2 → Sos1 → Ras → Raf1 → Mek → Erk

- Egf (EGF) binds to the Egf receptor (EgfR, EGFR) and stimulates its protein tyrosine kinase activity to cause auto-phosphorylation, 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 by exchanging GDP for GTP.
- 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) ...

The subnet of the PL Egf model activating **Hras**.

(Exchanging GDP for GTP.)





Rule Knowledge Base (RKB) Sample

- PL 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)

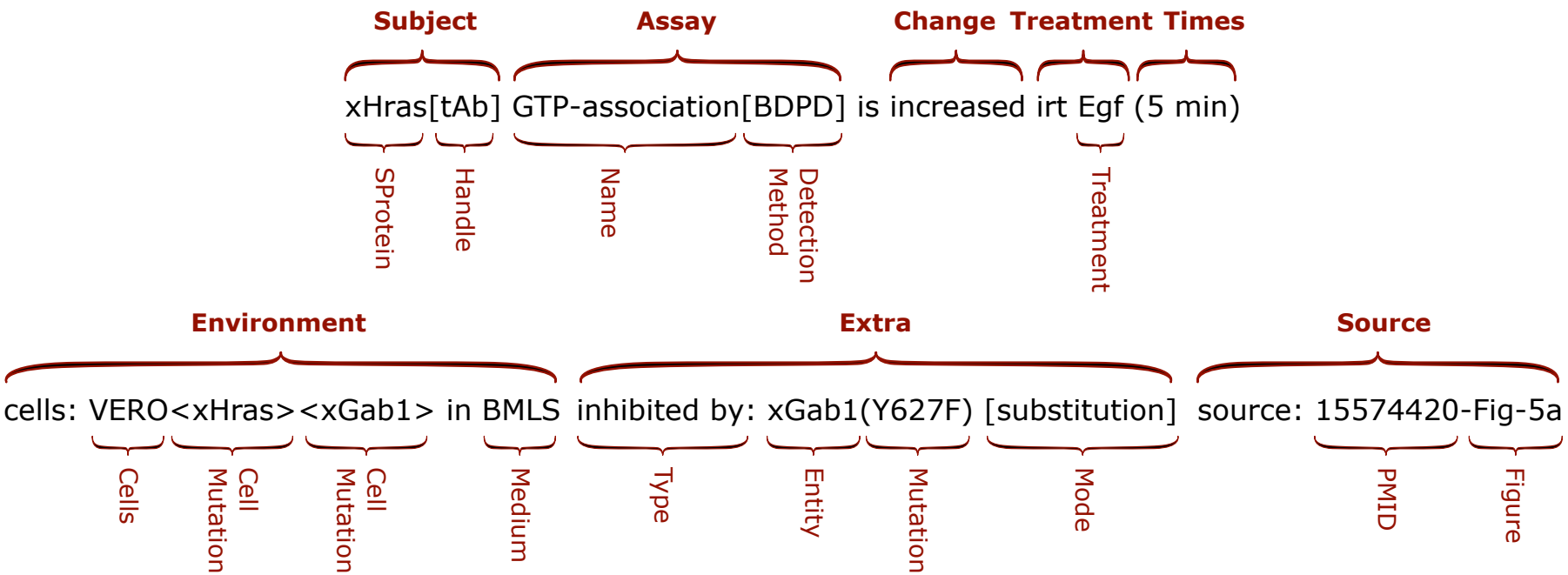
Supporting controlled vocabulary sample

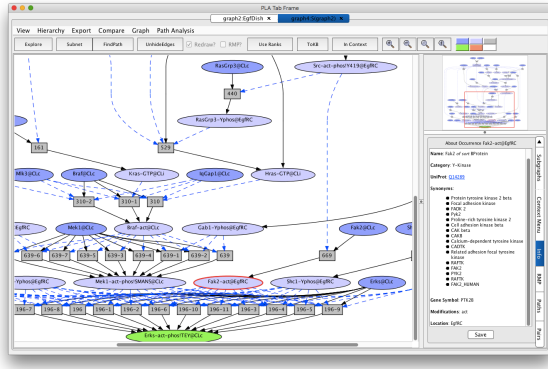
```
op Hras : -> HrasSort [ctor metadata "(\
  (category GTP-Binding-Protein)\
  (spnumber P01112)\
  (hugosym HRAS)\
  (synonyms \"GTPase HRas\"\
        \"Transforming protein p21\"\
        \"v-Ha-ras Harvey rat sarcoma viral oncogene homolog\"\
        \"Harvey murine sarcoma virus oncogene\"\
 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

subsort HrasSort < RasSort < BProtein .
```

Where do rules come from?

- They are inferred from experimental findings.
- These are collected using a formal data structure call datums
  - datums are available in text (readable) or json (computable)
- The datum below says that the amount of GTP bound to Hras is increased after addition of Egf (Epidermal Growth Factor) to VERO cells for 5 minutes. Also tyrosine phosphorylated Gab1 is involved (the extra).

The Elements of a Datum





The Pathway Logic Assistant (PLA)

Pathway Logic Assistant (PLA): a tool to visualize and interact with rule networks using symbolic analysis

