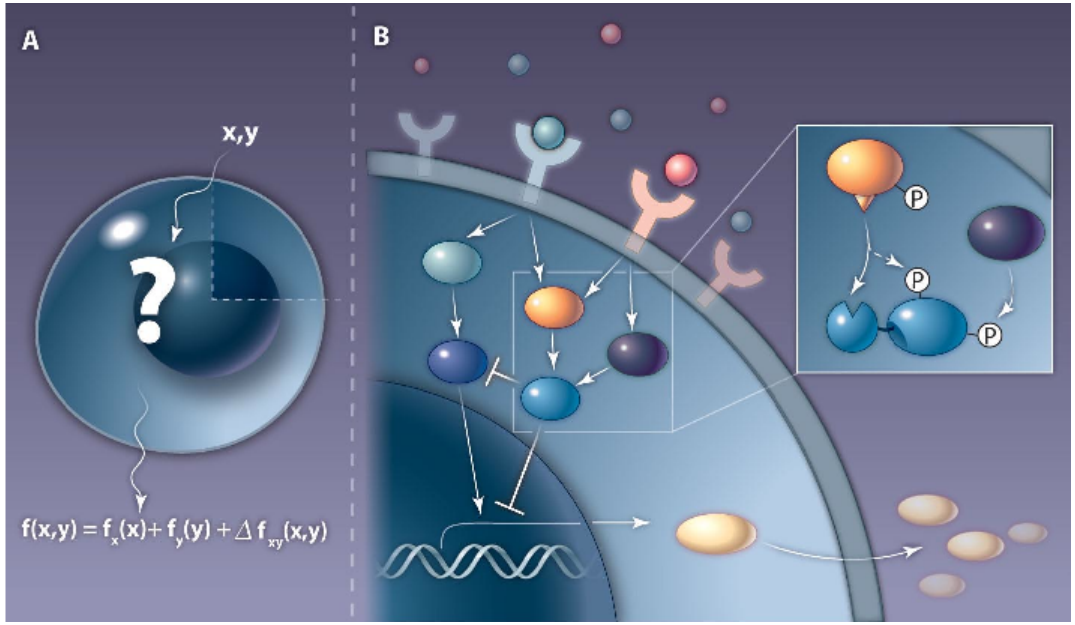
Atomizing BioModels

Transforming reaction-network models into rule-based models

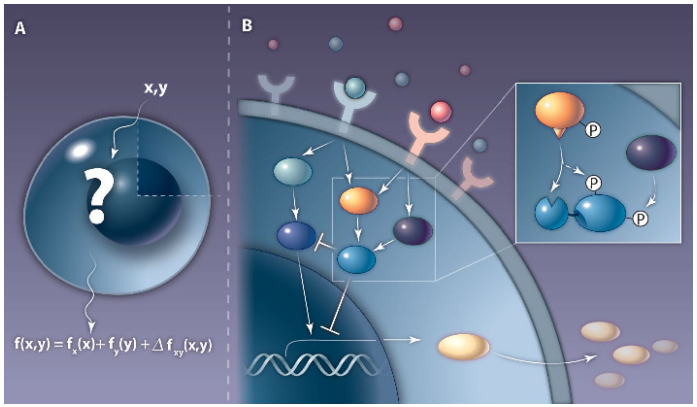
Jose-Juan Tapia
James R. Faeder

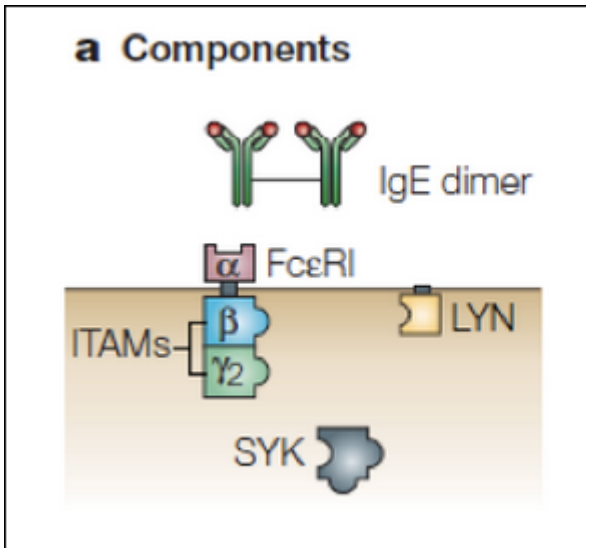
Rule-based models are …



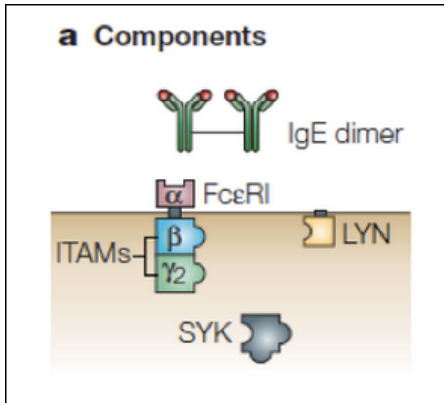
W. S. Hlavacek and J. R. Faeder, Sci. Signal. 2, pe46 (2009).

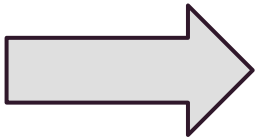
Mental model

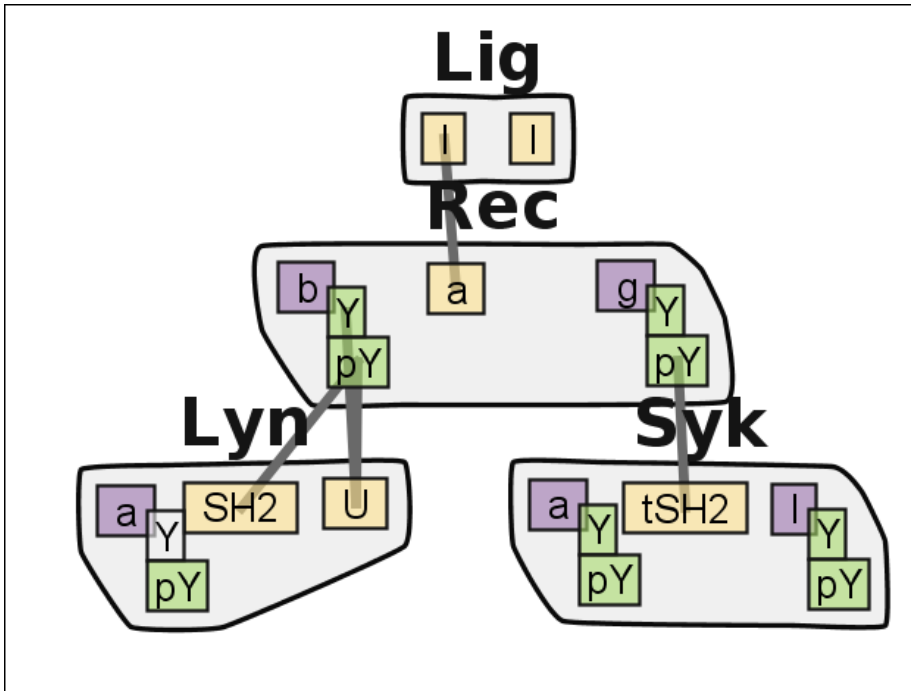




Rule Based Modeling (RBM)

