Using Atomizer to analyze and compare Reaction Network models

Jose Juan Tapia Valenzuela
University of Pittsburgh

What’s in a model?

What’s in a model means...

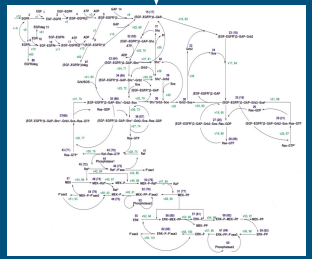
- I want to **understand** a model in the literature
- I want to **compare** a model against others in the literature
- I want to **reuse** models in the literature

What's in a model?

SBML

Annotation information

Paper



The challenges of model understanding are...

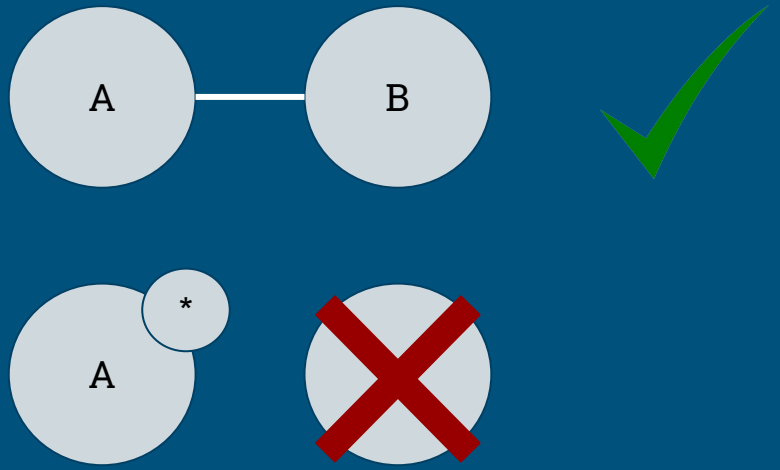
- How are elements inside a model related to each other?
- How do elements in a model compare to elements in other models?
- How do elements in a model compare to real-world objects?

The challenges of model understanding are...

- **How are elements inside a model related to each other?**
- How do elements in a model compare to elements in other models?
- How do elements in a model compare to real-world objects?

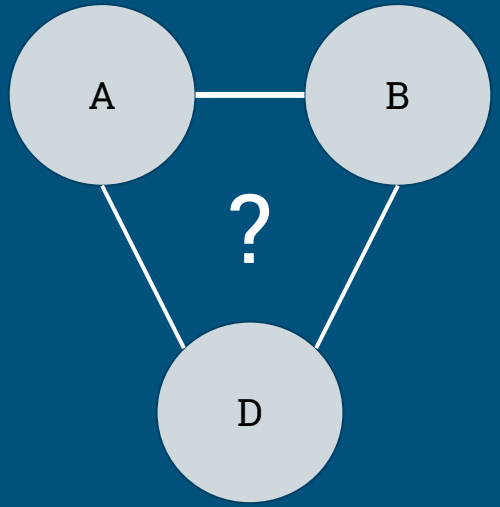
A pure RNM representation makes this a non-trivial problem...

A + B -> ~~C~~ A_B



It gets more interesting….

A_B + D -> A_B_D



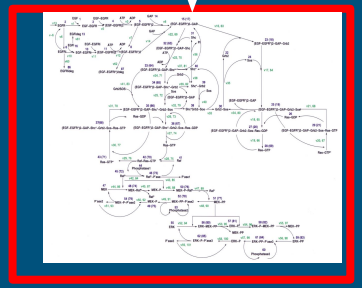
Some of these questions can be answered through model visualization

What's in a model?