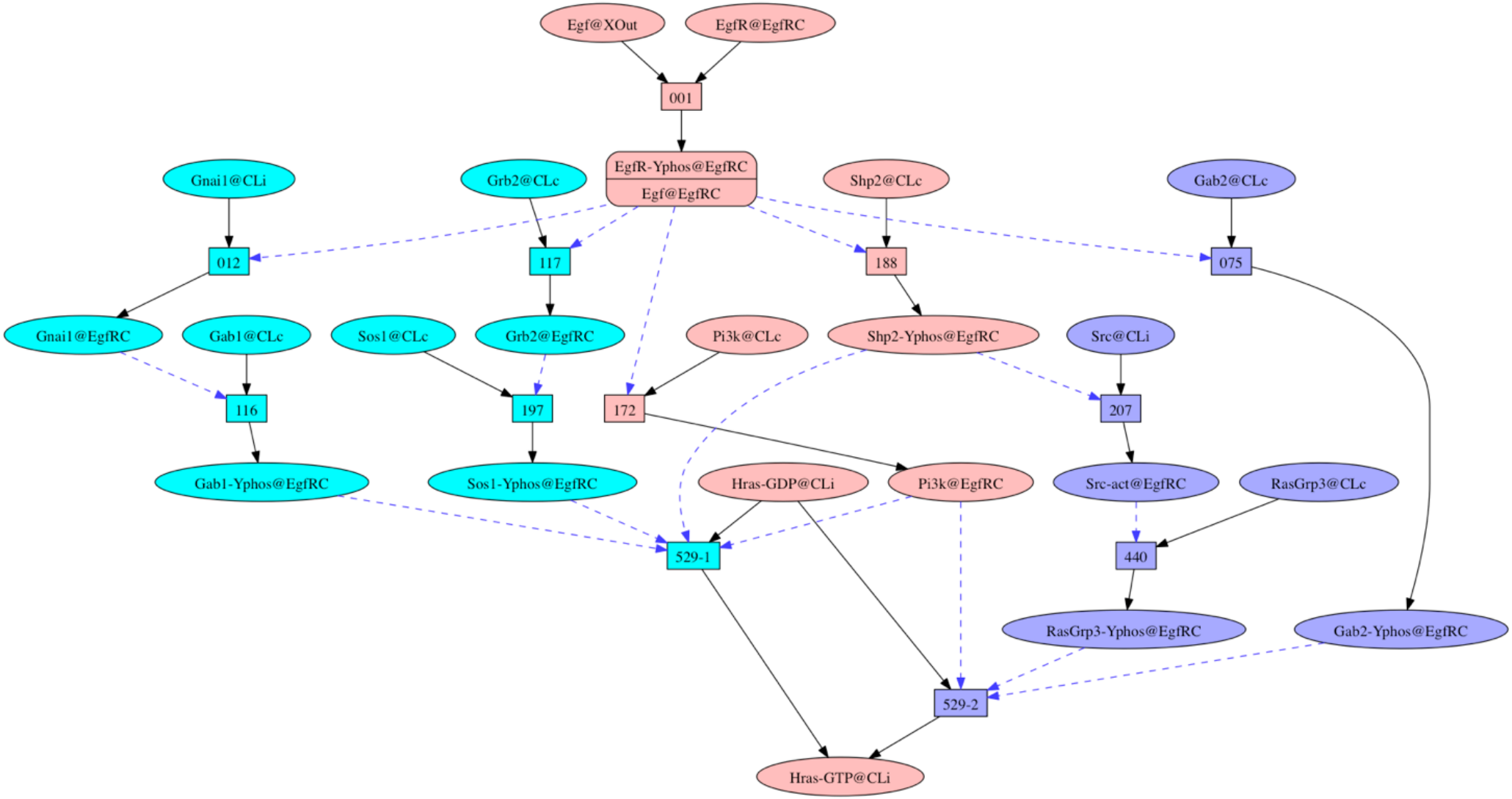
---

- Create specific models
  - forward collection
- Derive subnet relevant to given goals/avoids
  - backwards collection
- Find an execution (Path)
  - model checking
- Compare paths
- Explore -- what is upstream/downstream of X
    - symbolic execution
- All Paths analysis — essential rules, knockouts, uses
  - petri net algorithm

Using PLA to generate the Hras subnet.

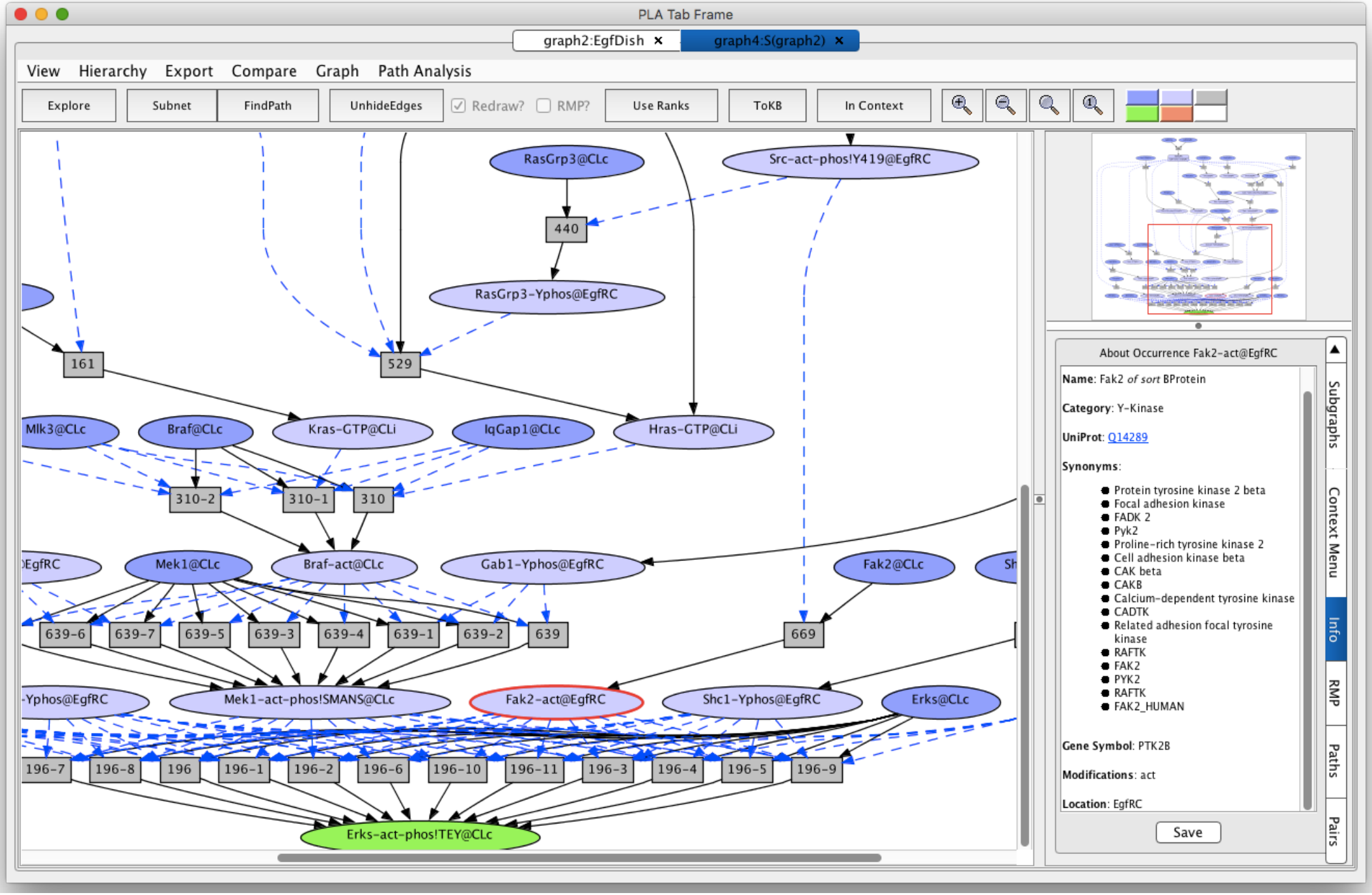
- Derive the network of rules modeling response to Egf
  - Forward reasoning generates a model of Egf signaling, the network of rules reachable from the EfgDish
- Find Hras activated (bound to GTP) in the picture (where is waldo?) and mark as a goal.
- Compute (and display) the relevant subnet.
  - Backward reasoning generates the network of all execution paths leading to activation of Hras in the Egf model
- All paths analysis finds 6 pathways