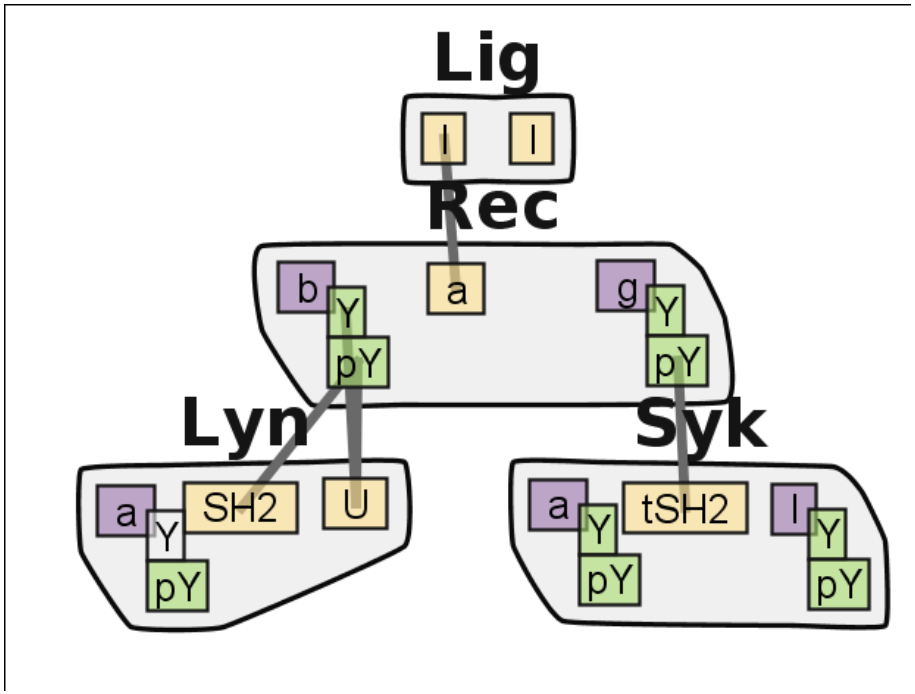






Contact map

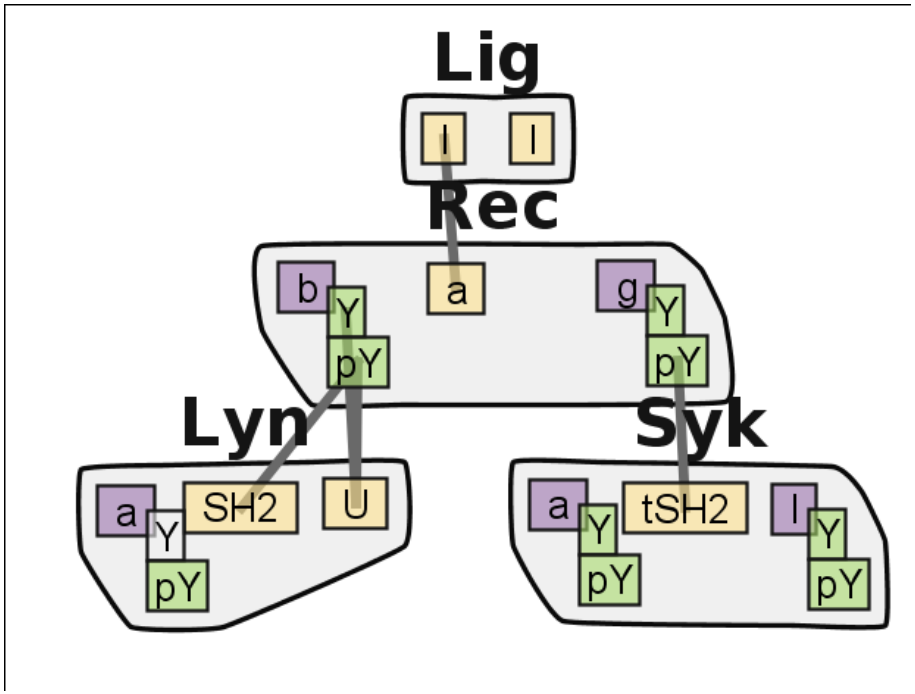
Terminology



Contact map

- Molecules: Indivisible entities that associate with other indivisible entities
- Components: Molecule's functional attributes
- Species: Unique configuration of one or more molecules

Terminology (2)



- Reaction: Transformation(s) applied to one or more species
- Rule: Compressed representation of a set of reactions. Uses patterns.

Reaction Network Model(RNM): reactions
Rule-Based Model(RBM): rules

Rule-based modeling

Pros:

- Rich syntax that allows the modeler to encode structural and contextual information
- Very scalable

Cons:

- Syntax may be overkill for smaller models or phenomenological models.

Reaction Network Modeling

$$dS/dt = \sigma R(t) - acS(t)I(t) - \mu S(t)$$

$$dI/dt = acS(t)I(t) - \rho I(t) - \delta I(t)$$

$$dR/dt = \rho I(t) - \sigma R(t) - \mu R(t)$$

$$dD/dt = \mu S(t) + \delta I(t) + \mu R(t)$$

RNM representation

SBML

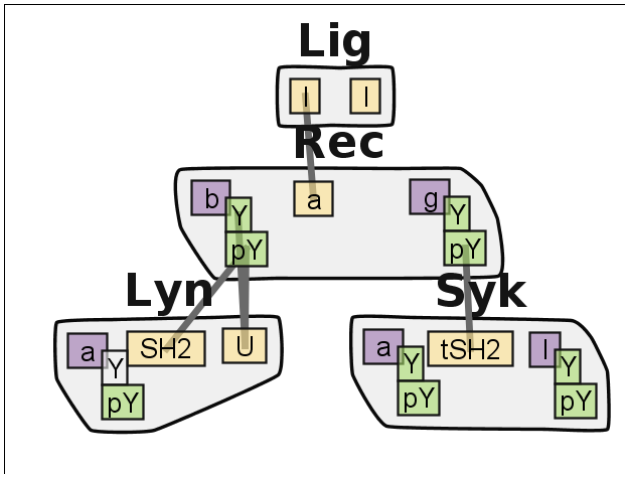
```
A + B -> A_B
```

BNGL

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

RNM is adequate for small models.

Network expansion



19 rules
4 molecules

SBML File

354 species
3680 rxns

This phenomenon is known as combinatorial complexity

Reaction Network Modeling (2)

Pros:

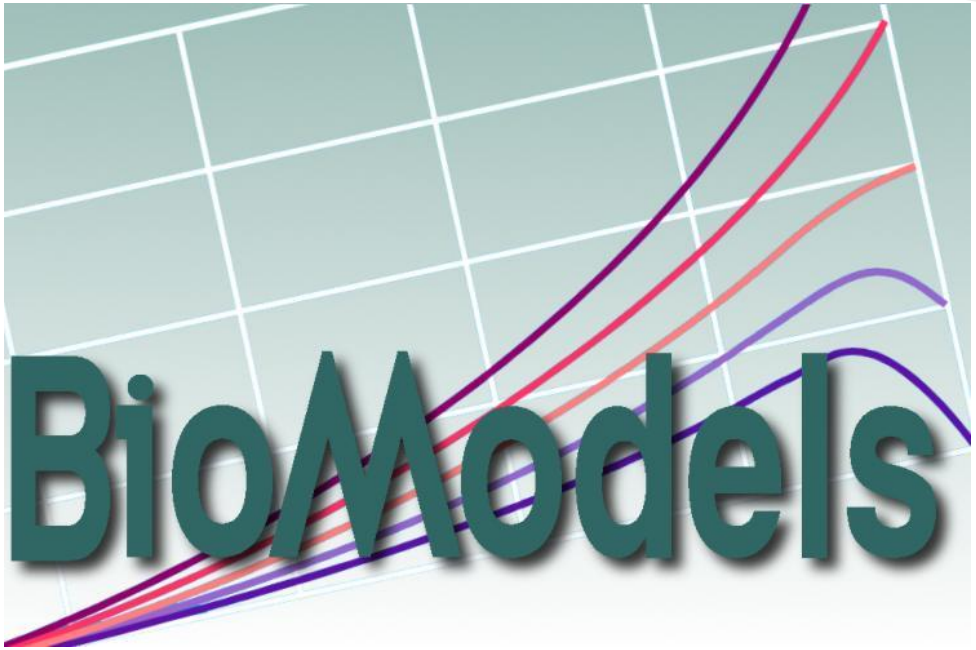
- Well understood theory of differential equations based chemical kinetics
- Simple representation suitable for smaller models
- Suitable for smaller models

Cons:

- Biological structural and contextual information is lost
- Non scalable

BioModels

There exist a large body of knowledge already encoded in SBML.



Our goals are

We could deeply understand and build upon years of RNM modeling knowledge in a semi-automated way.

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

Make the benefits of RBM accessible to a larger public.

Presenting…

The

Atomizer

RBM

RNM

How does it work?

