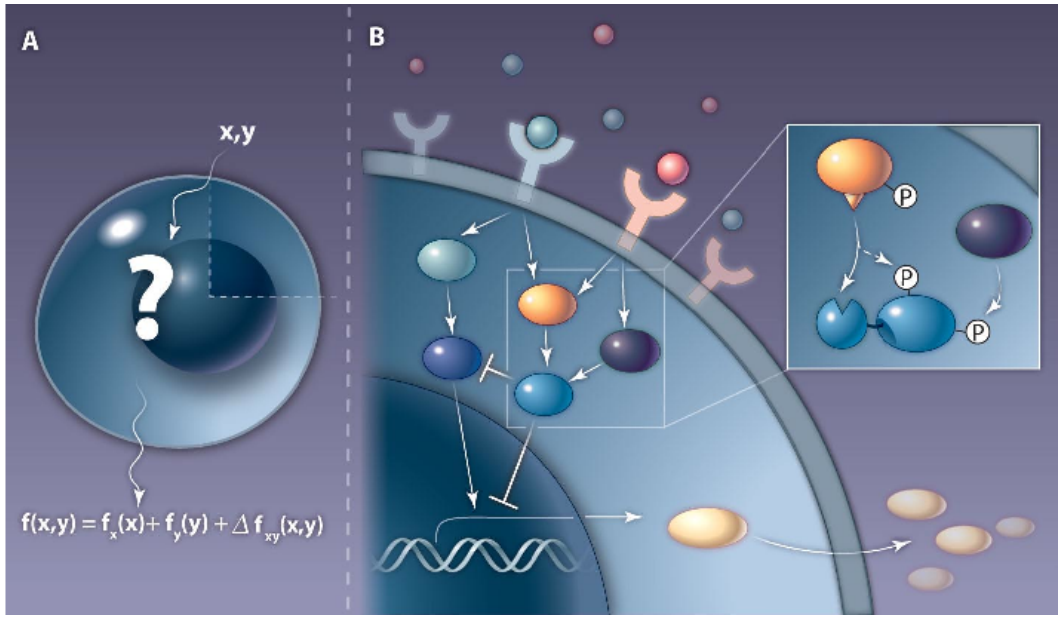
The atomizer: RNM2RBM

Jose Juan Tapia
James R. Faeder
University of Pittsburgh

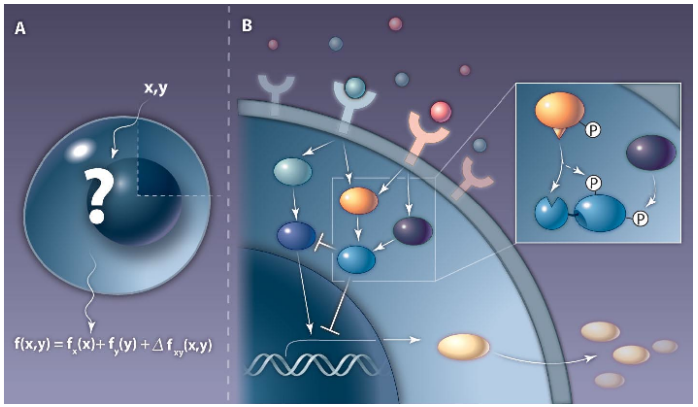
Rule-based models are …

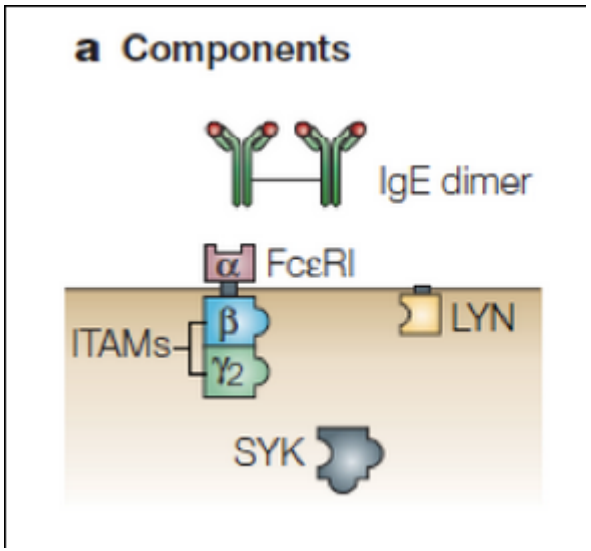


W. S. Hlavacek and J. R. Faeder, Sci. Signal. 2, pe46 (2009).

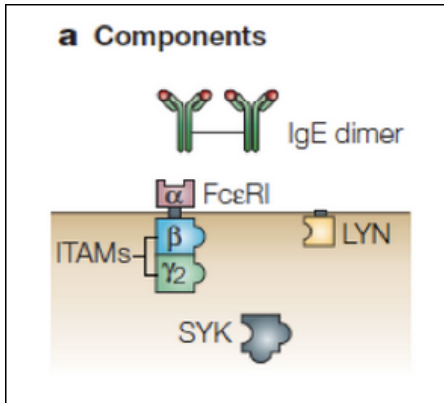
Reaction-network modeling
Rule-based modeling

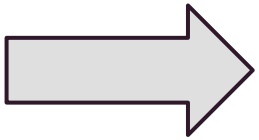
RNM RBM

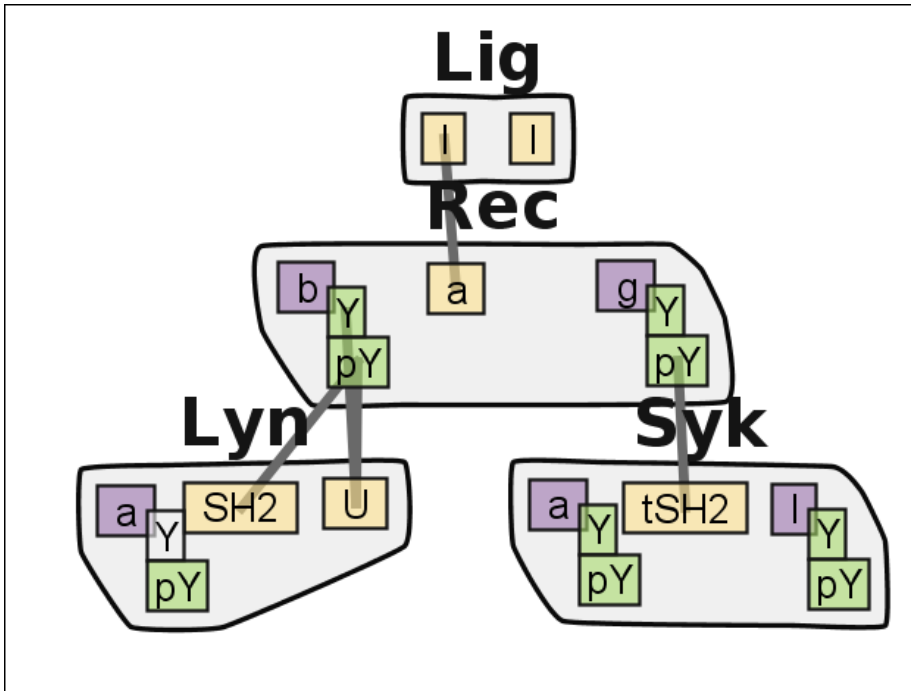




Rule Based Modeling (RBM)







Contact map

Terminology comparison

RBM

- Molecule types: Defines a category of molecules and its full component structure
- Molecules: Indivisible entities that associate with other indivisible entities
- Components: Molecule's functional attributes
- Species: Unique configuration of one or more molecules

RNM

- Species type: A type of entities that can participate in reactions.
- Species: A pool of entities of the same species type located in a particular compartment

Syntax comparison

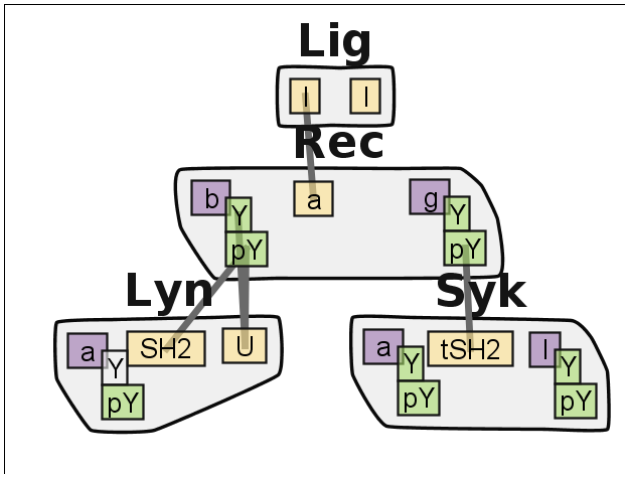
RNM

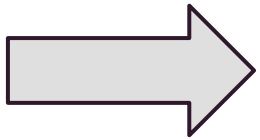
```
A + B -> A_B
```

RBM

```
A(b) + B(a) -> A(b!1).B(a!1)
```

Network expansion





SBML File

354 species
3680 rxns

19 rules
4 molecules types

This phenomenon is known as combinatorial complexity

Rule-based modeling

- Rich syntax that allows the modeler to encode structural and contextual information
- Very scalable
- Syntax may be overkill for smaller models or phenomenological models.

Reaction Network Modeling

- Well understood theory of differential equations based chemical kinetics.
- Streamlined representation suitable for medium-sized models.
- Large body of models encoded using RNM
- Biological structural and contextual information is lost.

Our goals are

Find a way to recover structural and contextual information that is no longer explicit in RNM models.

Obtain an RBM representation of the same model with this information.

Understand and build upon years of RNM modeling knowledge from an RBM point of view in a semi-automated way.