Comparing two pathways in the Hras subnet, found by model checking.



Pink: both pathways, cyan. blue different pathways.

Erk subnet in PLA



Using Pathway Logic to Explain Data

The Experiment [Korkut et. al]

- Exponentially growing SKMEL133 cells were treated with 12 drugs including
  - AKTI12 — Akt Inhibitor
  - PD0325901 — Mek1 Inhibitor
  - PLX4720 — BrafV600E inhbitor
  - Temsirolimus — Mtorc1 (Mlst8:Mtor:Raptor) inhibitor
  - ZSTK474 — inhibits catalytic subunit of Pi3k
- Change in protein expression/phosphorylation was measured for 138 entities at 24 hours using Reverse Phase PhosphoProteomics Analysis (RPPA)

[Korkut, et.al.] Perturbation biology nominates upstream-downstream drug combinations in Raf inhibitor resistant melanoma cells. Elife, 18(4), 2015

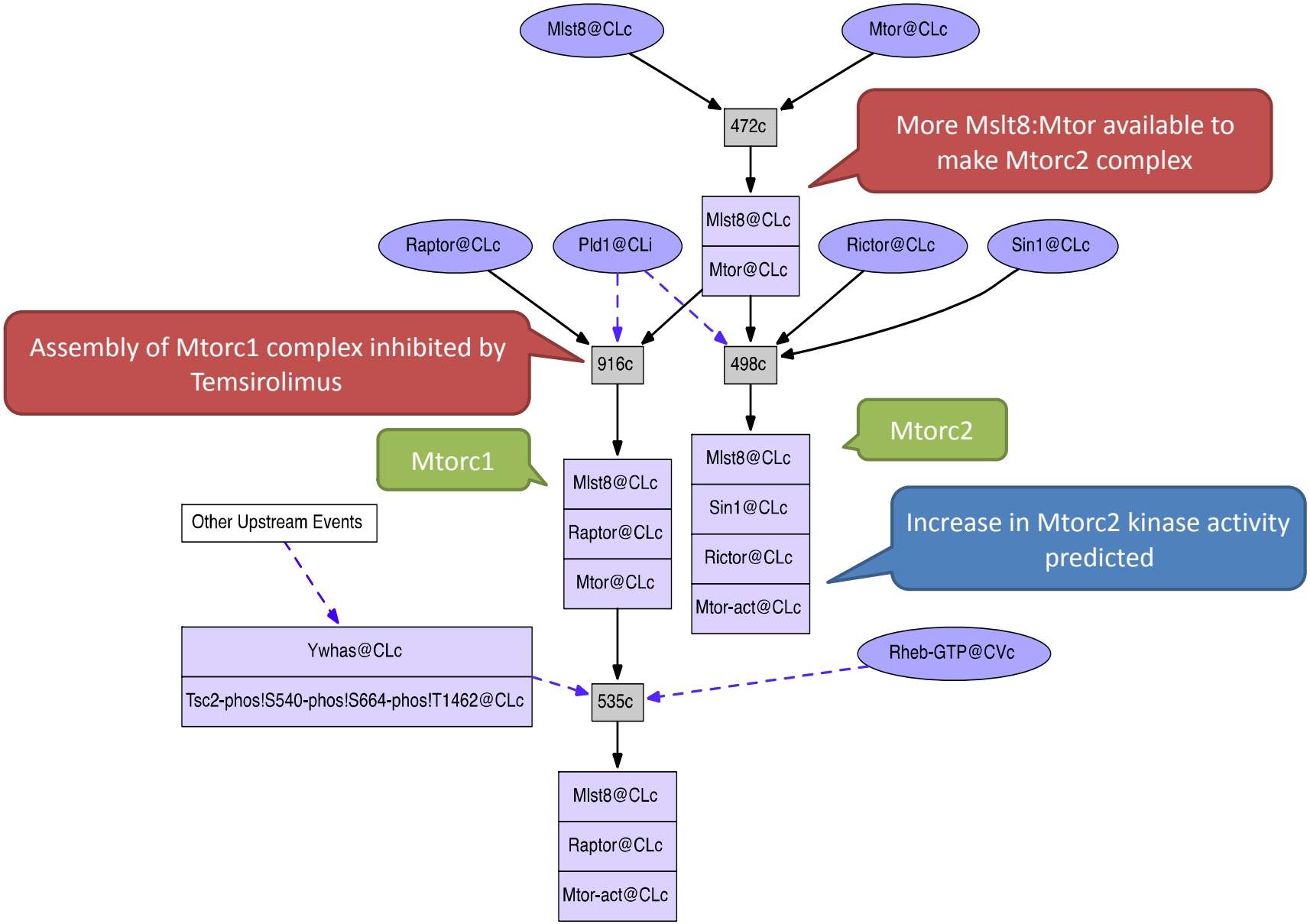
Developing a model

- Derive a model of the experiment
  - Identify rules that would cause the modifications seen in the data.
  - Identify rules that that would meet the requirements of the first set of rules.
  - Iterate until there are no more requirements to be met (backward collection).
  - Infer an initial state: for each entity in the relevant rules - determine the locations and modifications that cannot be created by any rules.
- Model the treatment effect
  - Generate the unperturbed network from the rules and initial state (forward collection).
  - For each drug, remove the occurrence inhibited by the drug from the network.
  - Compare the reachable and unreachable occurrences in the absence of drug target – visually or computationally.