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

Atomizing in a nutshell
Identify what every reaction does

- Reaction stoichiometry information

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

- Naming conventions

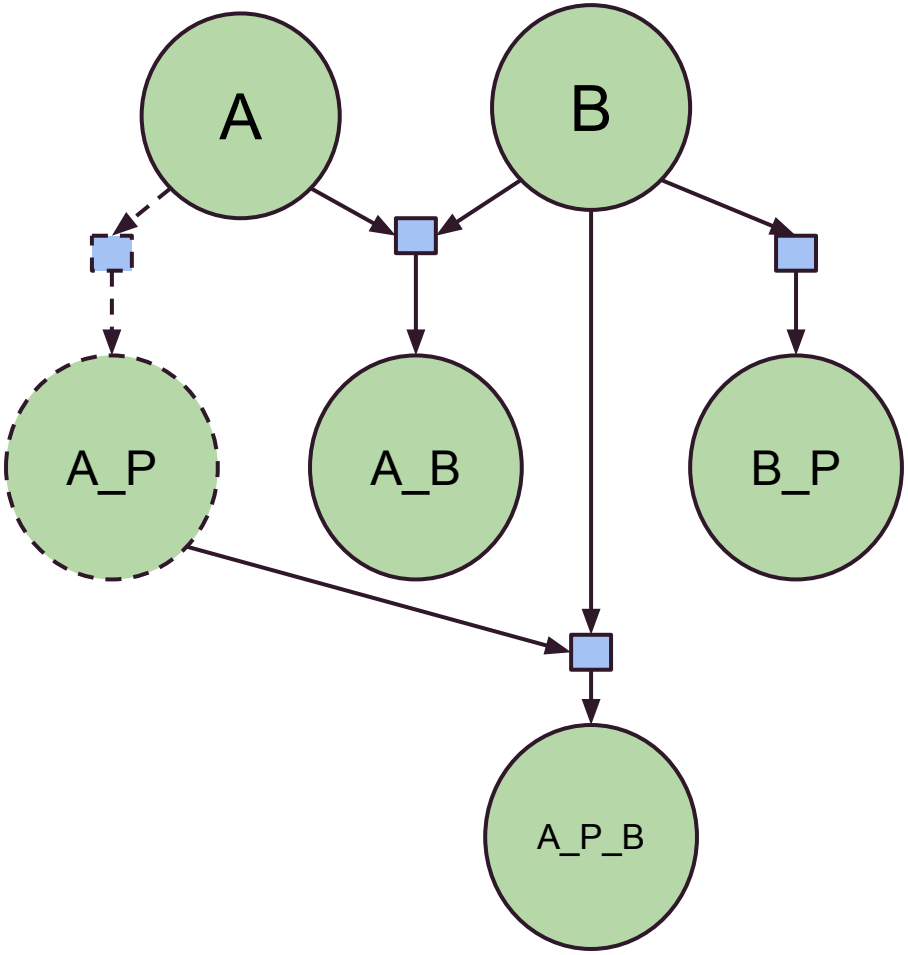
`B -> B_P` *Phosphorylation*

- Lexical analysis

`A + B_P -> A_P_B`

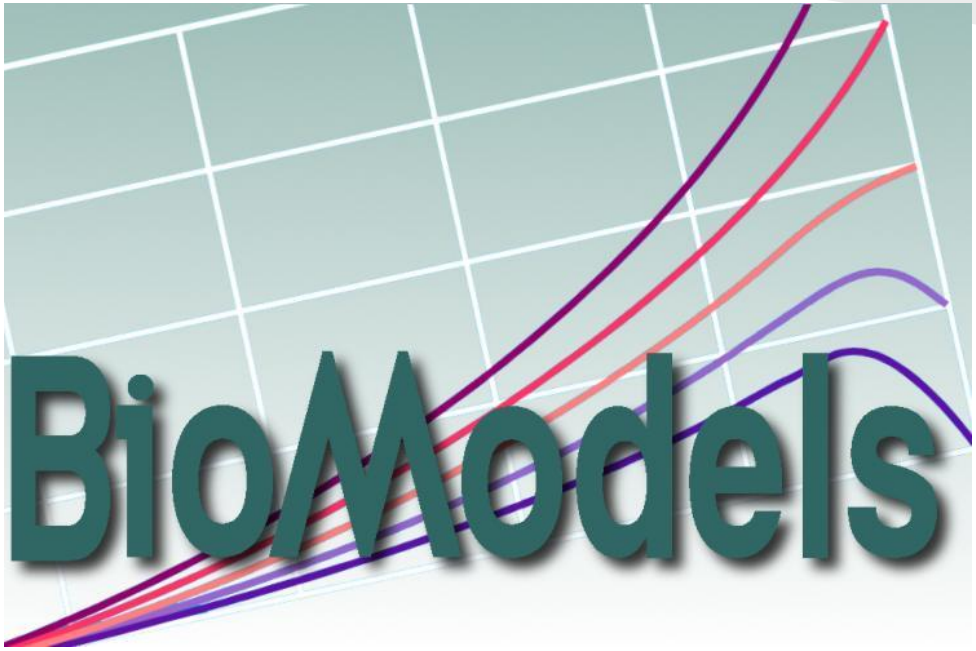
Complexation and phspaiyuiyiuyiu

Composition Graph

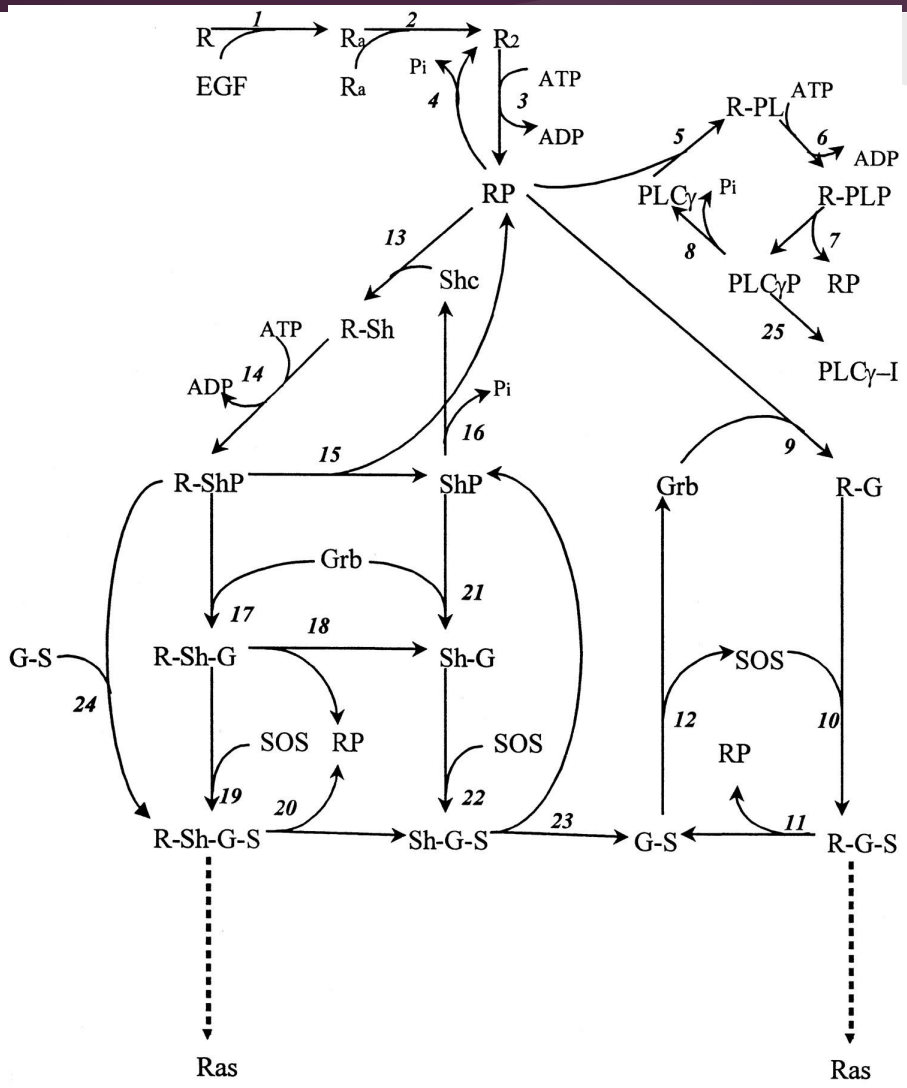



```
A(b,phospho~U~P)
B(a,phospho~U~P)
```