Presenting…

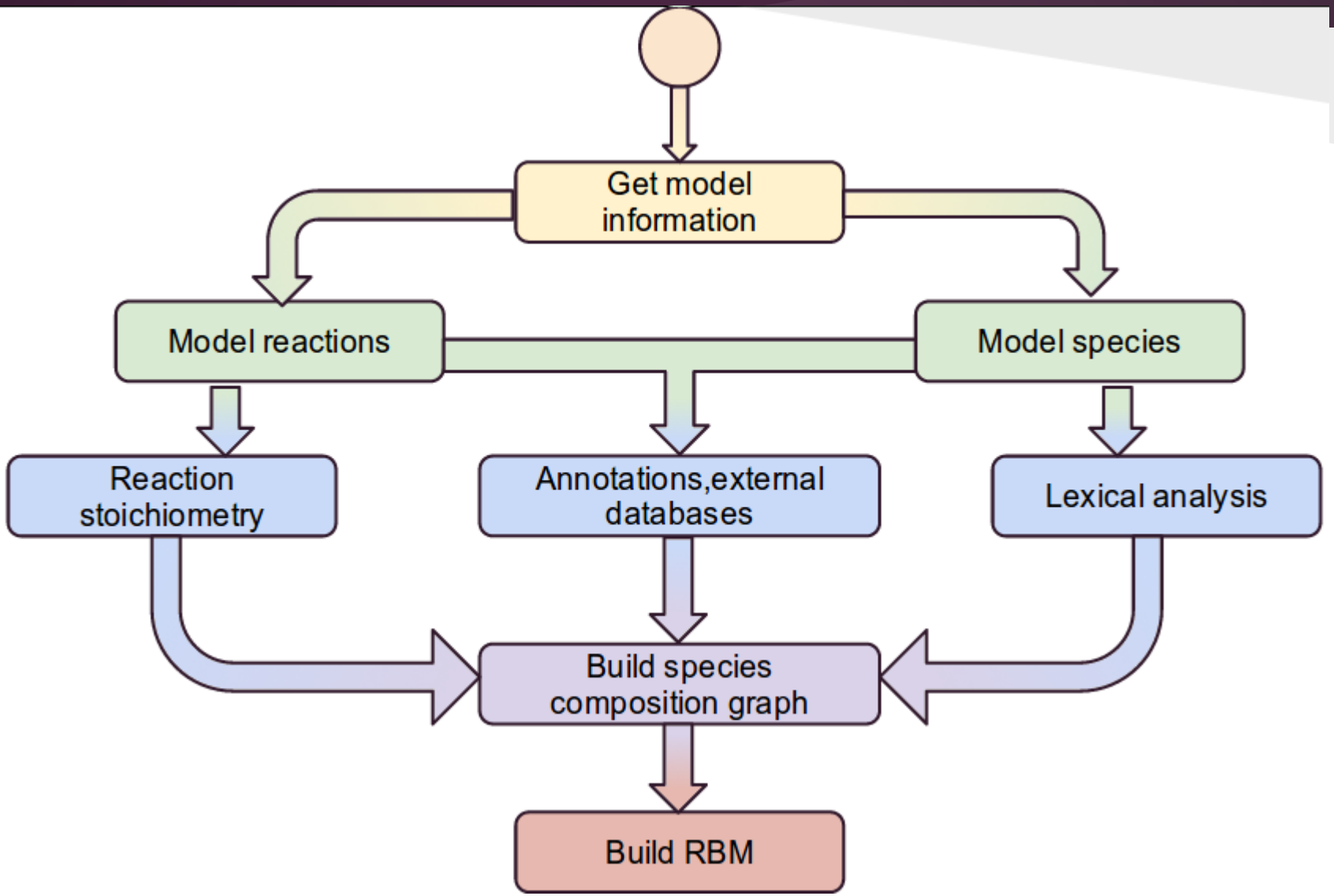
The

Atomizer

RBM

RNM

How does it work?



Toy example

```
A + B <-> A_B
B -> B_P
A + B_P -> A_P_B
```

Atomizing in a nutshell

- Reaction stoichiometry information

`A + B <-> A_B` *Complexation*

- Lexical analysis

`B -> B_P`

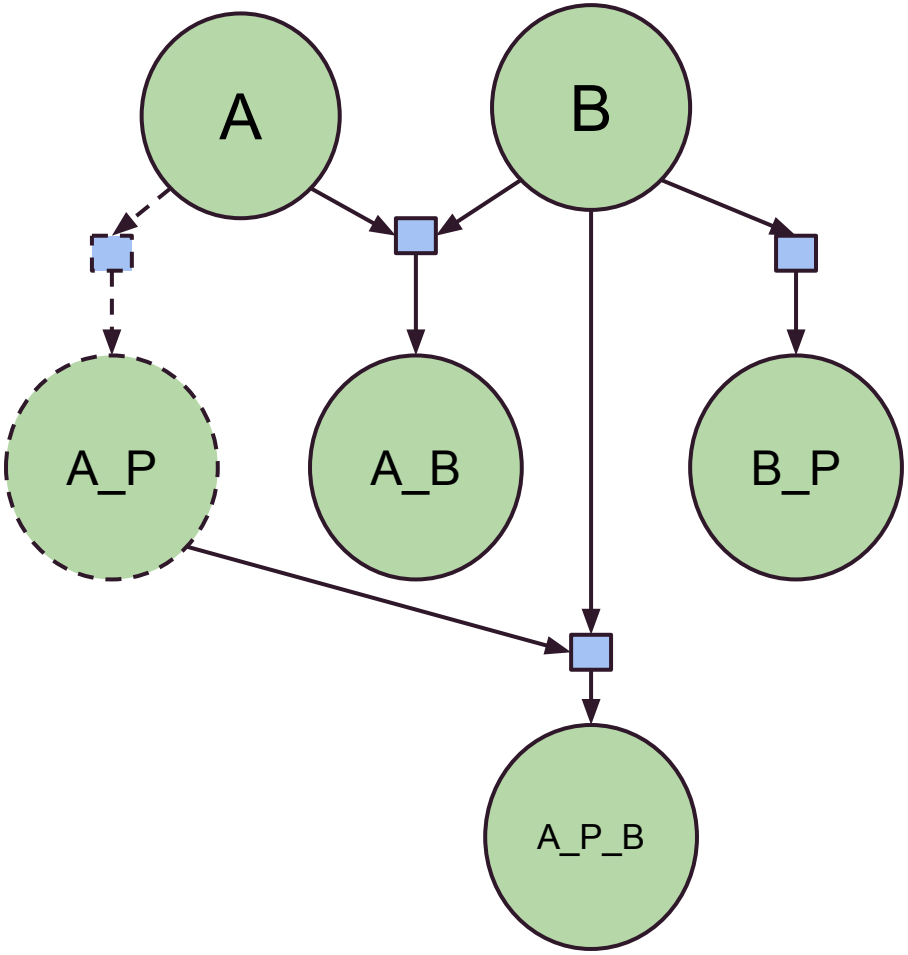
Phosphorylation

`A + B_P -> A_P_B`

Complexation and phosphorylation transfer

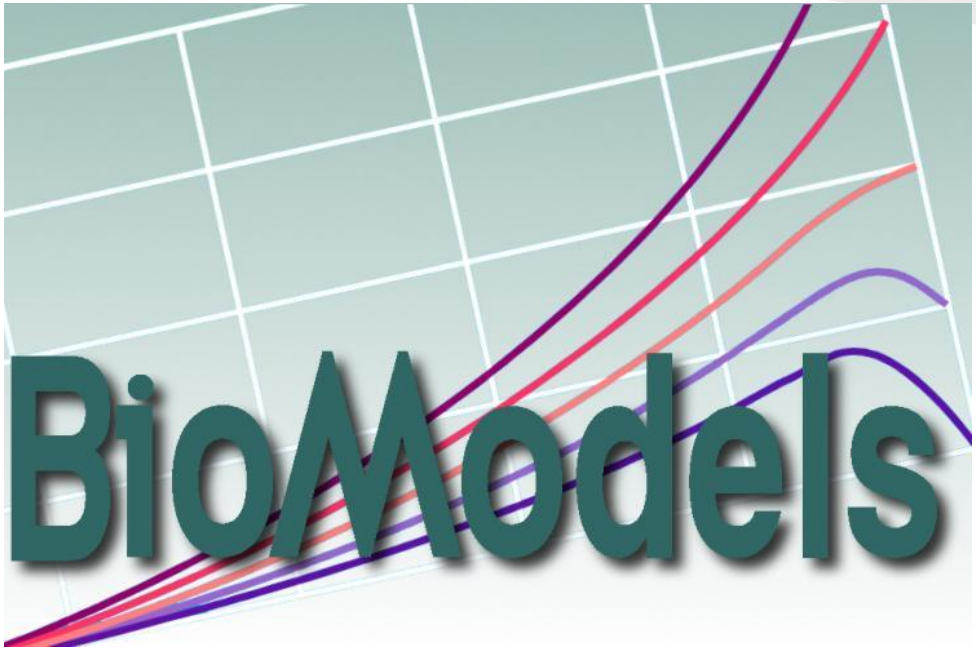
- Annotation information,interaction databases