Explanation principles

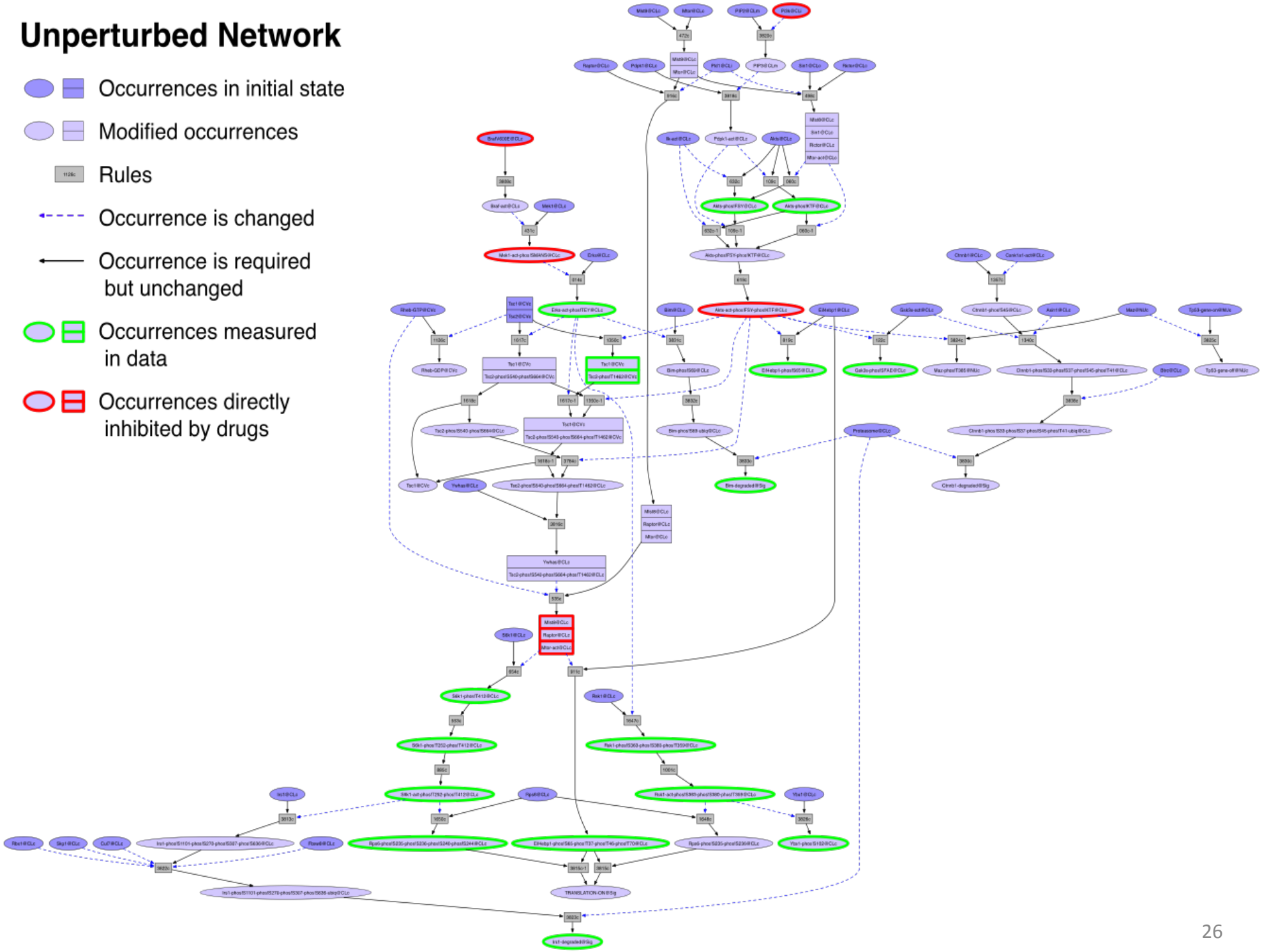
- A phosphorylation state no longer reachable predicts a decrease in that state.
- A degradation state no longer reachable predicts an increase in expression.
- A blocked branch suggests an increase in the non-blocked branch.