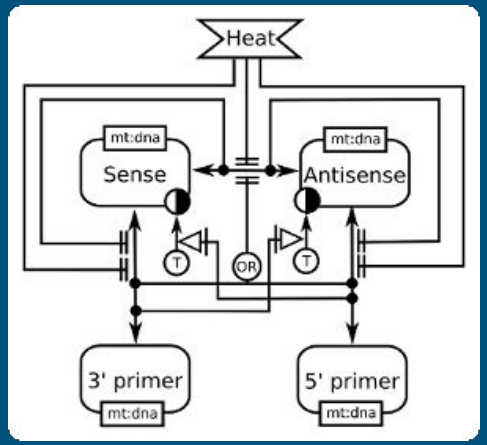
SBML

Annotation information

Paper

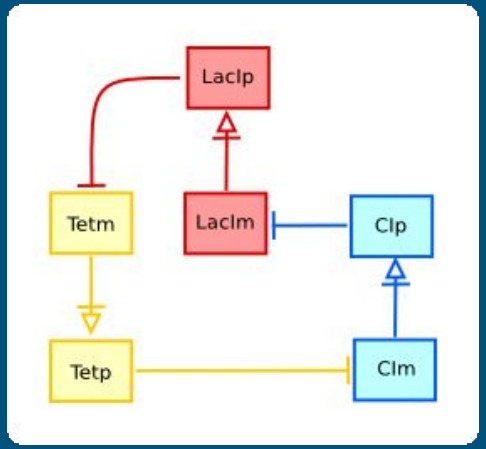


Several visualization methodologies (SBGN)



Entity-relationship diagrams

Process diagrams



Activity flow diagrams

So we ask...

Does this scale for large models?

What happens if I want to understand and compare a large number of models?

Toward “model informatics”

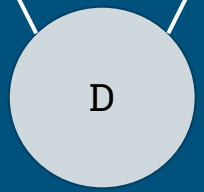
There’s a limit to what we can understand without computer assistance.

A model description should enable (semi) automated analysis of a single model and comparison with other models.

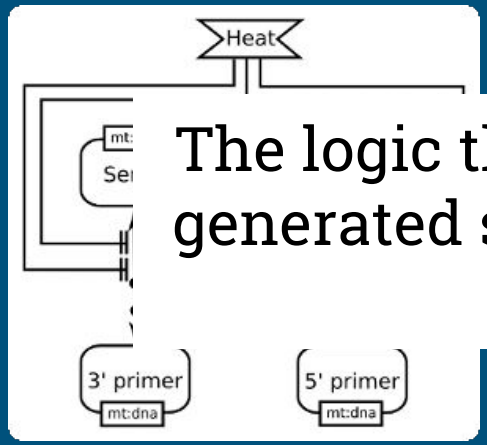
It gets more interesting….

A + B + D -> A + B + D

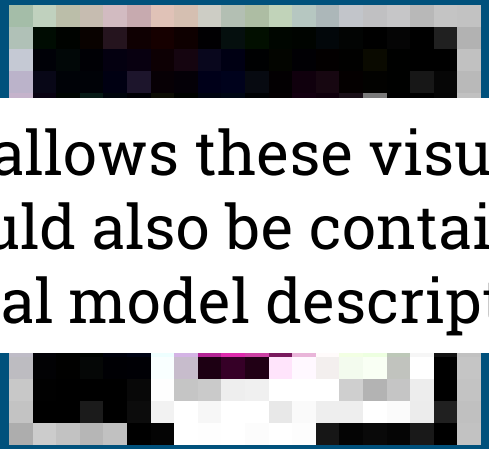
These questions should be answerable from the model description alone



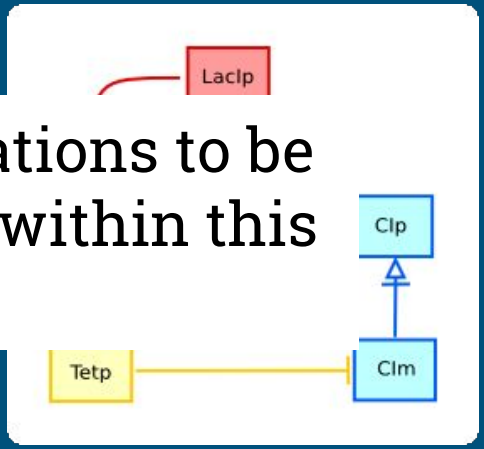
Several visualization methodologies (SBGN)