Compositional Graph

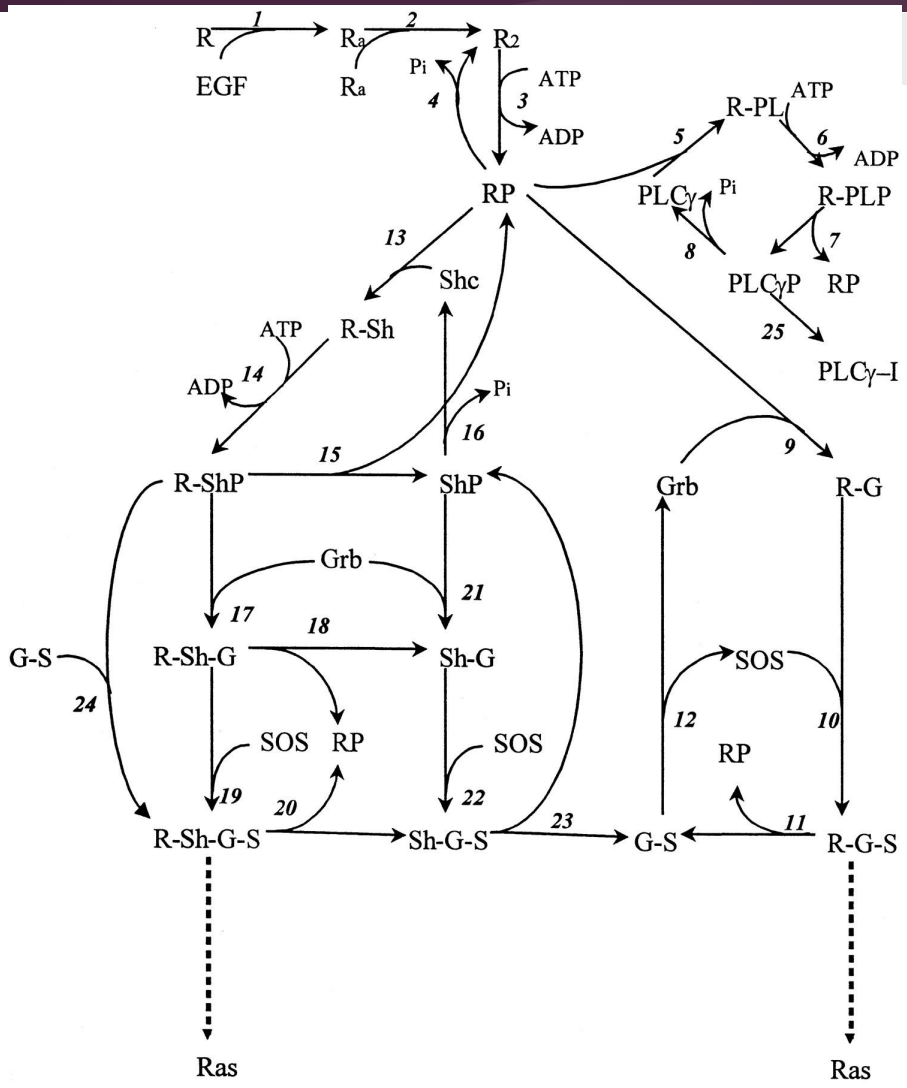



```
A(b,phospho~U~P)
B(a,phospho~U~P)
```