A blocked branch suggests an increase in the non-blocked branch



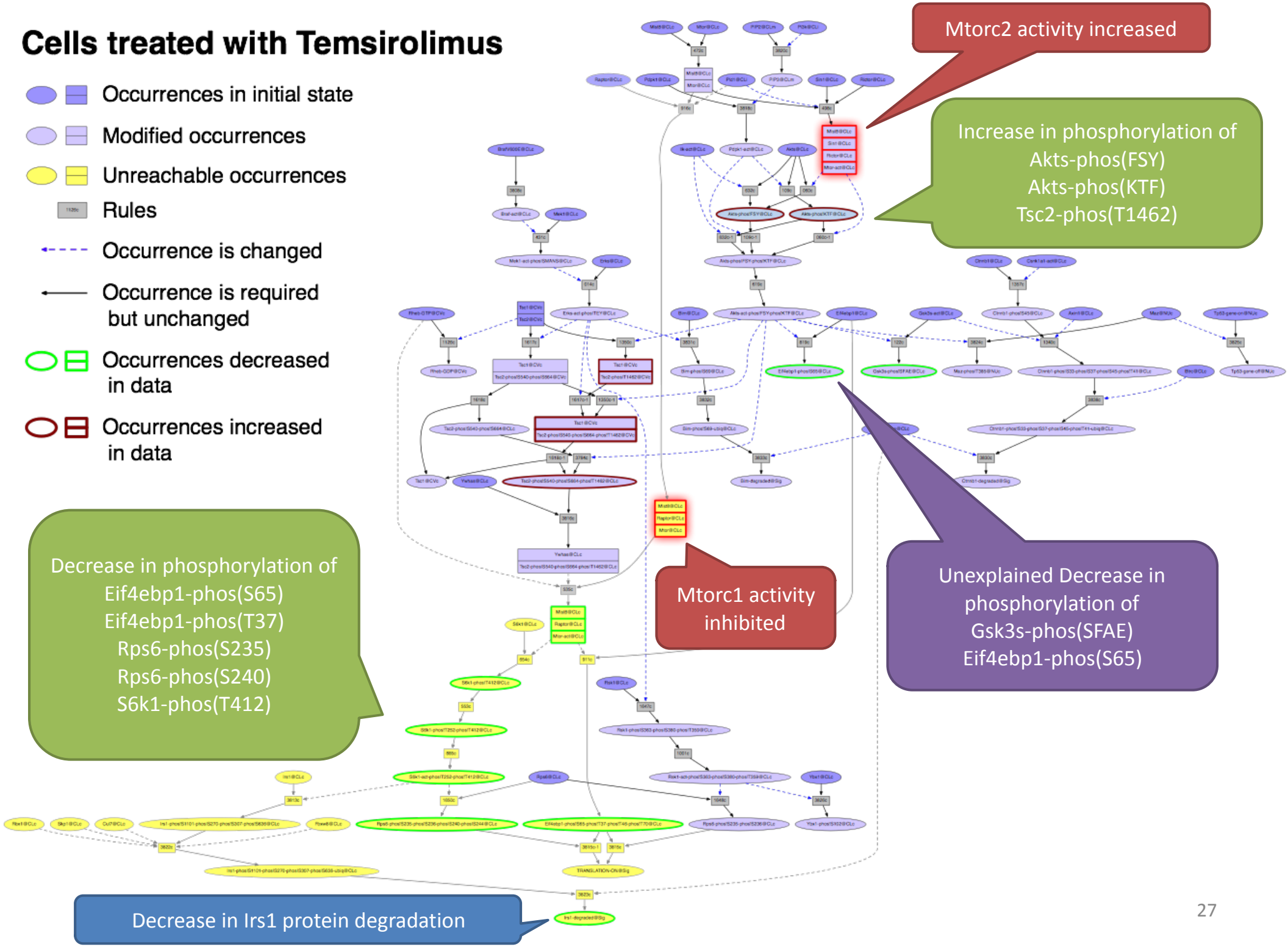
Unperturbed Network

- Occurrences in initial state
- Modified occurrences
- Rules
- Occurrence is changed
- Occurrence is required but unchanged
- Occurrences measured in data
- Occurrences directly inhibited by drugs



Cells treated with Temsirolimus

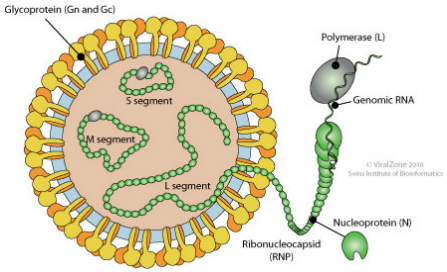
- Occurrences in initial state
- Modified occurrences
- Unreachable occurrences
- Rules
- Occurrence is changed
- Occurrence is required but unchanged
- Occurrences decreased in data
- Occurrences increased in data



Results

- Using the PL STM model and our 3 `principles' we could explain 42 out of 107 changes in response to 5 drugs.
- Unexplained changes were mainly in protein expression. Effects on transcription, translation and cell cycle are not the focus of the current model.
- Some discrepancies were found, indicating likely missing controls in rules, could be variance in data.

- For more details see: C. Talcott and M. Knapp. Explaining response to drugs using Pathway Logic. In Computational Methods in Systems Biology, 2017.
- The SKMEL133 model is available at pl.csl.sri.com/online.html as part of the Pathway Logic suite of models.



Host pathogen interactions

Context

- Pathway Logic plays an integrating role in an IARPA FunGCAT project.
- FunGCAT = Functional Genomic and Computational Assessment of Threats

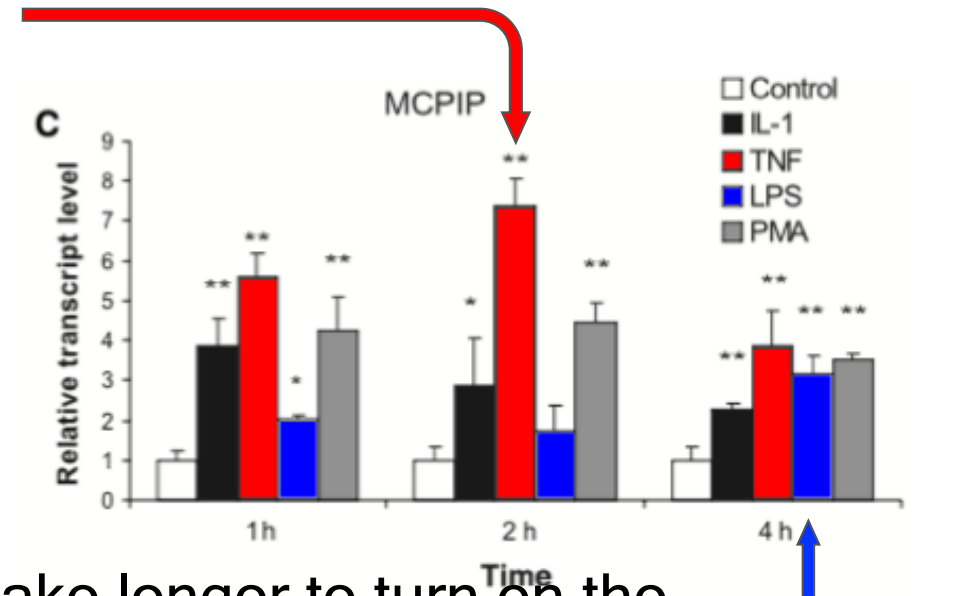
- Challenge: determine if a given DNA/RNA/peptide fragment is part of a pathogenic entity.