Entity-relationship diagrams



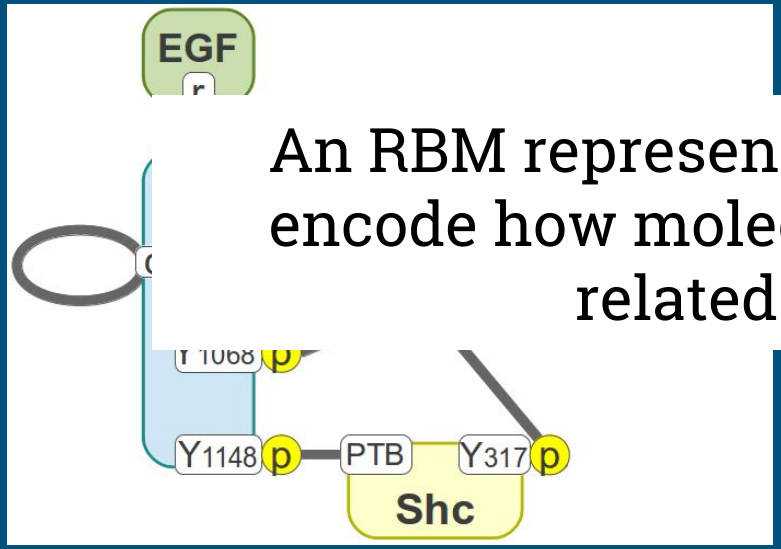
Process diagrams



Activity flow diagrams

The logic that allows these visualizations to be generated should also be contained within this ideal model description

Enter Rule-based modeling (SBML multi)



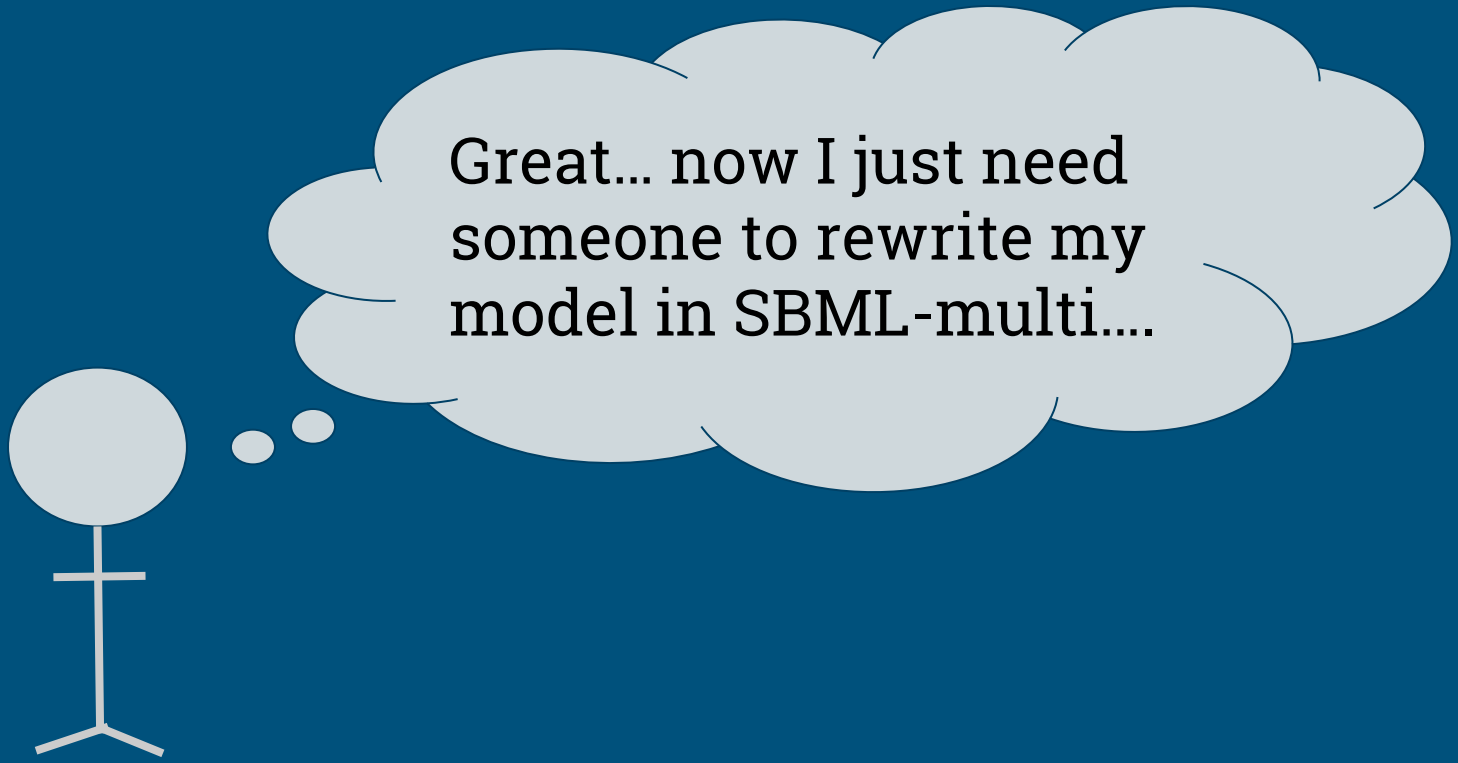
An RBM representation allows the user to encode how molecules inside a model are related to each other