Atomize BioModels

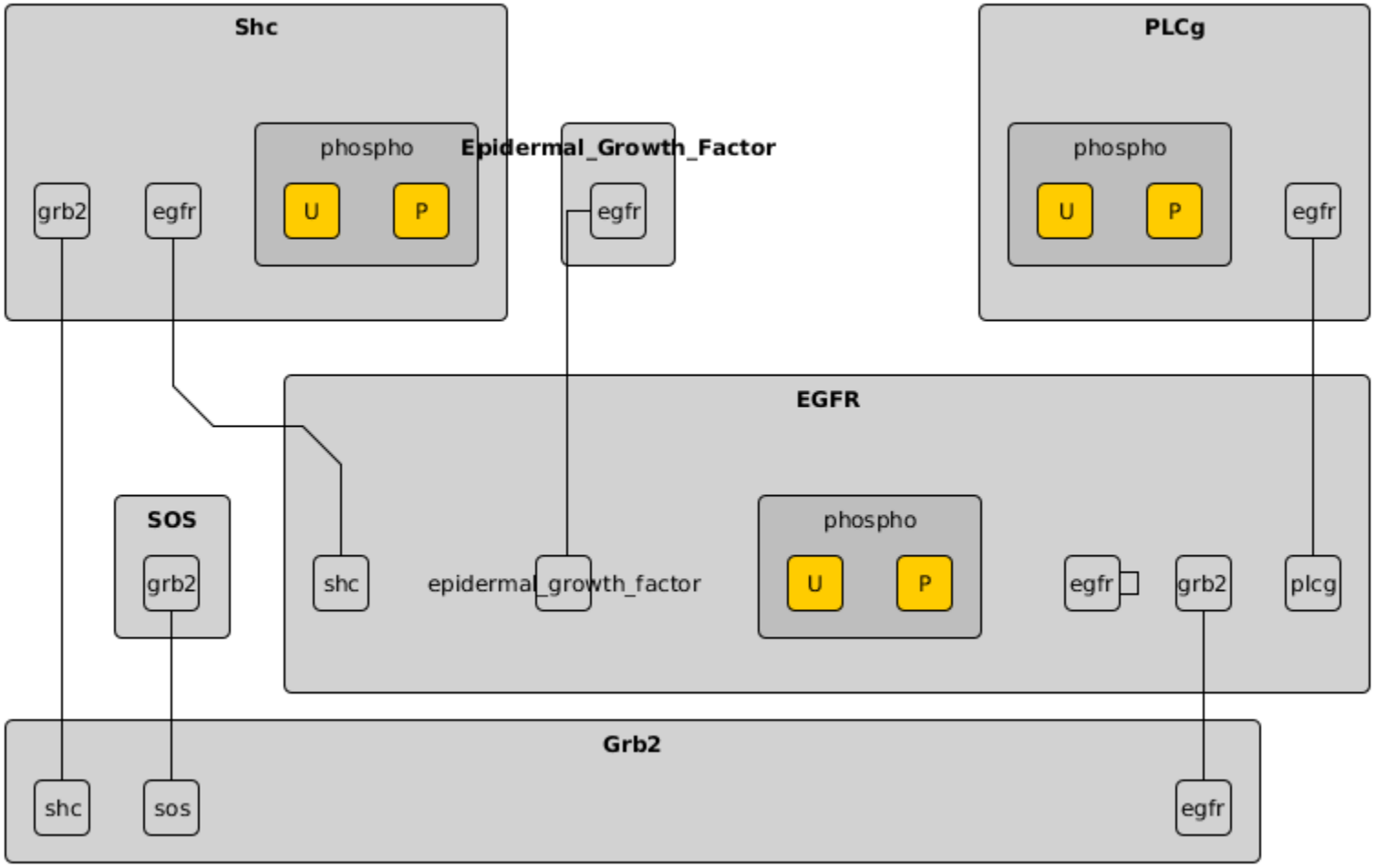


Epidermal Growth Factor

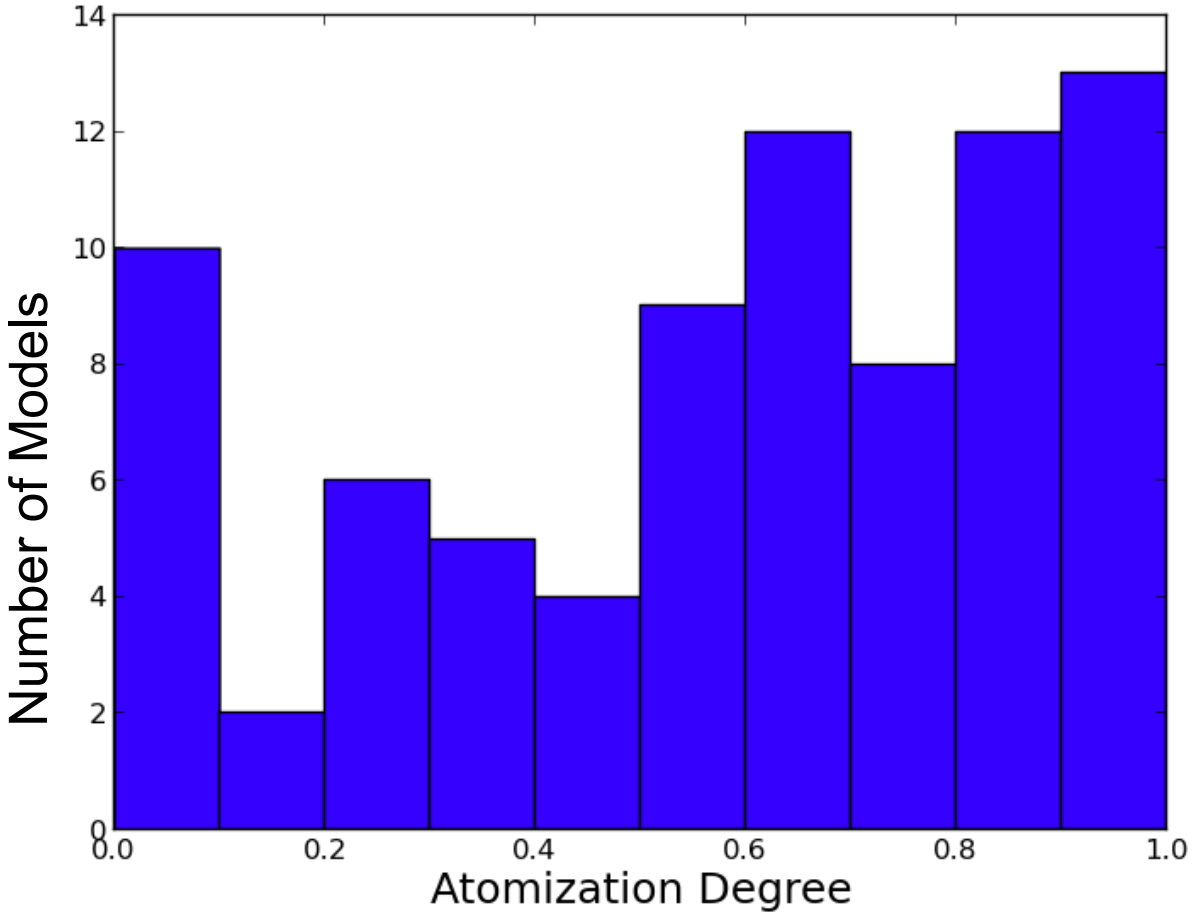


Kholodenko 1999

BioModels 48 atomized



Atomization degree = structured / total molecules



Models with >20 reactions

Model fusion

Model aggregation: a building-block approach to creating large macromolecular regulatory networks

Ranjit Randhawa[1], Clifford A. Shaffer[1,*] and John J. Tyson[2]

[1]Department of Computer Science and [2]Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA

Received on April 29, 2009; revised on August 25, 2009; accepted on September 28, 2009

Advance Access publication October 29, 2009

Associate Editor: Trey Ideker

"Process that combines two or more submodels into a single unified model that contains the combined information (without redundancies) across the original collection."

How do I map species?