- The project involves:
  - Machine learning to classify patterns
  - Experimental exploration and validation
  - KBs and executable network models
    - to organize data
    - identify attack surfaces/points

Explaining experimental observation

The combined LPS and TNF PL response networks were used to explain the data from PMID:19909337-Fig-1c.

Addition of TNF to U937 cells causes an increase in MCPIP1 mRNA by 1 hour and a peak at 2 hours.

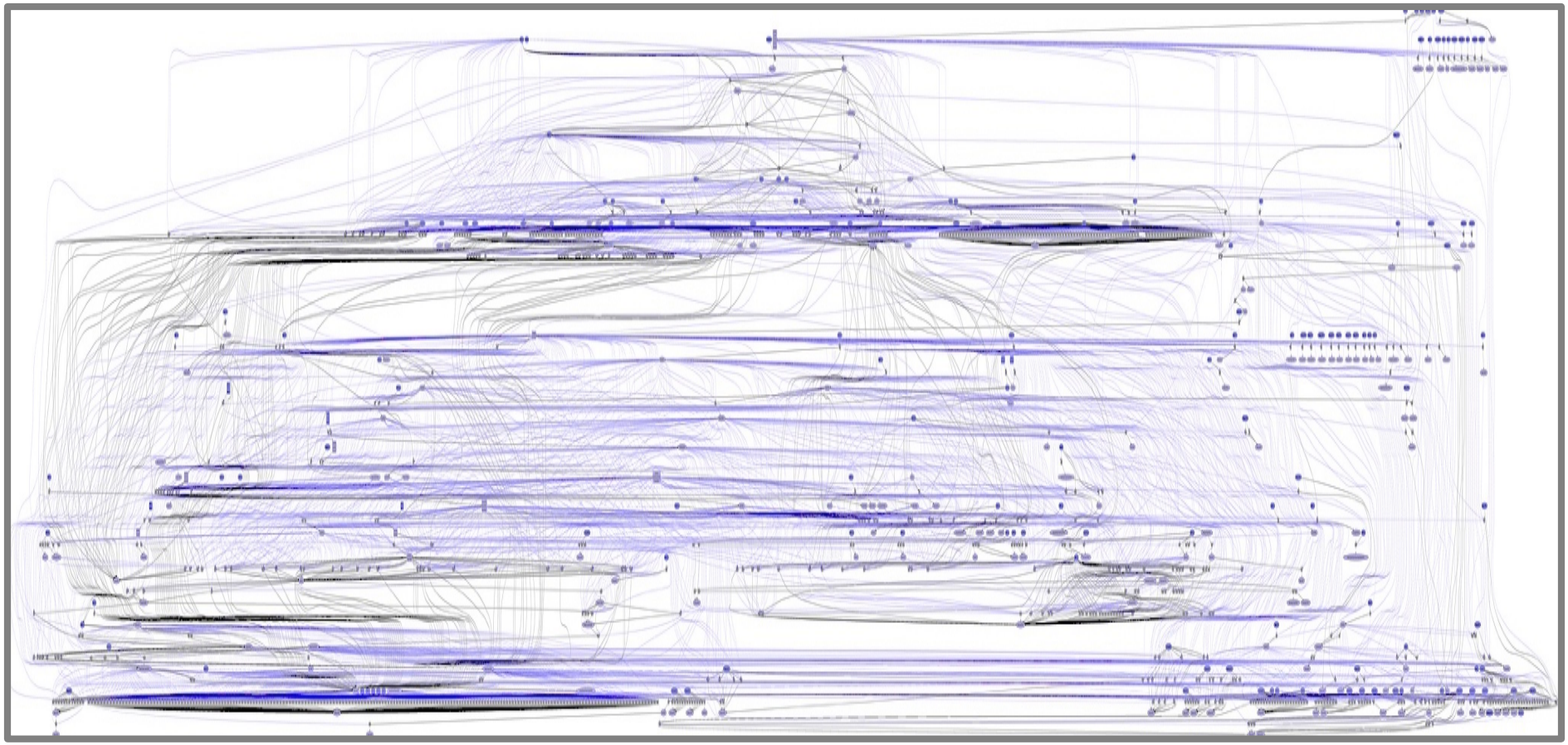


MCPIP1 mRNA does not increase significantly until 4 hours after addition of LPS.

Why does LPS take longer to turn on the MCPIP1 gene than TNF?