- Graph based representation

ins

interactions

- **Graph-based syntax allows a modeler to directly encode structural information**


Great… now I just need someone to rewrite my model in SBML-multi….

Presenting...

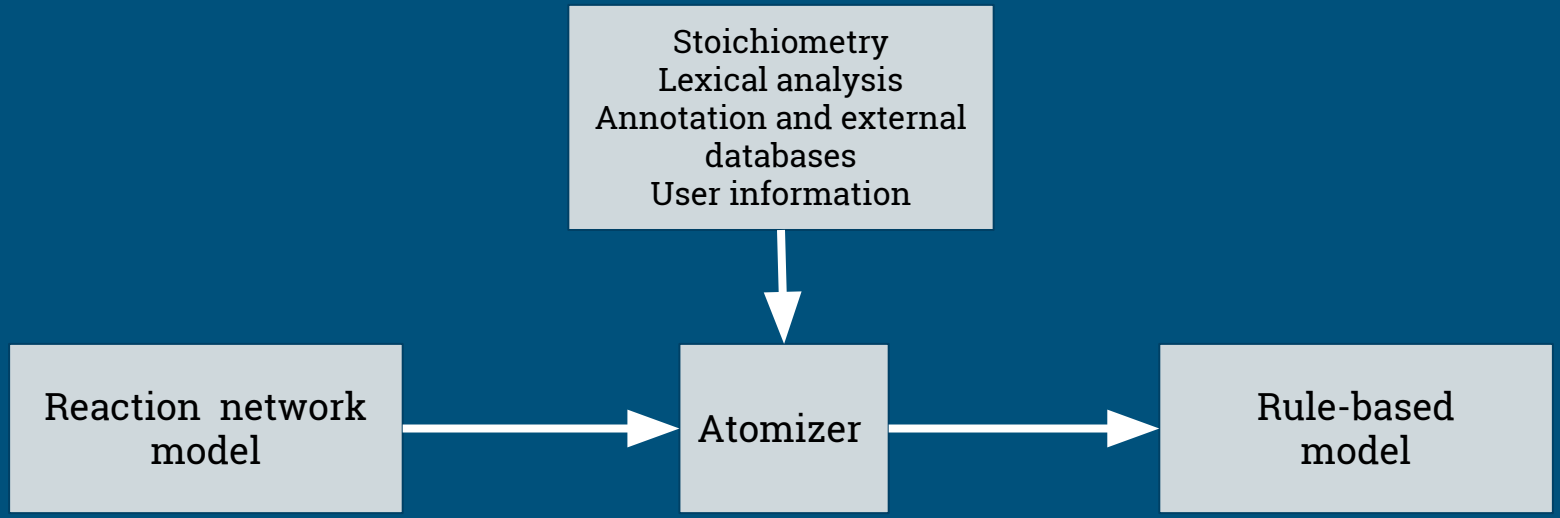
The

Atomizer

RBM

RNM

Now without the cats



How does Atomization work?

Stoichiometry

```