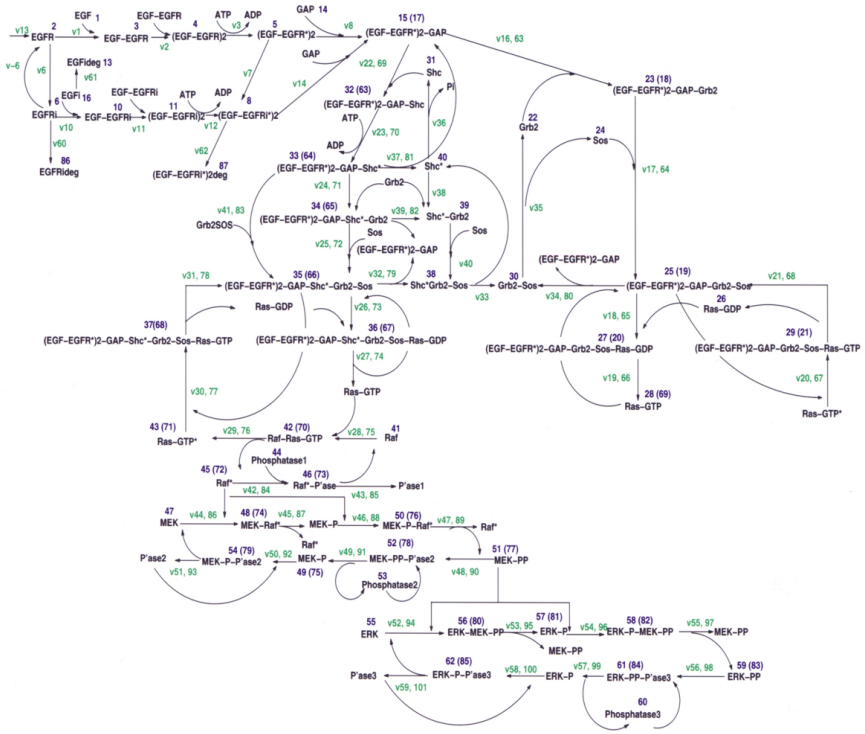


Image taken from Schoeberl et. al.

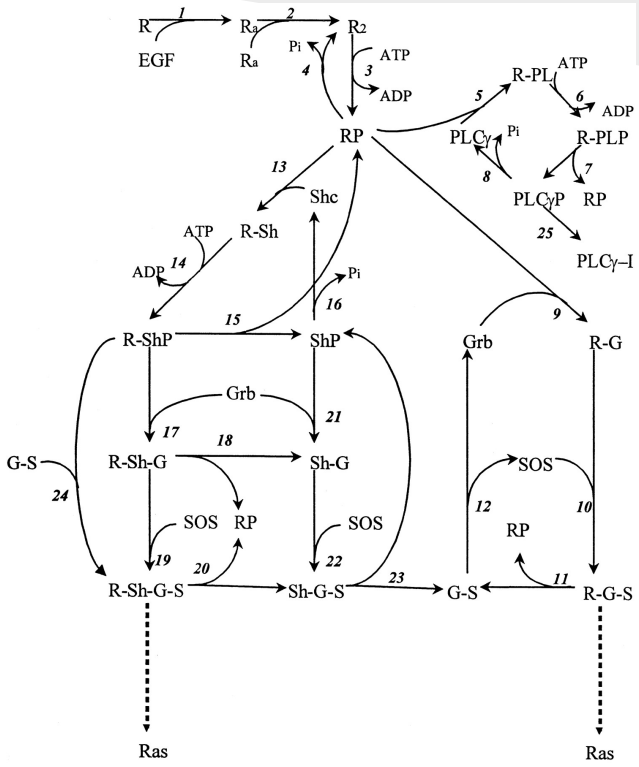
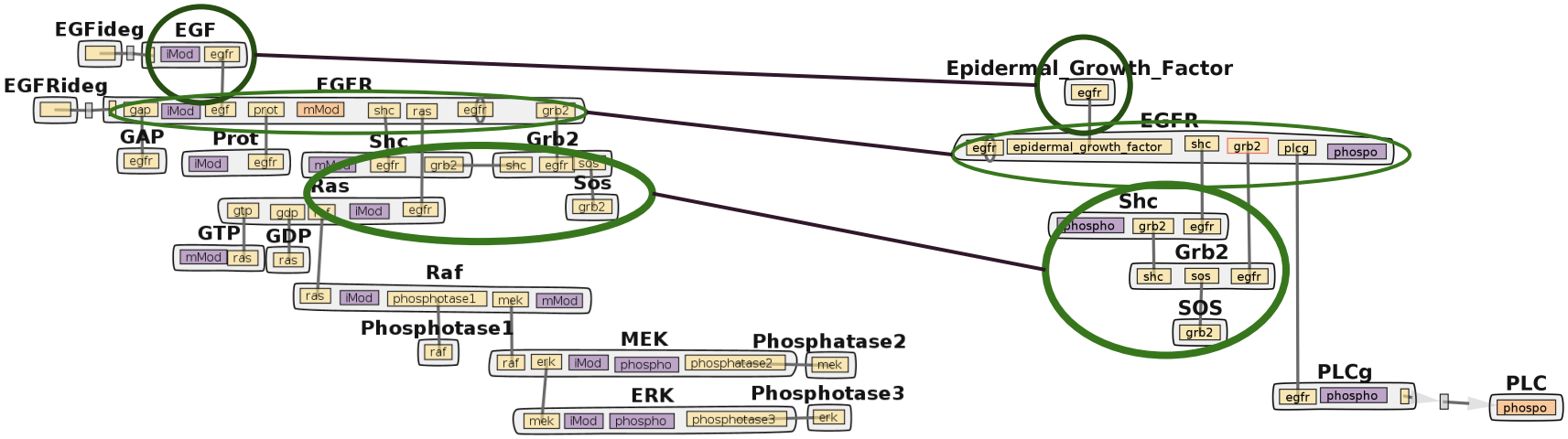


Image taken from Kholodenko et. al.

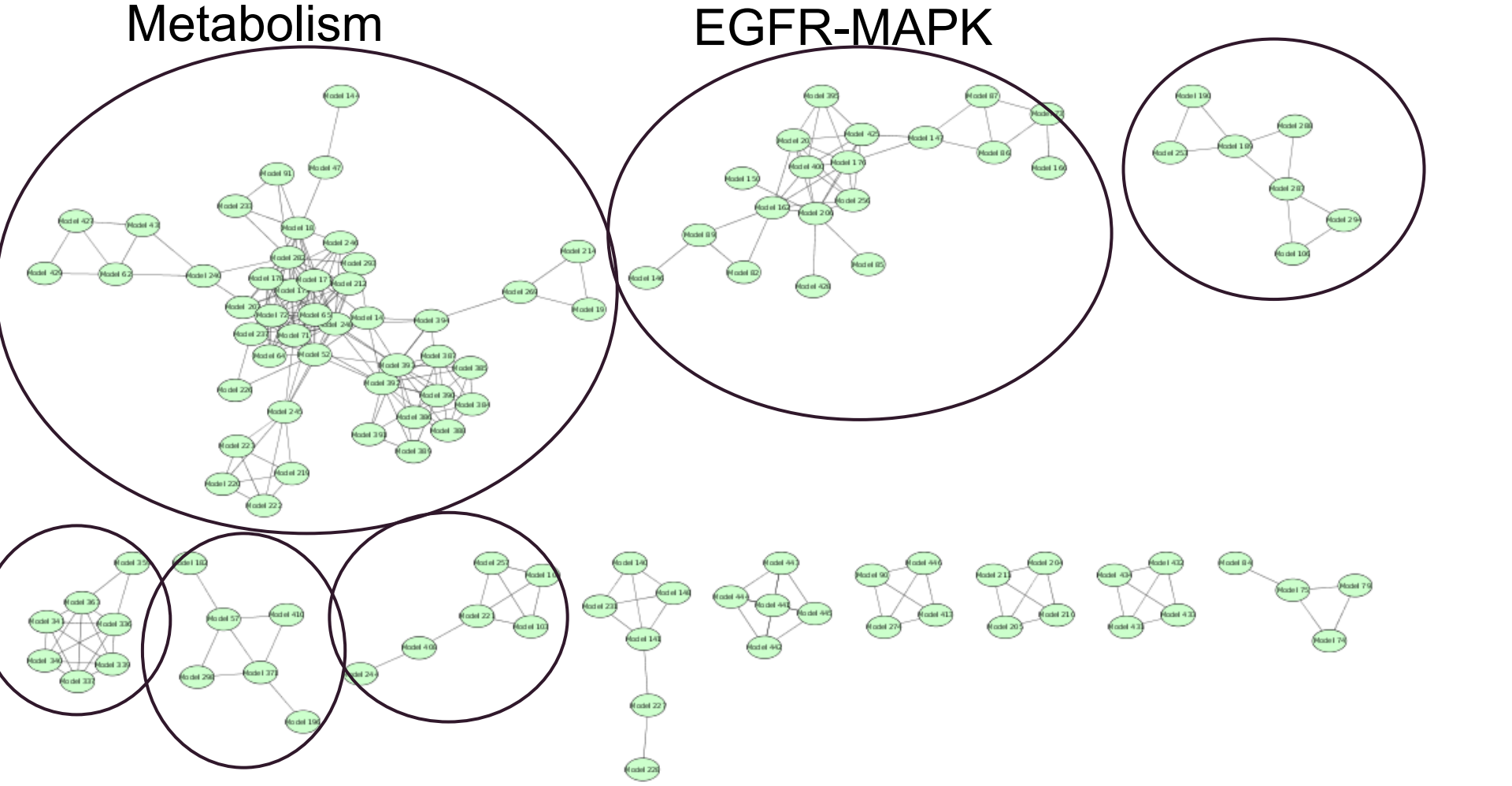
RBM Model fusion (2)



Schoeberl et. al.