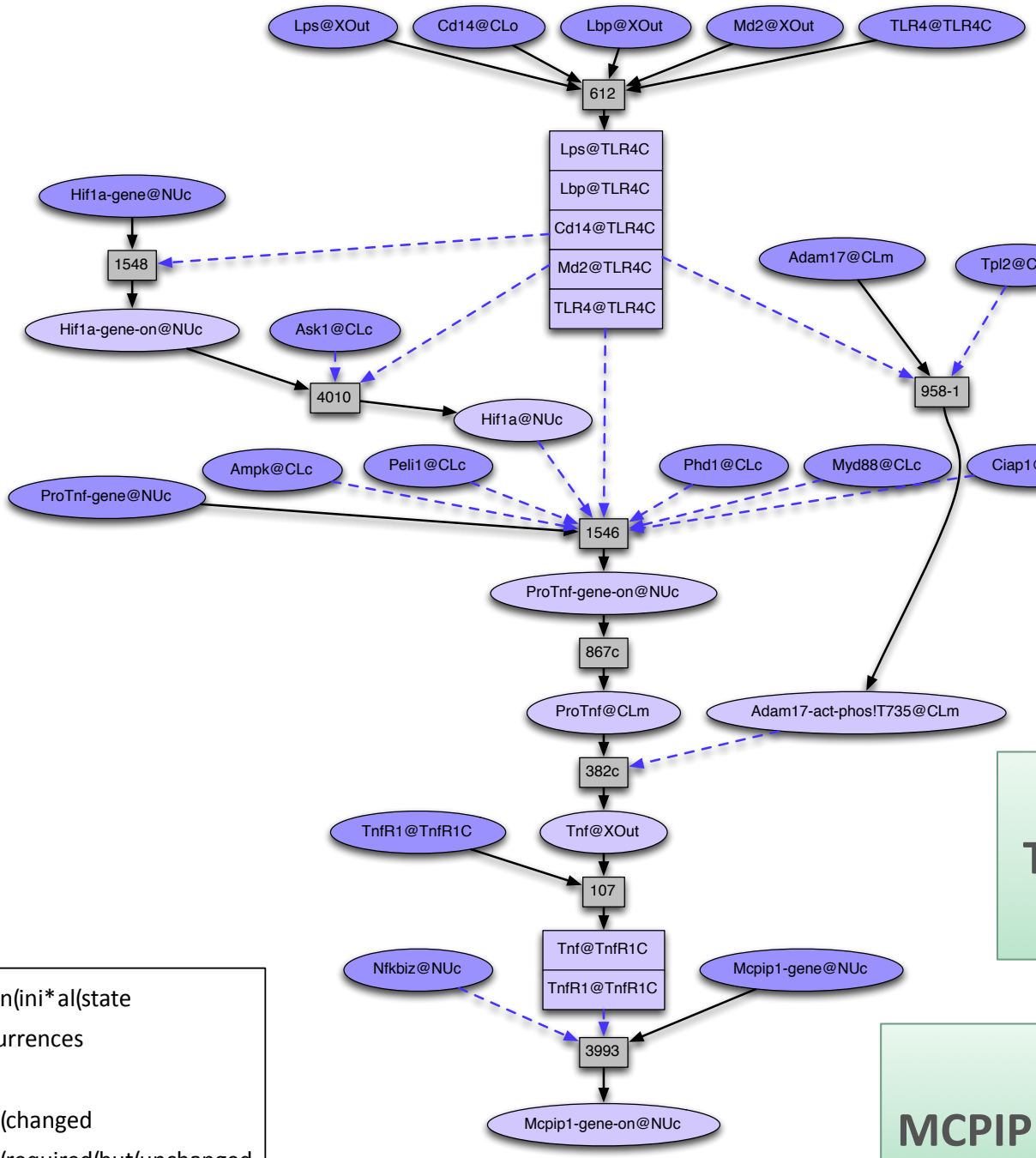
MCPIP1 modulates the immune response and inflammation by regulating the decay of specific mRNA molecules, including IL6, IL12B.

The Combined LPS and TNF Response Networks



PLA was used to fish out the path from LPS binding TLR4 to induction of the gene for MCPIP1.

Pathway Found



Rule 612
LPS binds to TLR4

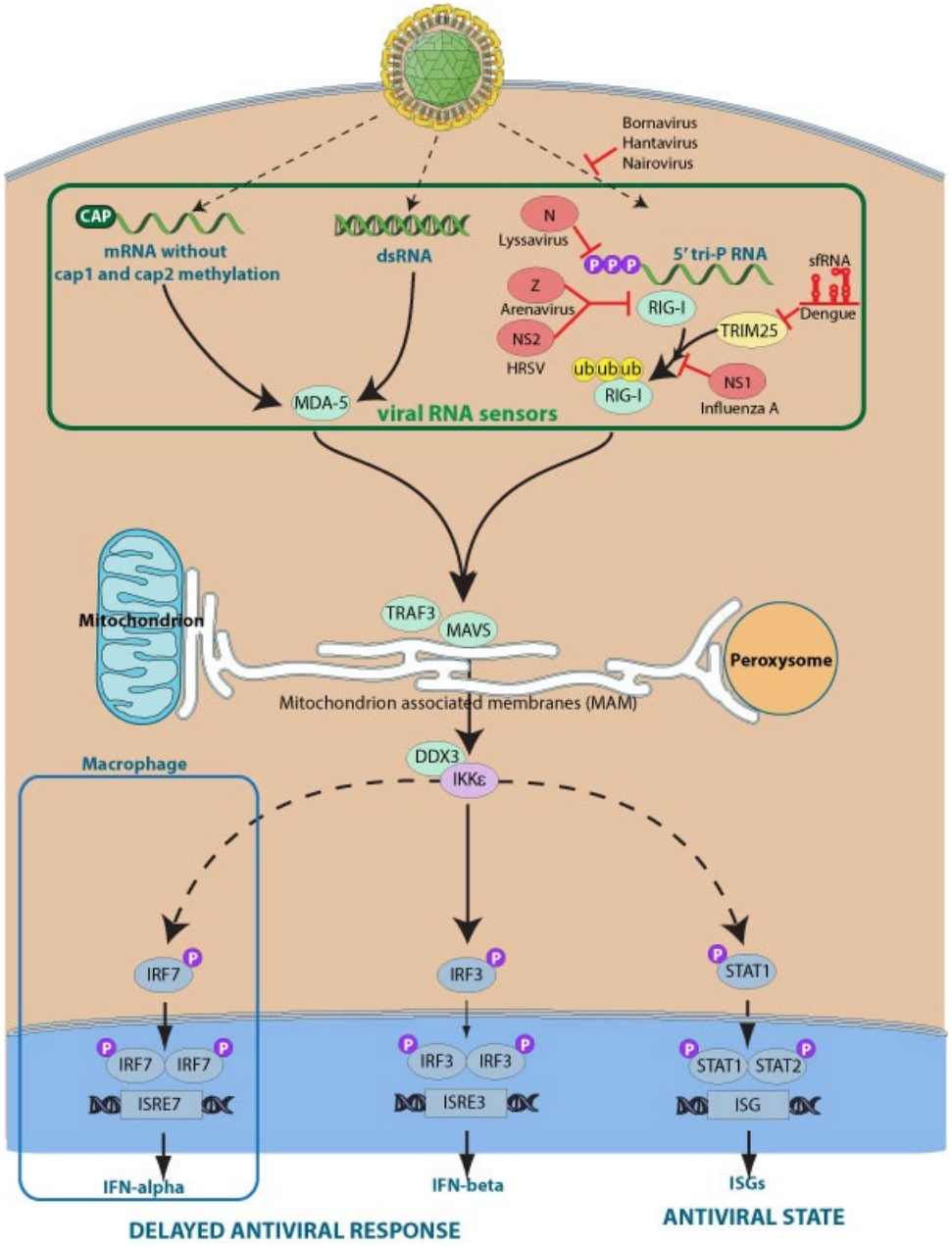
Rule 1546
ProTNF gene is turned on

Rule 382c
Tnf is released from cell membrane

Rule 3993
MCPIP1 gene is turned on

Occurrences(in(ini*al(state
Modified(occurrences
Rules
Occurrence(is(changed
Occurrence(is(required(but(unchanged