EGF + EGFR <-> EGF-EGFR
```

How does Atomization work?

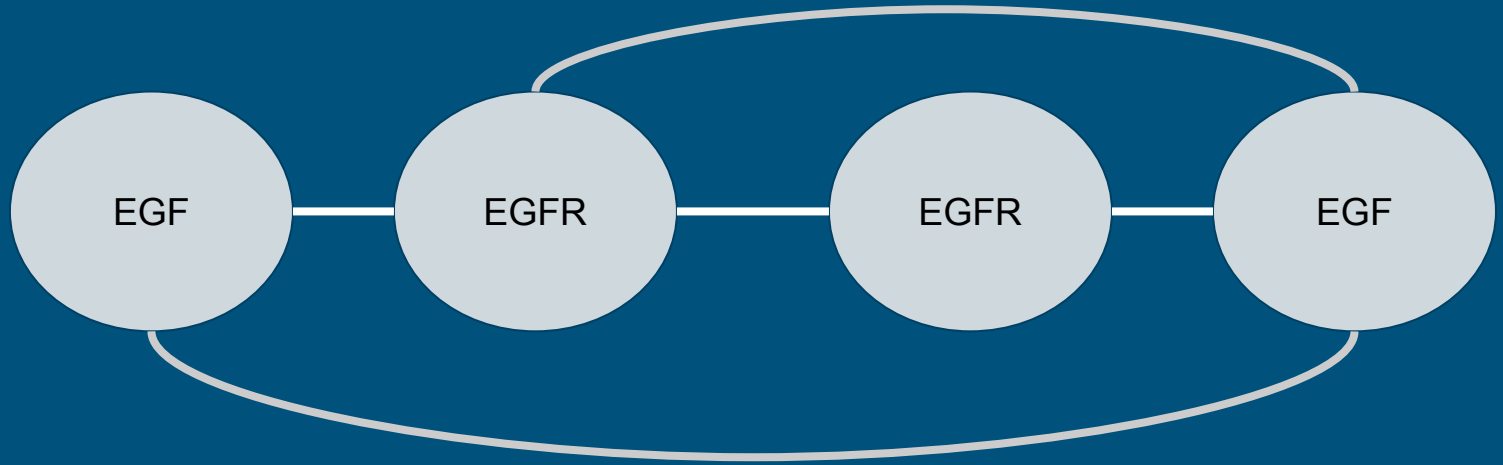
Lexical analysis

```
EGFR <-> EGFRi
```

How does Atomization work?

Protein interaction databases

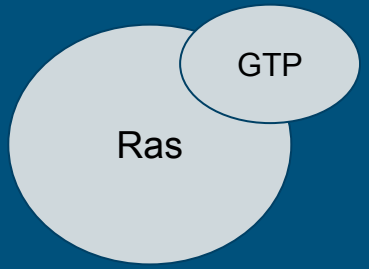
```
EGF-EGFR + EGF-EGFR <-> EGF-EGFR^2
```



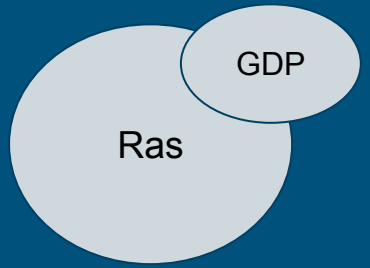
How does Atomization work?