Annotation information is used to reliably map species with different names across models

Model interaction(each node is a BioModels model)



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

Model verification: BioModels 109

Model annotation tells us that…

Cdk1Y10:

- Cyclin-dependent kinase

Cdk1Y11:

- Cyclin-dependent kinase
- Cyclin A

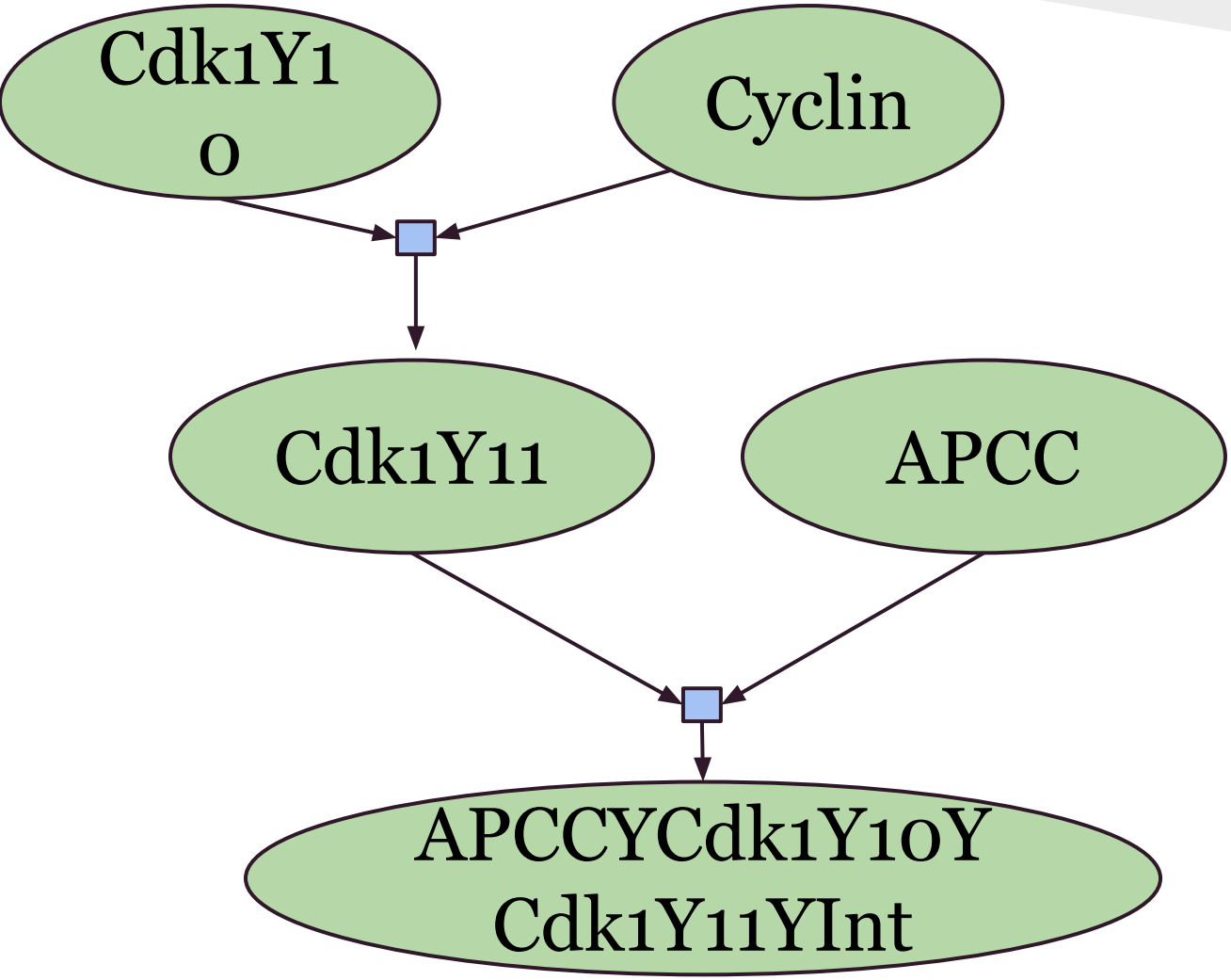
APCC:

- ubiquitin ligase

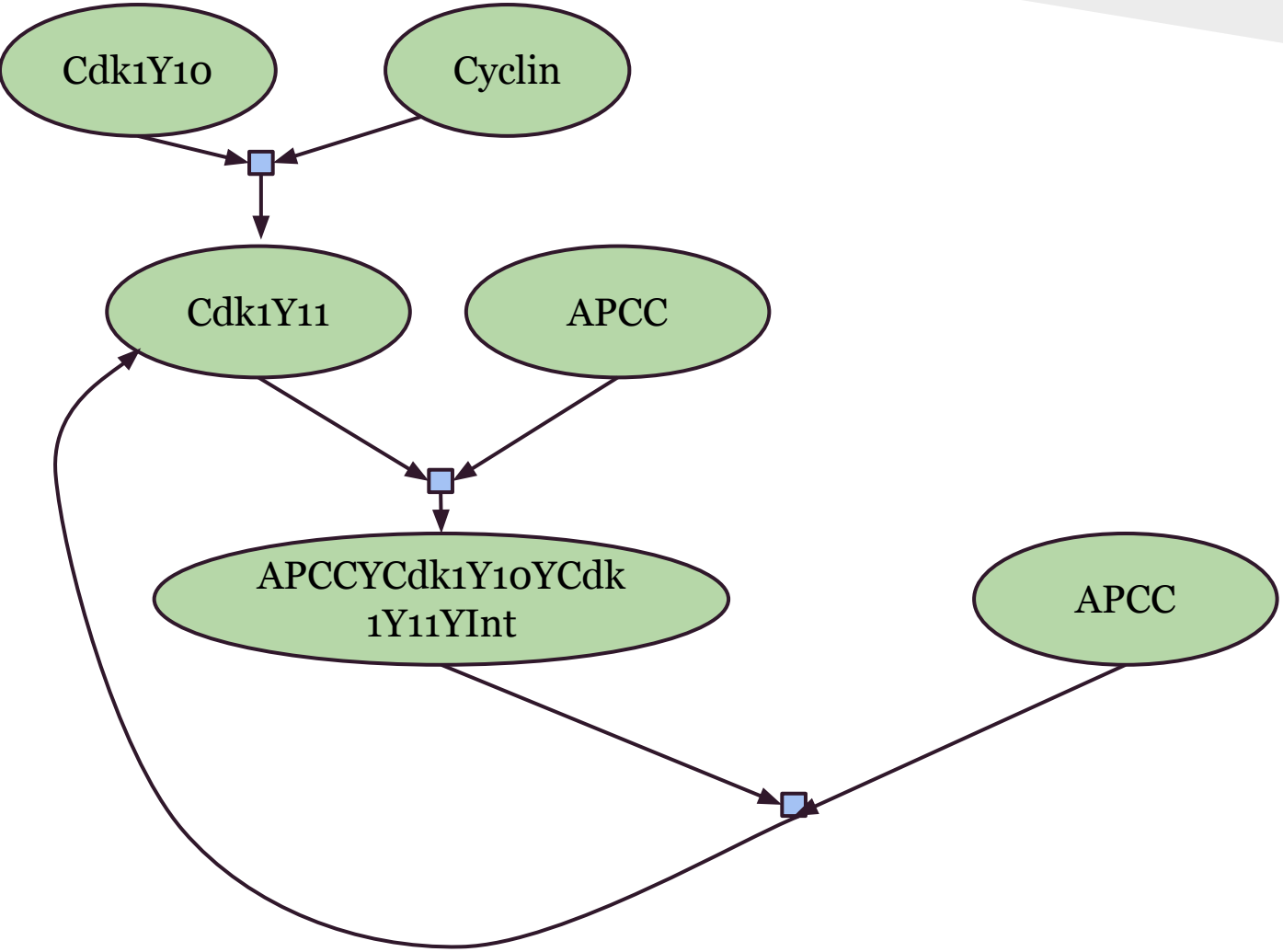
APCCYCdk1Y10YCdk1Y11YInt:

- Cyclin-dependent kinase
- ubiquitin ligase
- Cyclin A

Model composition (BM 109)

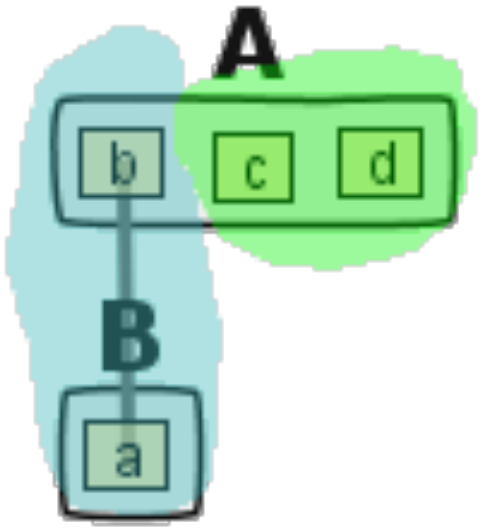


And then… ??


Cdk1Y10
Cyclin
Cdk1Y11
APCC
APCCYCdk1Y10YCdk
1Y11YInt
APCC

Reaction center and context

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



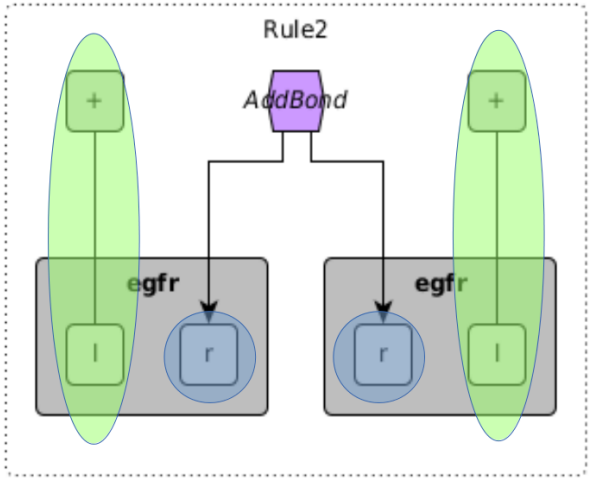
SBML Molecules contain minimal context information

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

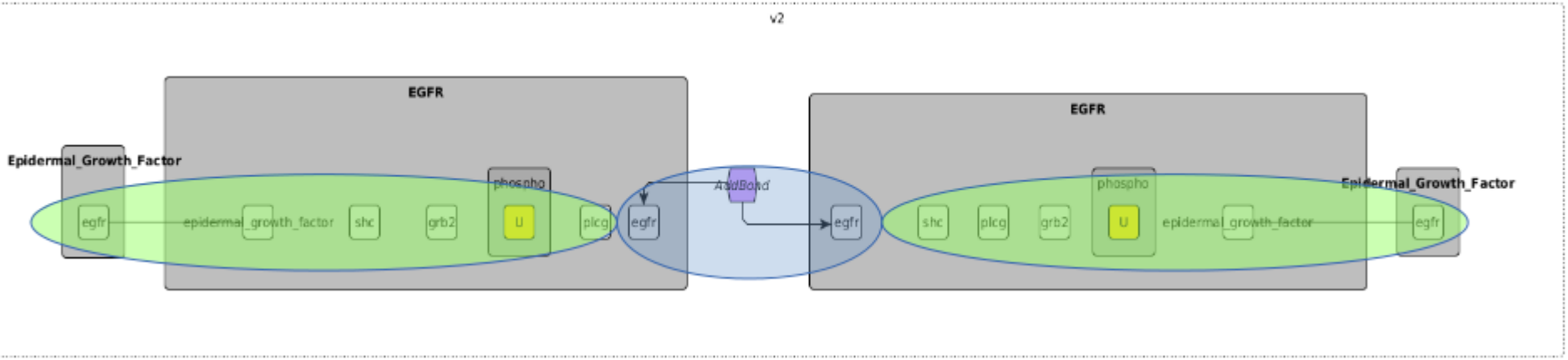
What is A? (Single molecule, umbrella name for a series of compound series of A molecules, etc).

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

Kholodenko's reaction context

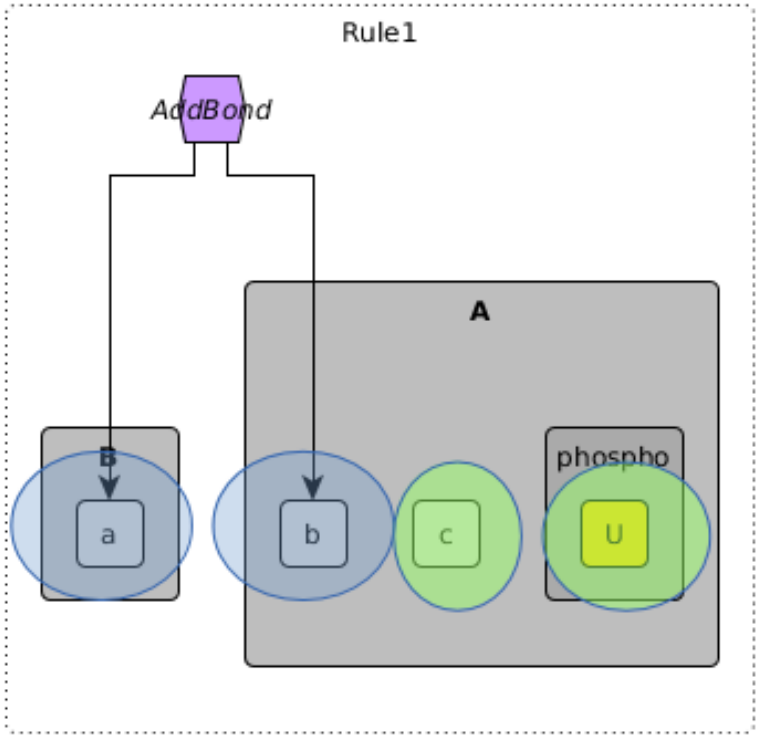


Manually constructed RBM version of Kholenko's model.