Atomize BioModels

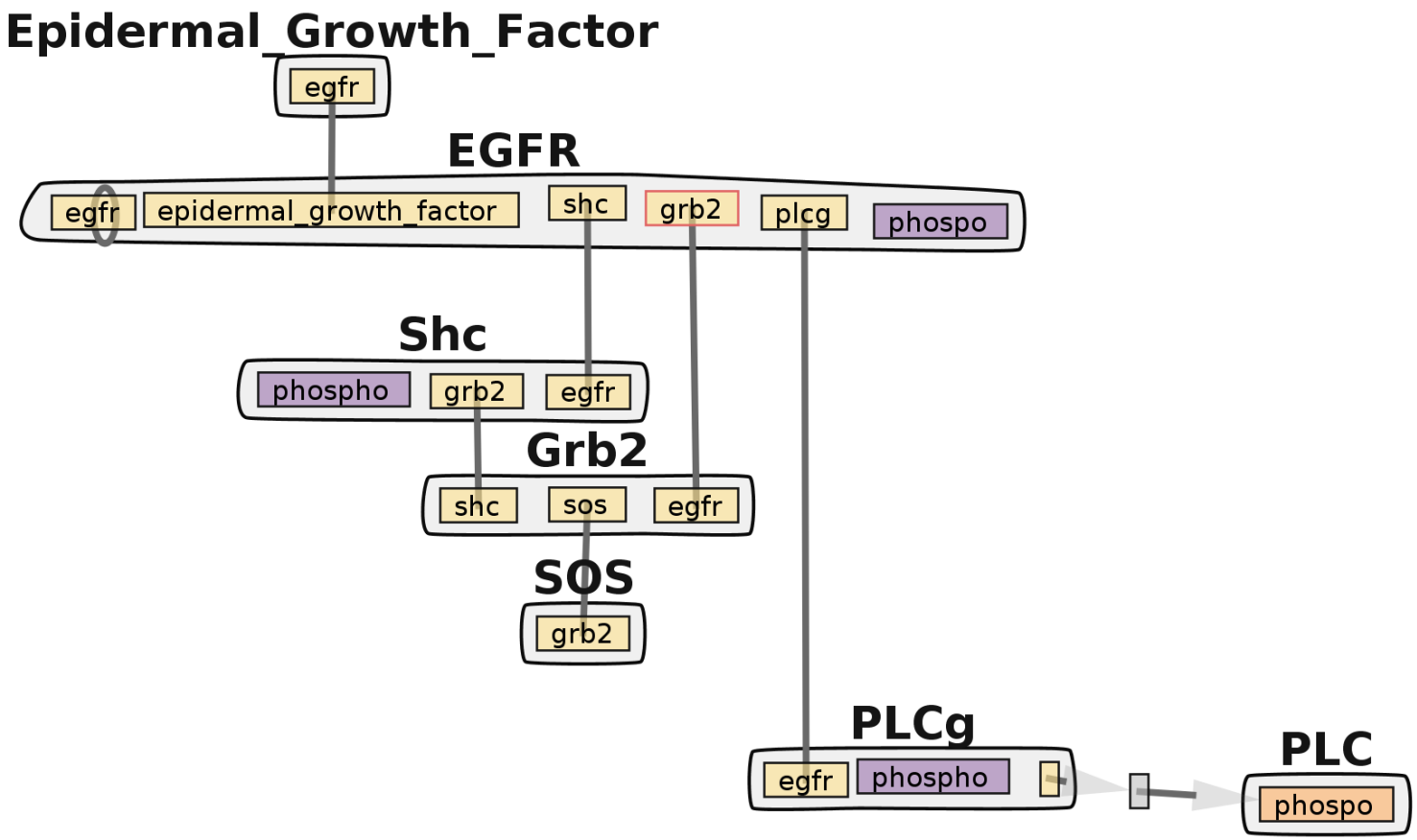


Epidermal Growth Factor

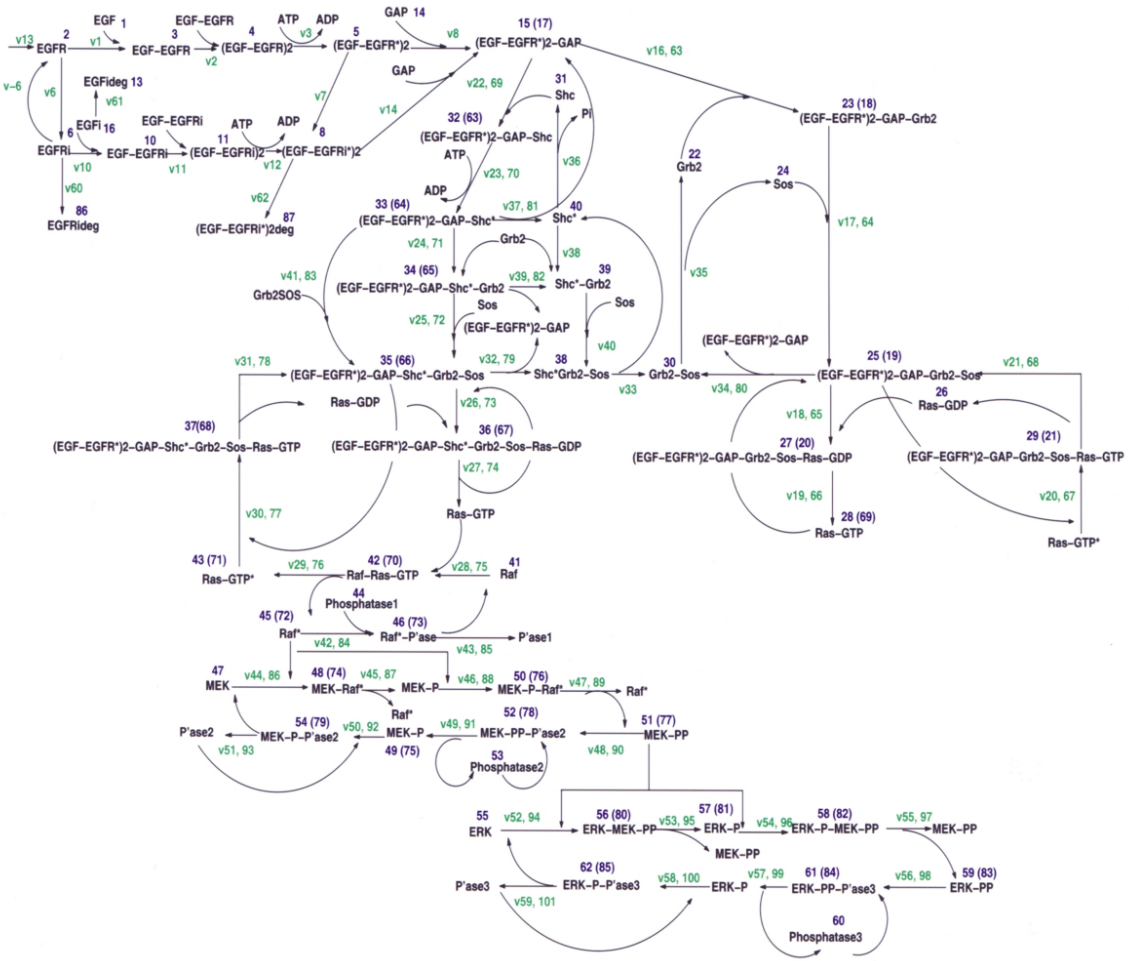


Kholodenko 1999

BioModels 48 atomized

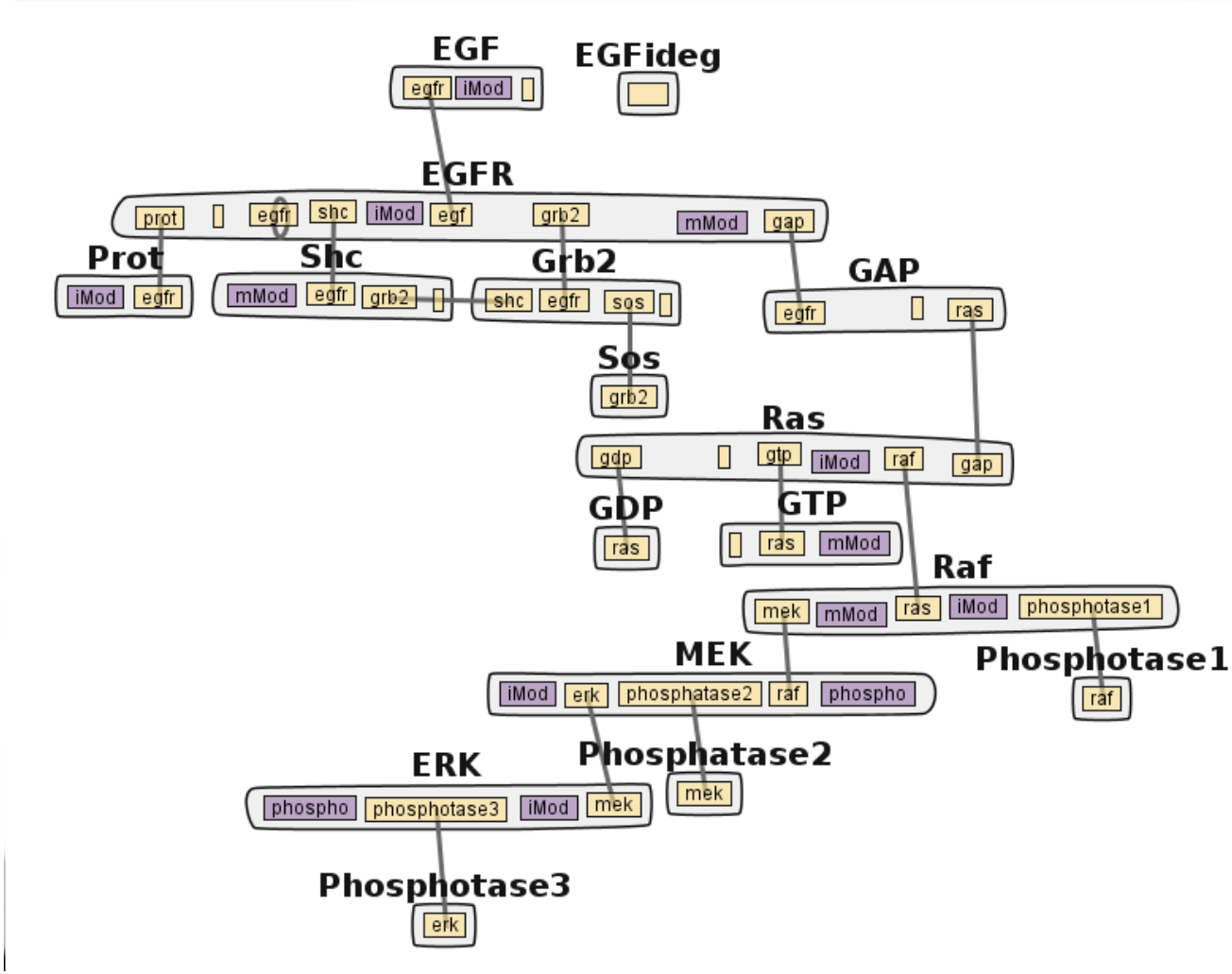


Bigger example



Schoeberl et. al.

BioModels 19 atomized



Molecule composition highlights: BioModels 109 (Cell cycle model)

Model annotation tells us that…

Cdk1Y10:

- Cyclin-dependent kinase

Cdk1Y11:

- Cyclin-dependent kinase
- Cyclin A

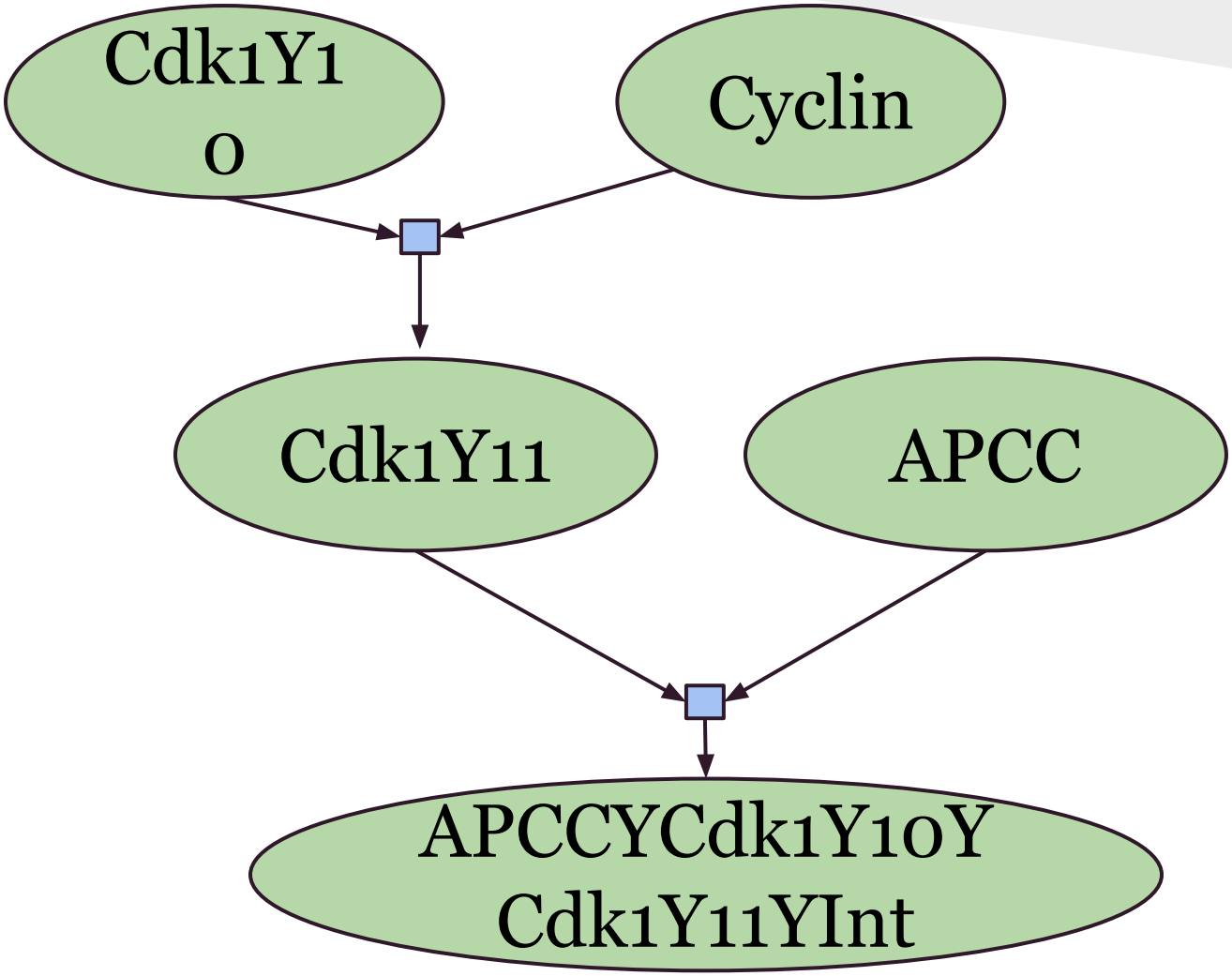
APCC:

- ubiquitin ligase

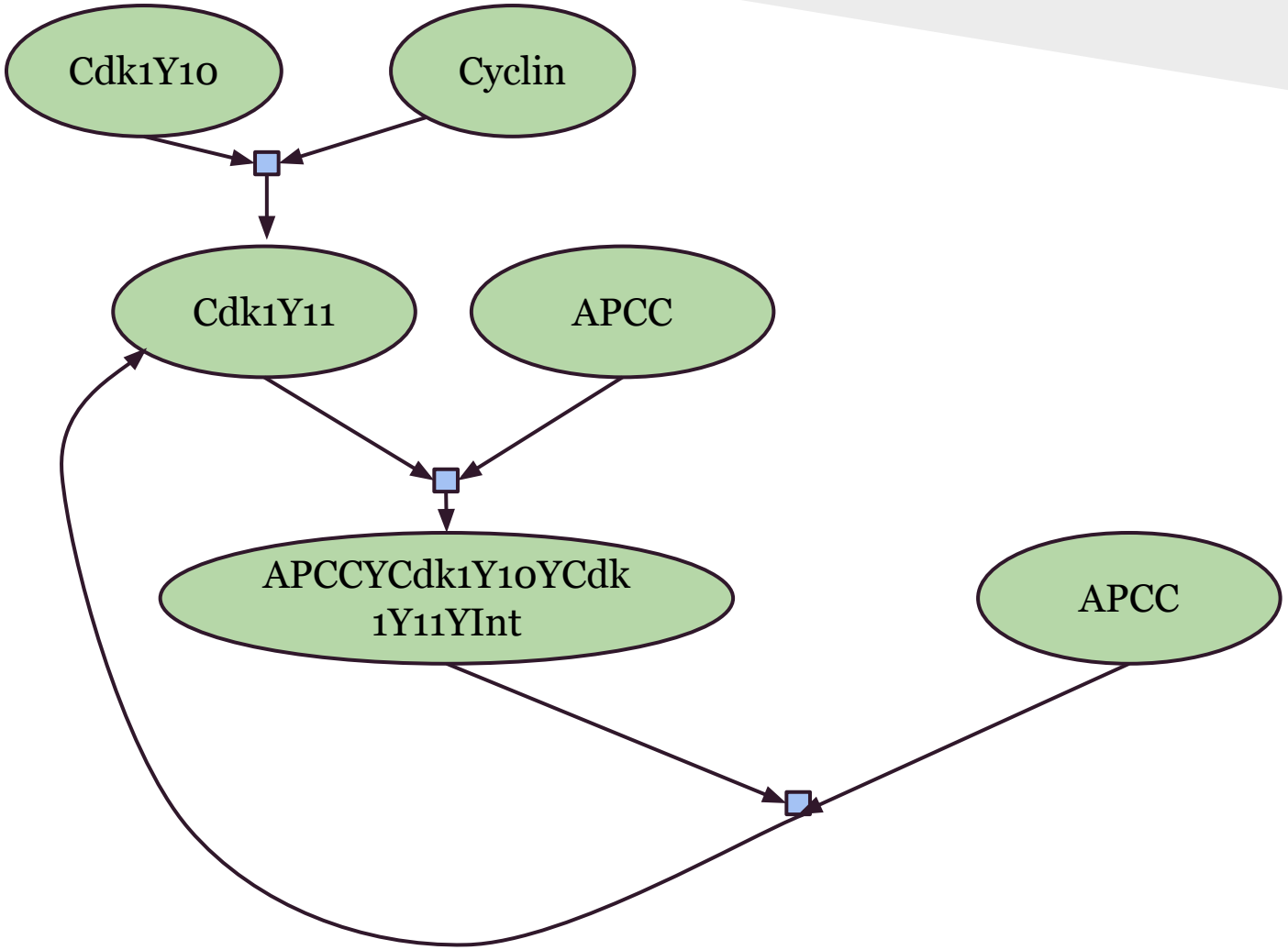
APCCYCdk1Y10YCdk1Y11YInt:

- Cyclin-dependent kinase
- ubiquitin ligase
- Cyclin A

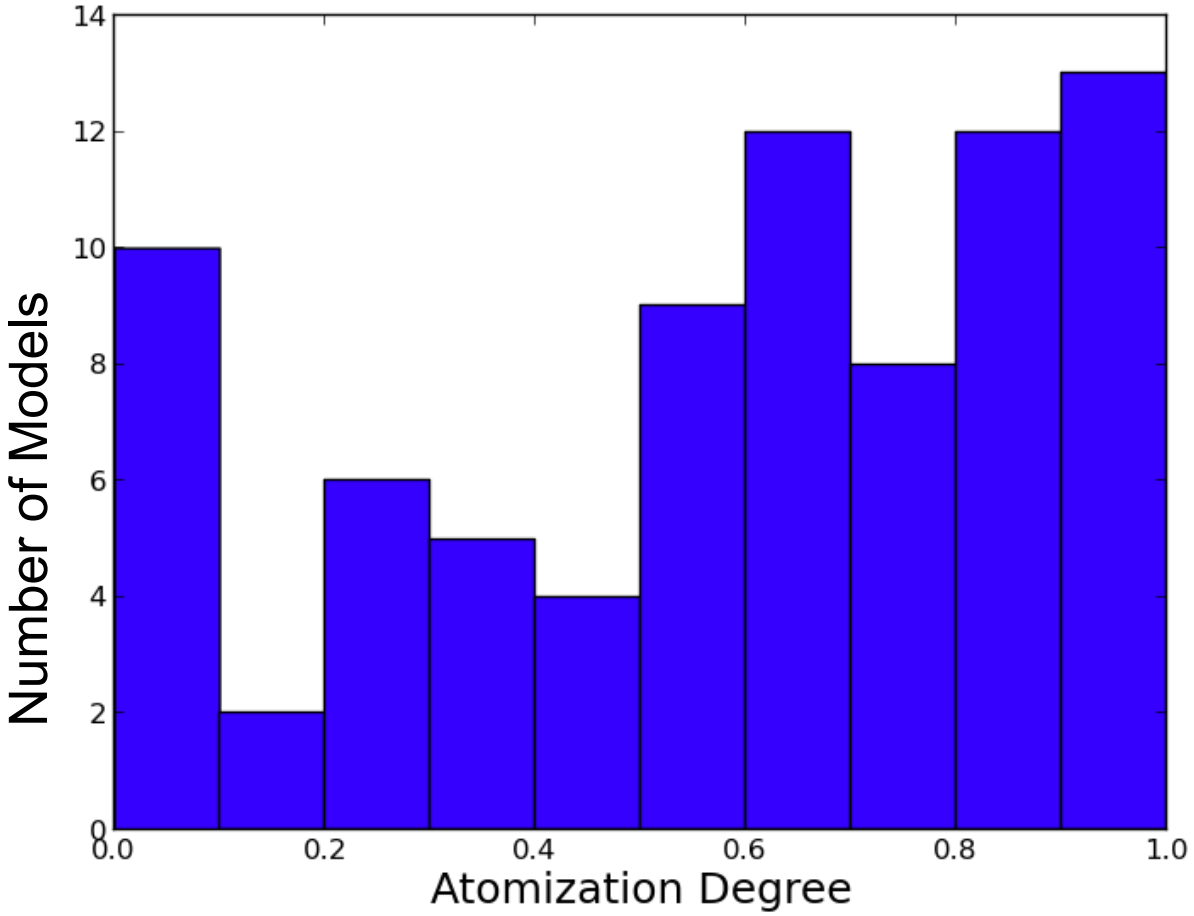
Model composition (BM 109)



And then… ??


Cdk1Y10
Cyclin
Cdk1Y11
APCC
APCCYCdk1Y10YCdk
1Y11YInt
APCC

Atomization degree = structured / total SBML species



Models with >20 reactions

SBML Limitations

Atomization works best with ODE-based models.

Molecular structure, cooperativity analysis and such can still be obtained for models with non ODE based dynamics (events, rate rules, delays, etc) however their simulation can not always be obtained.

How do I compare these models?

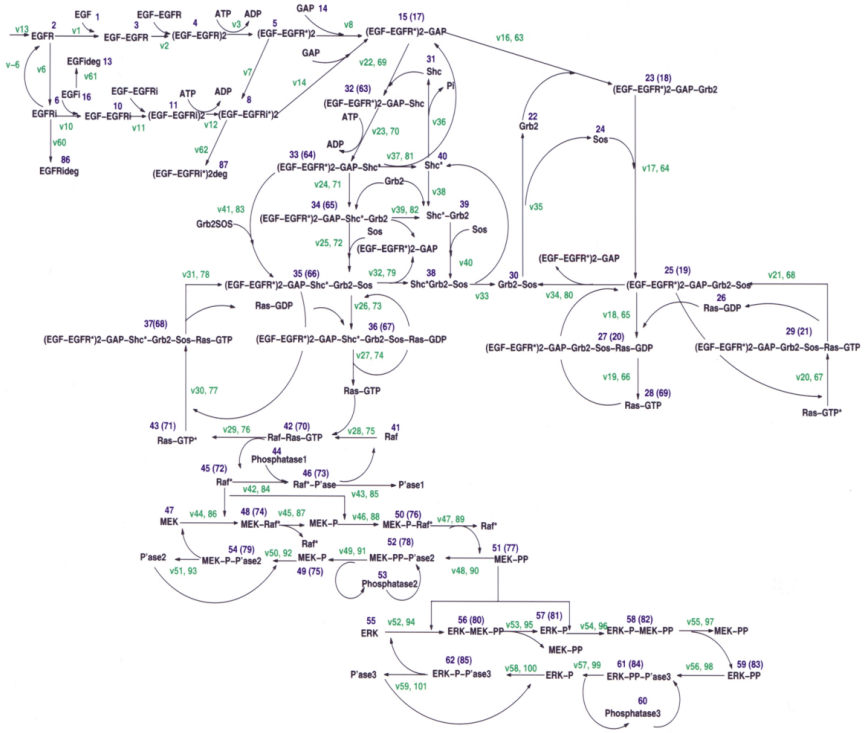


Image taken from Schoeberl et. al.