The Canonical RLR Pathway



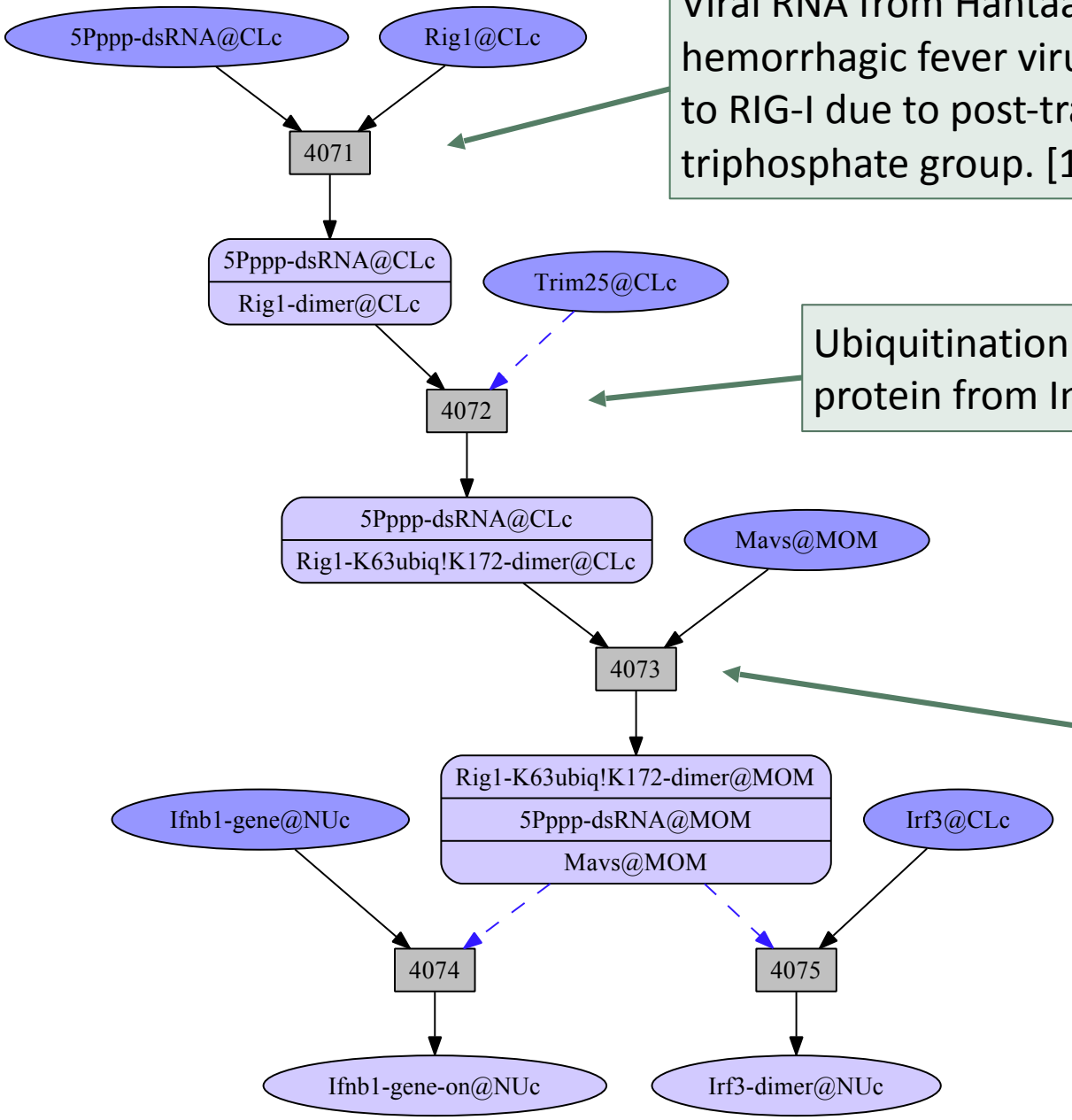
Source: ViralZone www.expasy.org/viralzone, Swiss Institute of Bioinformatics

Cellular RIG-like receptors belong to an innate sensor pathway that recognize RNA virus products and activates cellular antiviral state. Upon viral infection, RIG-I recognizes viral "foreign" RNA and triggers intracellular signaling events that induce innate immunity.

It is in the virus's best interest to avoid activation of this pathway. TOSV and SFTS NSs proteins (under study) bind RIG-I and hijack immune signaling.

This picture is fine as long as you are not interested the underlying interaction details.

Evasion of the Host Ifnb1 Response by Viruses



Viral RNA from Hantaan virus, Crimean-Congo hemorrhagic fever virus and Borna disease does not bind to RIG-I due to post-transcriptional cleavage of the 5' triphosphate group. [1]

Ubiquitination of RIG-I is inhibited by Ns1 protein from Influenza A [2]

Binding of RIG-I to Mavs is inhibited by RING finger protein Z from Junin mammarenavirus [3] and Ns2 protein from Ovine respiratory syncytial virus [4].

Summary

- Pathway Logic features include
  - formal representation of experimental evidence
  - links from symbols representing biomolecules to external reference resources
  - links from rules to supporting evidence.
  - assembly of models and pathways by symbolic reasoning
- Challenges going forward
  - scaling — curation, analysis, visualization
  - automating semi-quantitative reasoning
  - automating inference of rules from datums
  - fuzzy matching for assembling connected networks

PL extended team (past and present)*

- Robin Donaldson
- Steven Eker**
- Merrill Knapp**
- Keith Laderoute**
- Pat Lincoln**
- Ian Mason
- Jose Meseguer**
- Huaiyu Mi (USC, PANTHER)
- Anupama Panikkar
- Andy Poggio
- Malabika Sarker
- Carolyn Talcott
- Maneesh Yadav

*Alphabetical order
**Founders

Thats all folks!