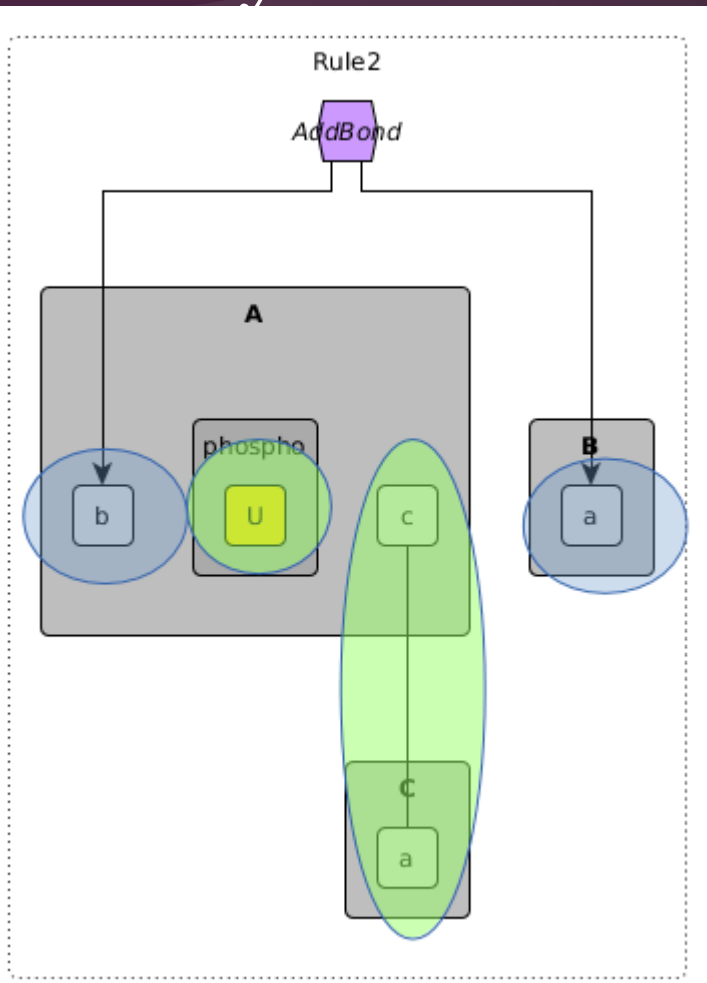


Automatically translated version of Kholodenko's model

Redundancy is the redundant way to redundantly go redundantly

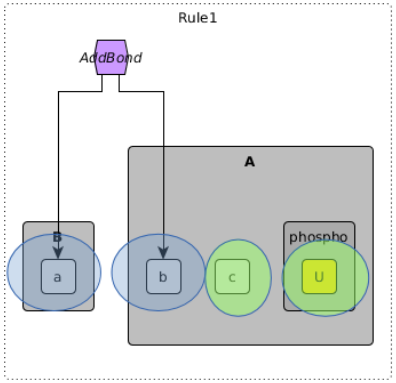


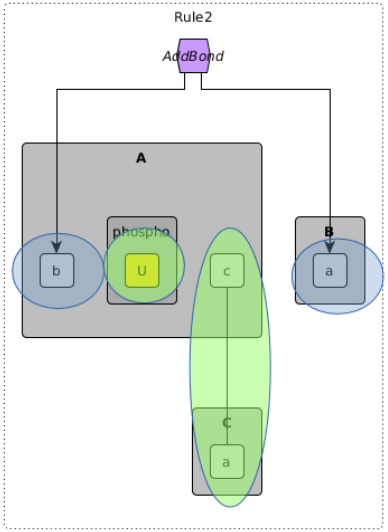
For ‘A’ and ‘B’ to bind all other states must be unoccupied.

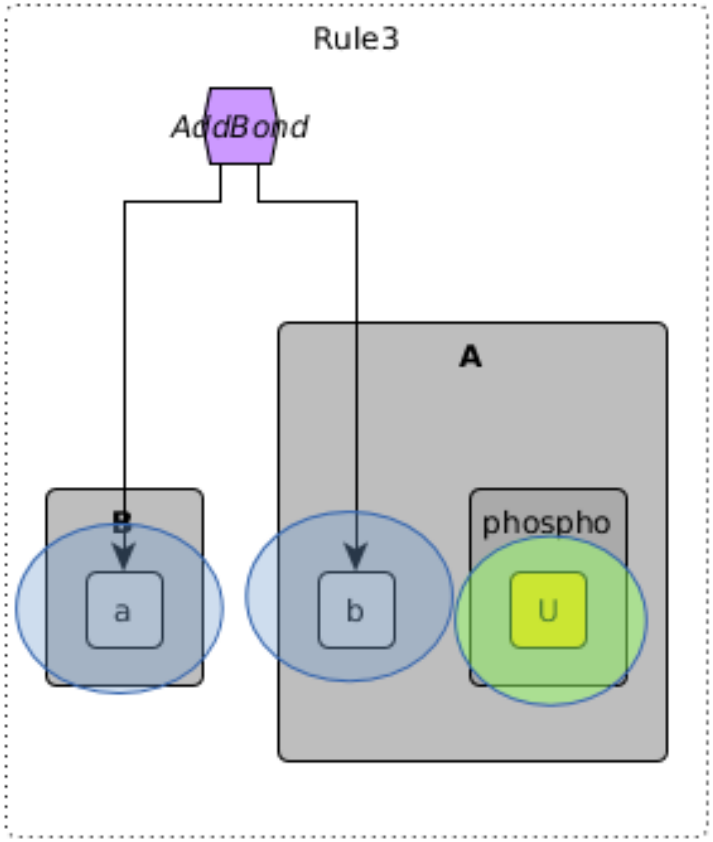


For ‘A’ and ‘B’ to bind ‘A’ must be bound to ‘C’ already

Context factorization

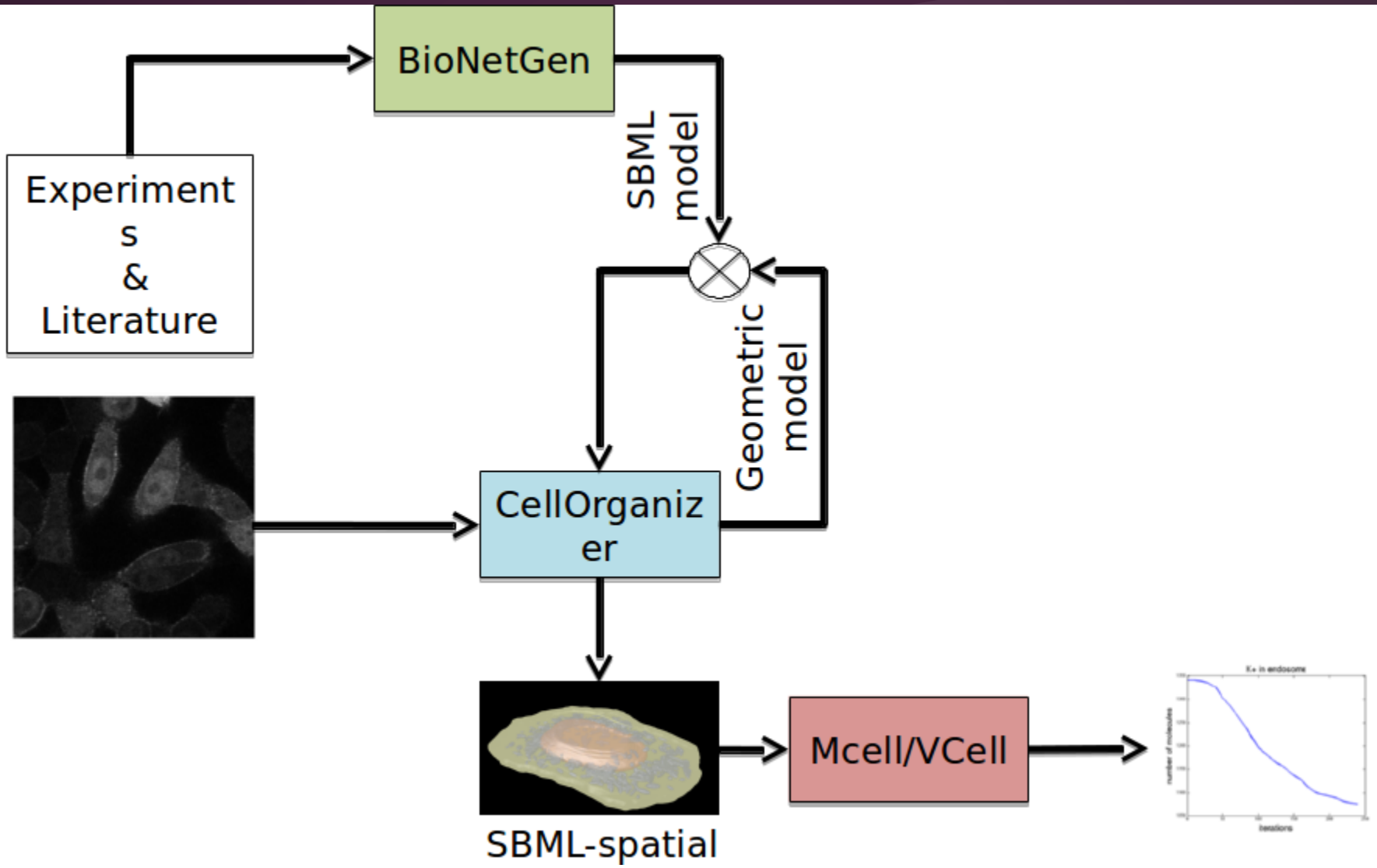






For A and B to bind it does not matter if C is part of the complex or not

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

Devin Sullivan
Aaron Wise

Q&A