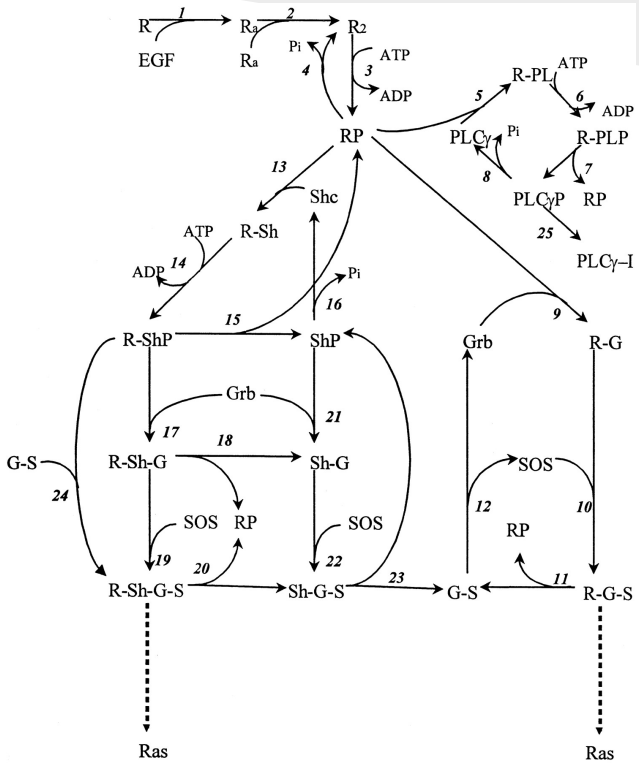


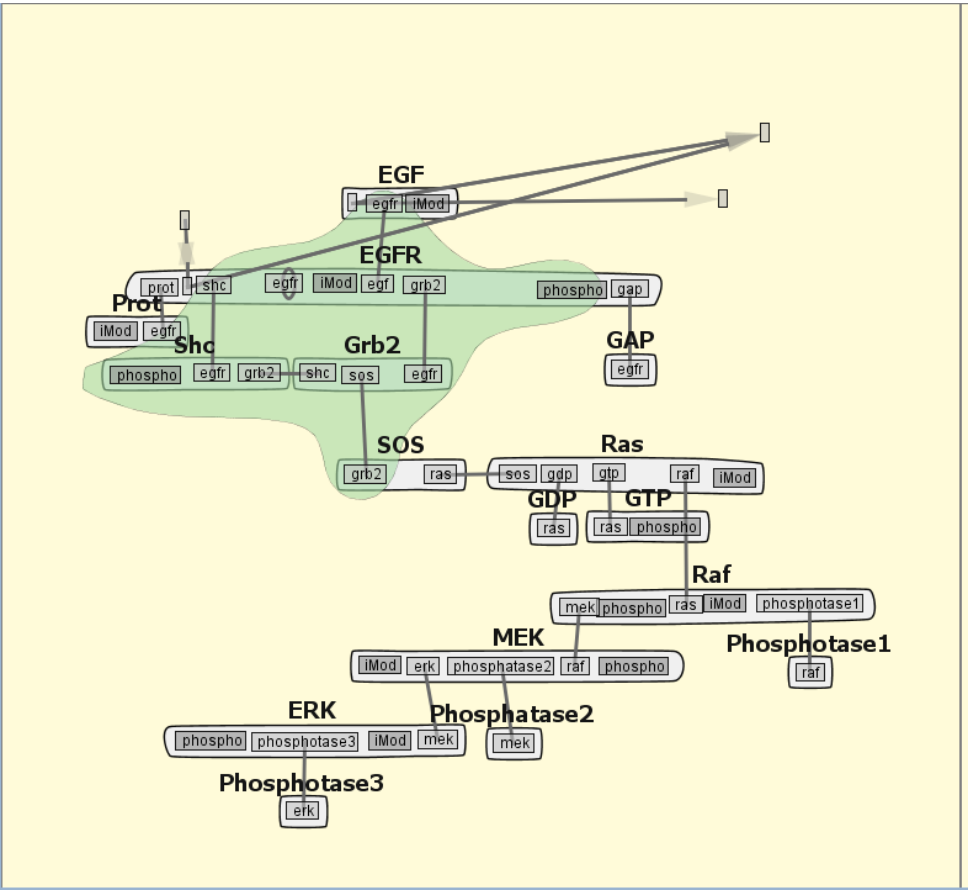
Image taken from Kholodenko et. al.

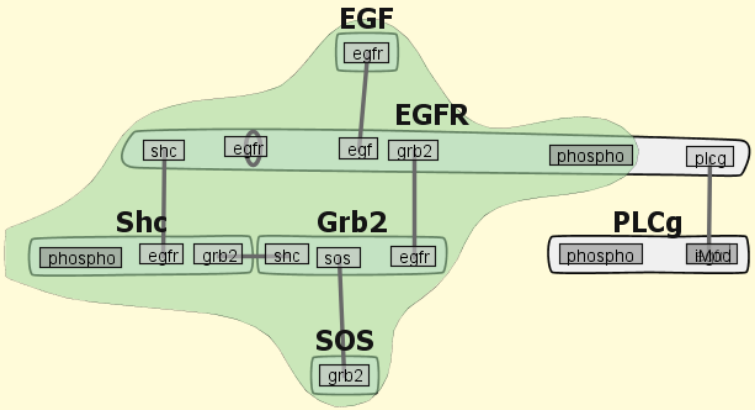
Challenges for model alignment

- Similarities between models (species mapping)
- Relationships and references within elements in the same model. (compositional network)

A successfully atomized model resolves a model compositional structure.

Atomized model alignment

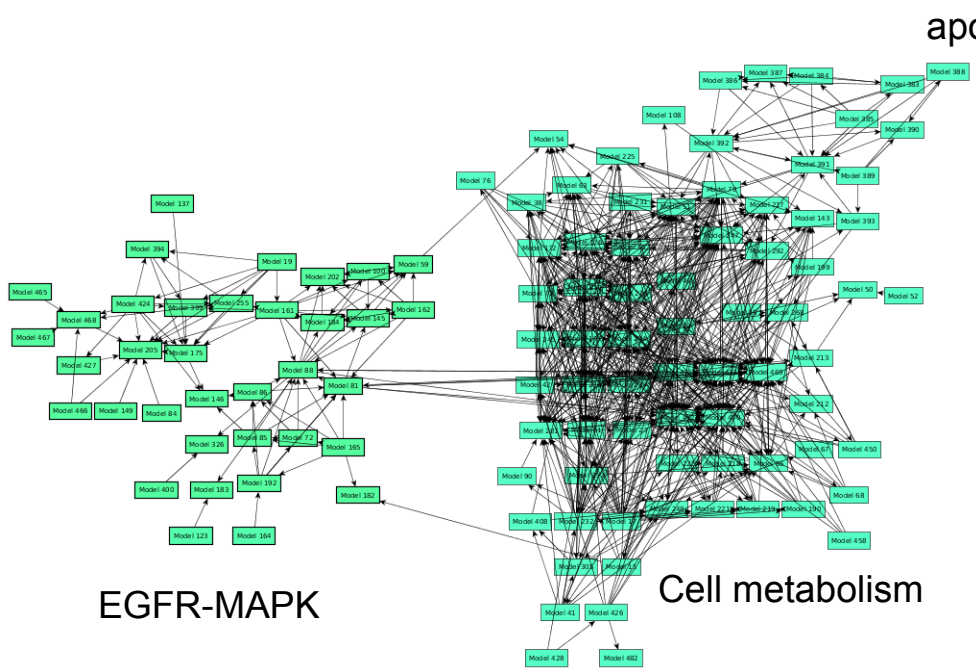




MOSBIE: A Tool of Comparison and Analysis of Rule-Based Biochemical Models
John E. Wenskovitch Jr., Leonard A. Harris, Jose-Juan Tapia, James R. Faeder and G. Elisabeta

Model interaction(each node is a BioModels model)

nodes: models
edges: models share >4 significant annotations

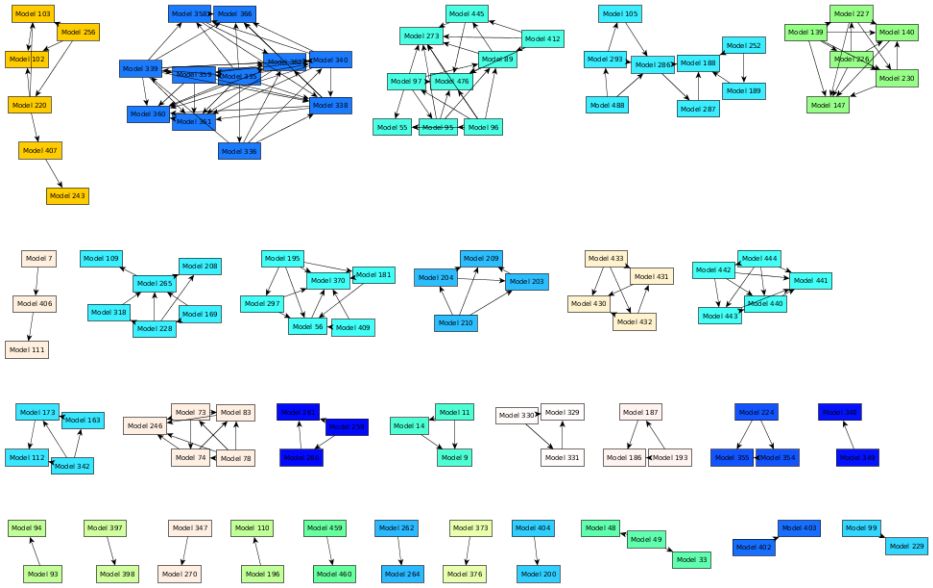


EGFR-MAPK

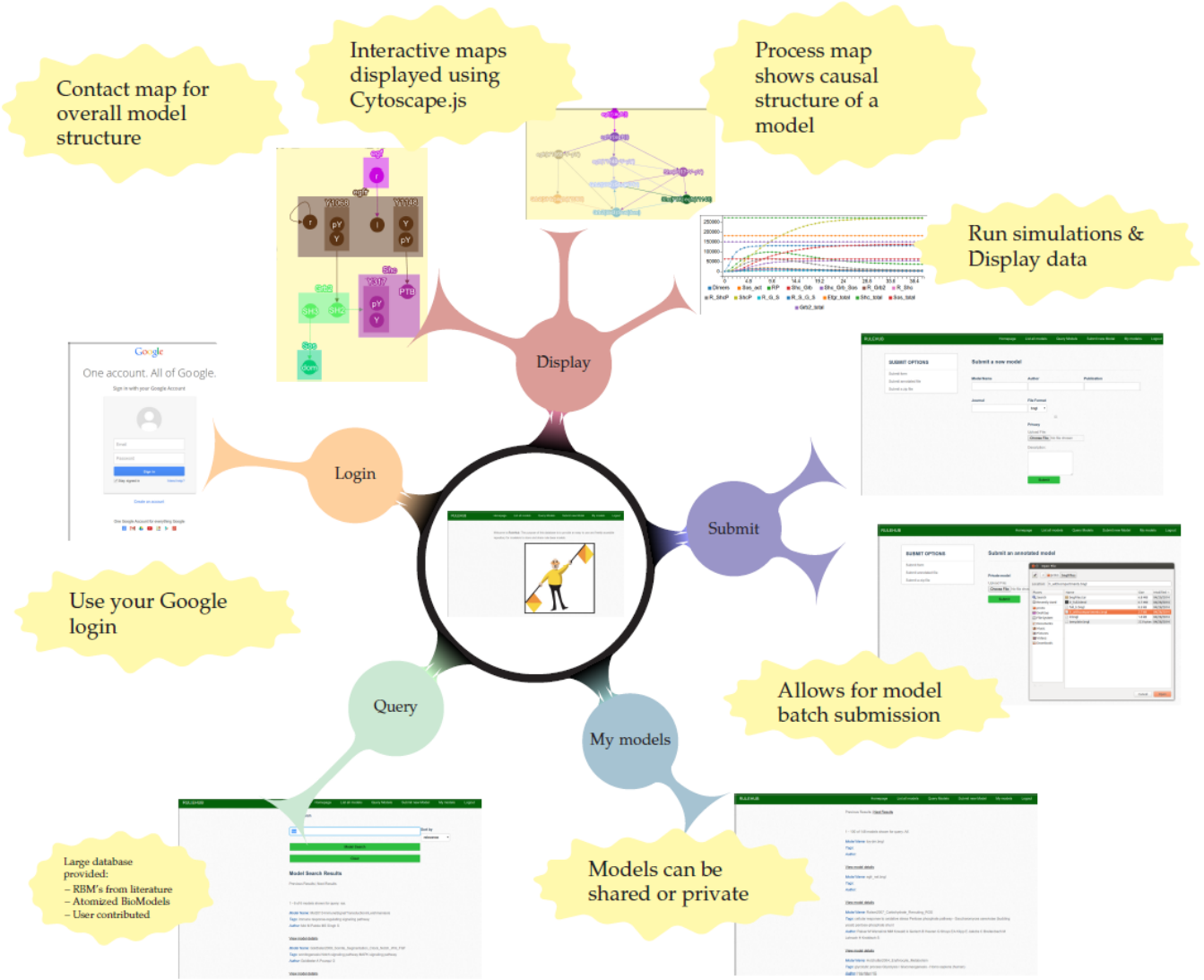
Cell metabolism

apoptosis

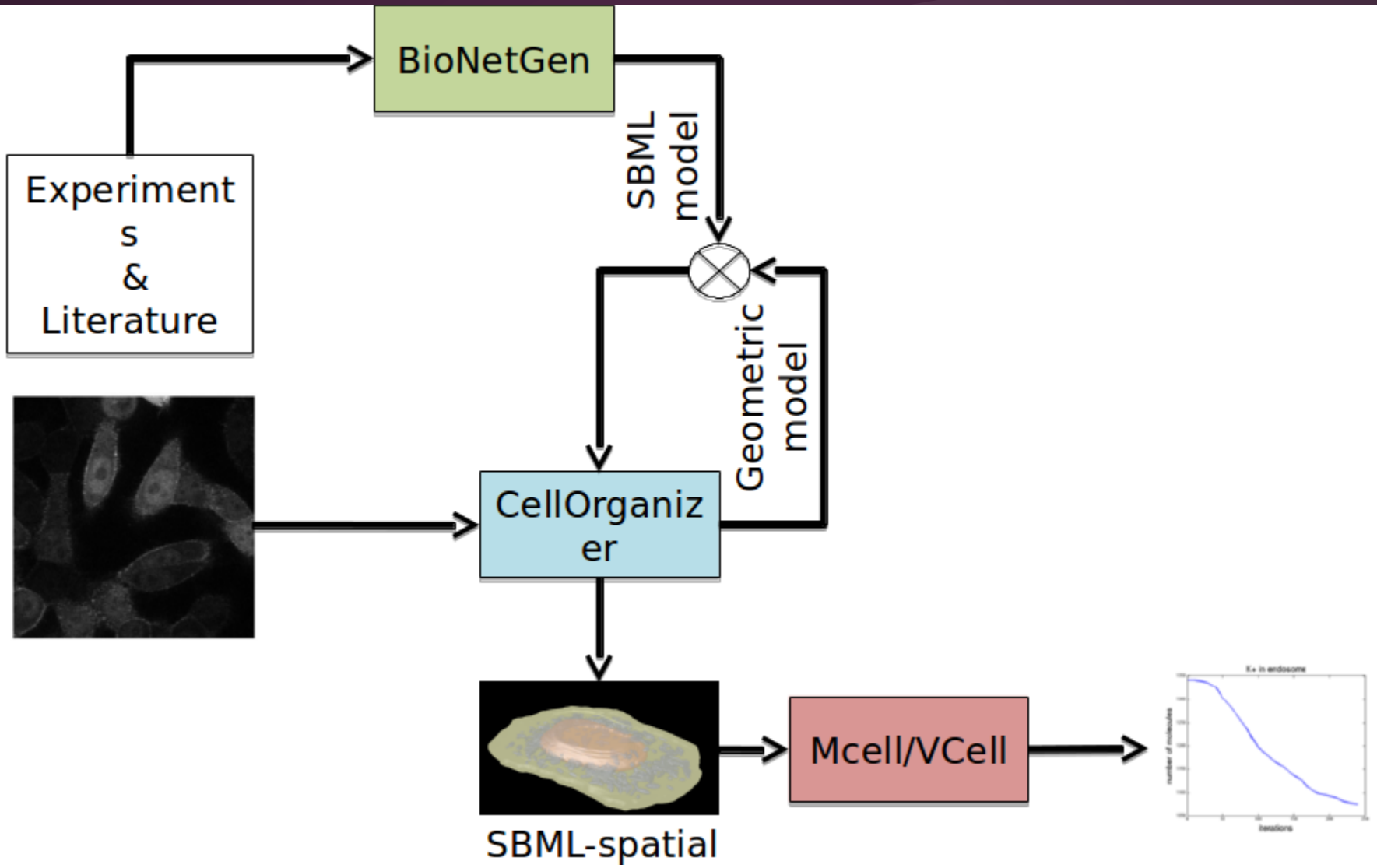
blood coagulation



RuleHub


Contact map for overall model structure
Interactive maps displayed using Cytoscape.js
Process map shows causal structure of a model
Run simulations & Display data
Display
Login
Use your Google login
Submit
Allows for model batch submission
Query
My models
Large database provided:
– RBM's from literature
– Atomized BioModels
– User contributed
Models can be shared or private

Pipeline


BioNetGen
Experiments & Literature
SBML model
Geometric model
CellOrganizer
SBML-spatial
Mcell/VCell

Now you can atomize too!

http://ratomizer.appspot.com/translate

BNG2.pl <sbmlfile.xml>

Thank you

Faeder Lab:

- Dr. James Faeder
- Dr. Justin Hogg
- Dr. Leonard Harris
- John Sekar

MCell Team

- Jacob Czech
- Markus Dittrich
- Devin Sullivan

NIH grant P41 GM103712 and NSF Expeditions in Computing Grant (award 0926181)

Q&A

Extra slides

Naming convention analysis

Patterns	Associated process	Example
$[`+P', `+p', `+P\_', `+\_P']$	Phosporylation	x → xP
$[`+PP', `+\_PP']$	Double phosporylation	xP → xPP
$[`+i']$	Internalization	x− > xi
$[`-\_n', `+\_c']$	Compartment transfer from nucleus to cytoplasm	x_n, x_c
$[`+\_ubiq']$	Ubiquitination	x− > x_ubiq
$['K', 'KK']$	Kinase, Kinase kinase	MAP, MAPK, MAPKK
$[`+H']$	Adding a hydrogen-related modification	NAD, NADH
$[`+R']$	Receptor	EGF, EGFR
$[`+c']$	[Cyclic version, cytoplasm, casp3 substrate]	x, cx
$[`+2', `+3', `+4']$	[Dimer, Trimer, Tetramer]-[Protein family]	x → x2

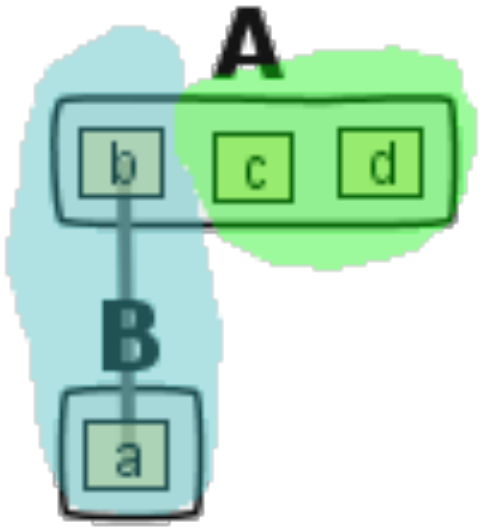
Naming convention analysis

+ P	27.4669509595
+ p	21.5778251599
- T+ D	9.0618336887
+ 2	7.4669509595
+ a	6.908315565
- D+ T	6.7356076759
- P+ M	5.5991471215
- n+ c	4.8614072495

Metric is the product of the number of times an annotation appears across the database multiplied by the percentage of models it appears in

Reaction center and context

A(b,c,d) + B(a) -> A(b!1,c,d).B(a!1)



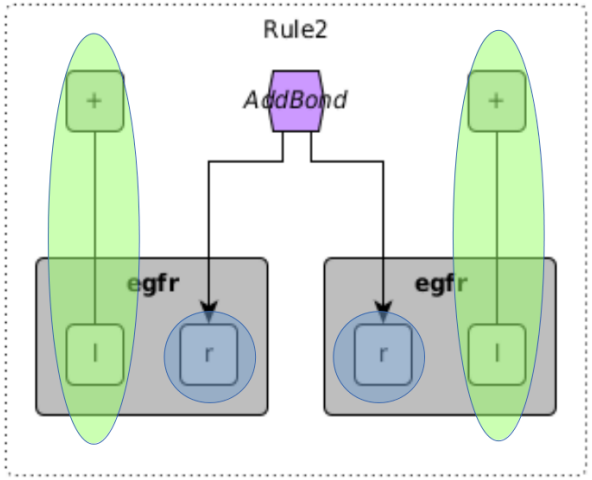
SBML Molecules contain minimal context information

```
A + B <-> C
```

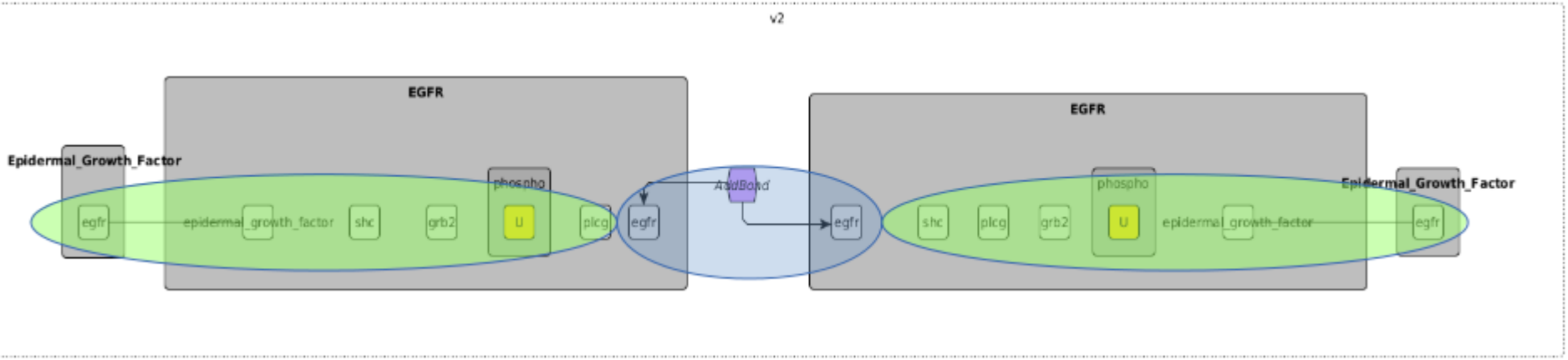
What is A? (Single molecule, umbrella name for a compound series of A molecules, etc).

Impossible to know without extensive annotation information. So we have to take them at face value.

Kholodenko's reaction context

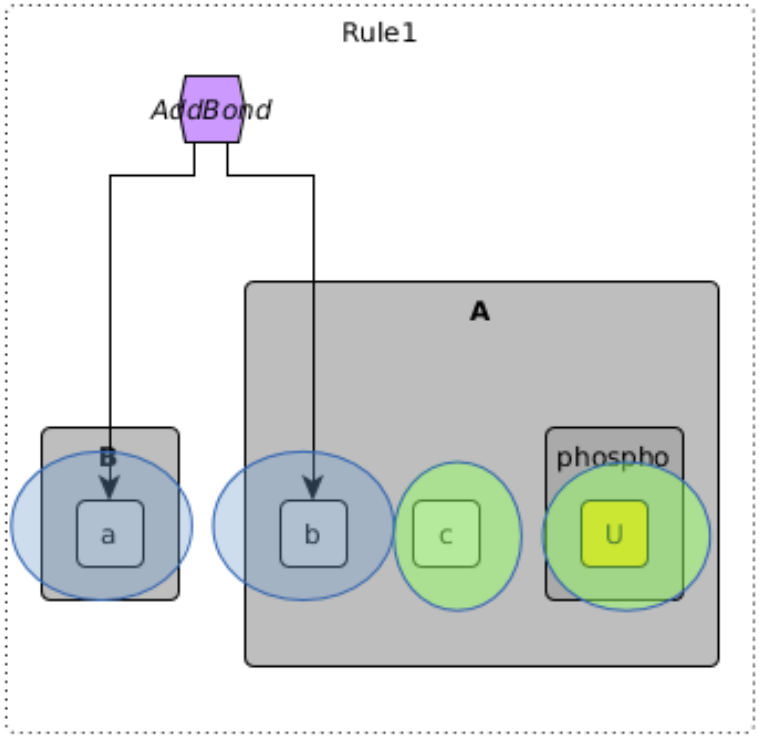


Manually constructed RBM version of Kholenko's model.

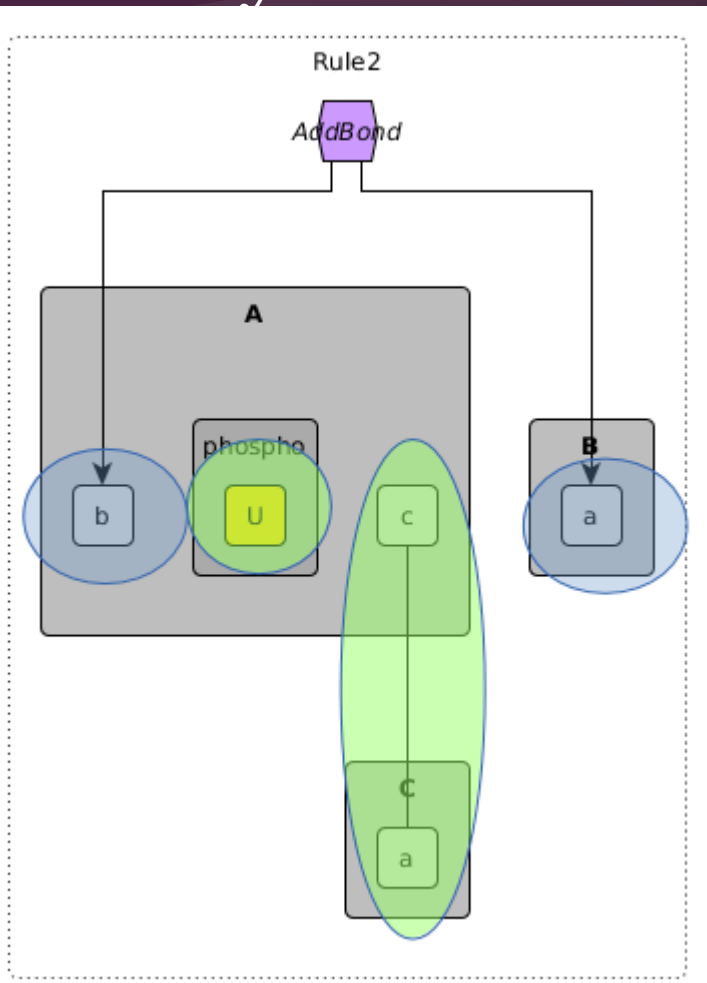


Automatically translated version of Kholodenko's model

Redundancy is the redundant way to redundantly go redundantly

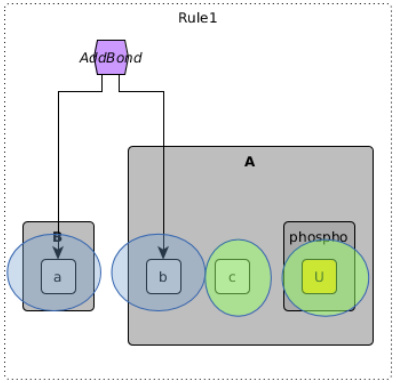


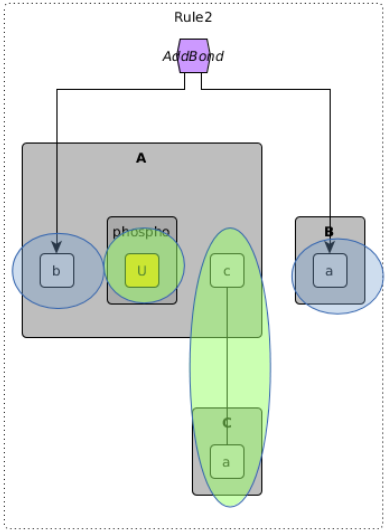
For ‘A’ and ‘B’ to bind all other states must be unoccupied.

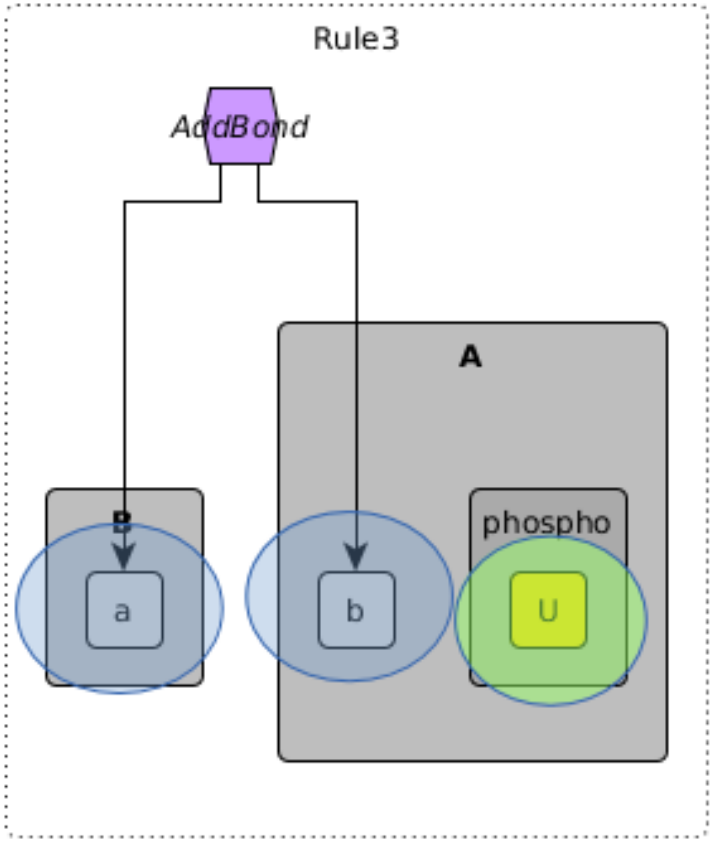


For ‘A’ and ‘B’ to bind ‘A’ must be bound to ‘C’ already

Context factorization







For A and B to bind it does not matter if C is part of the complex or not