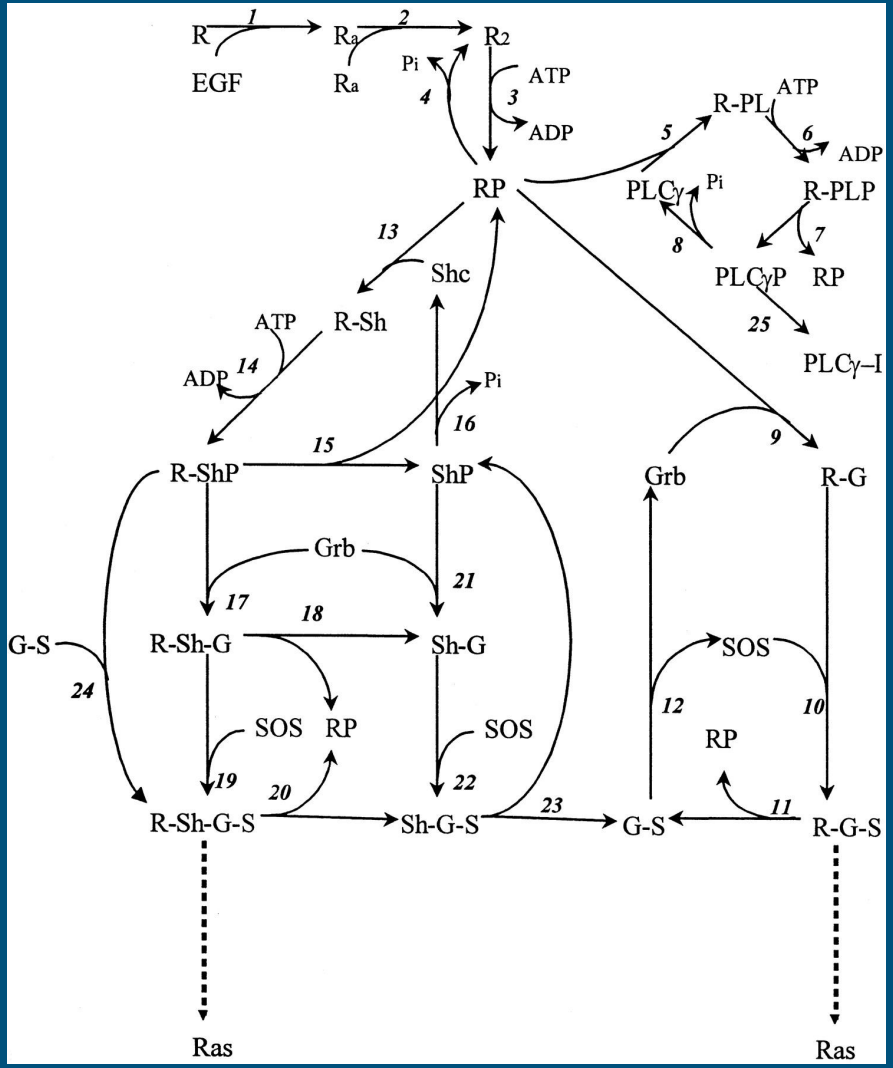
Annotation + User information

Ras-GTP



Ras-GDP

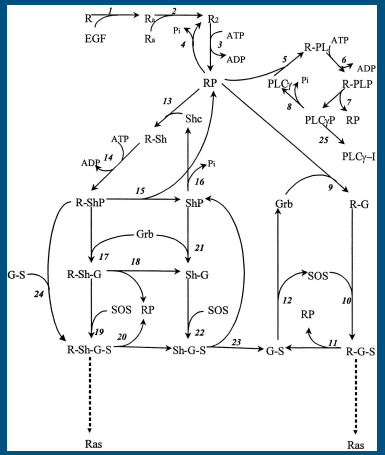


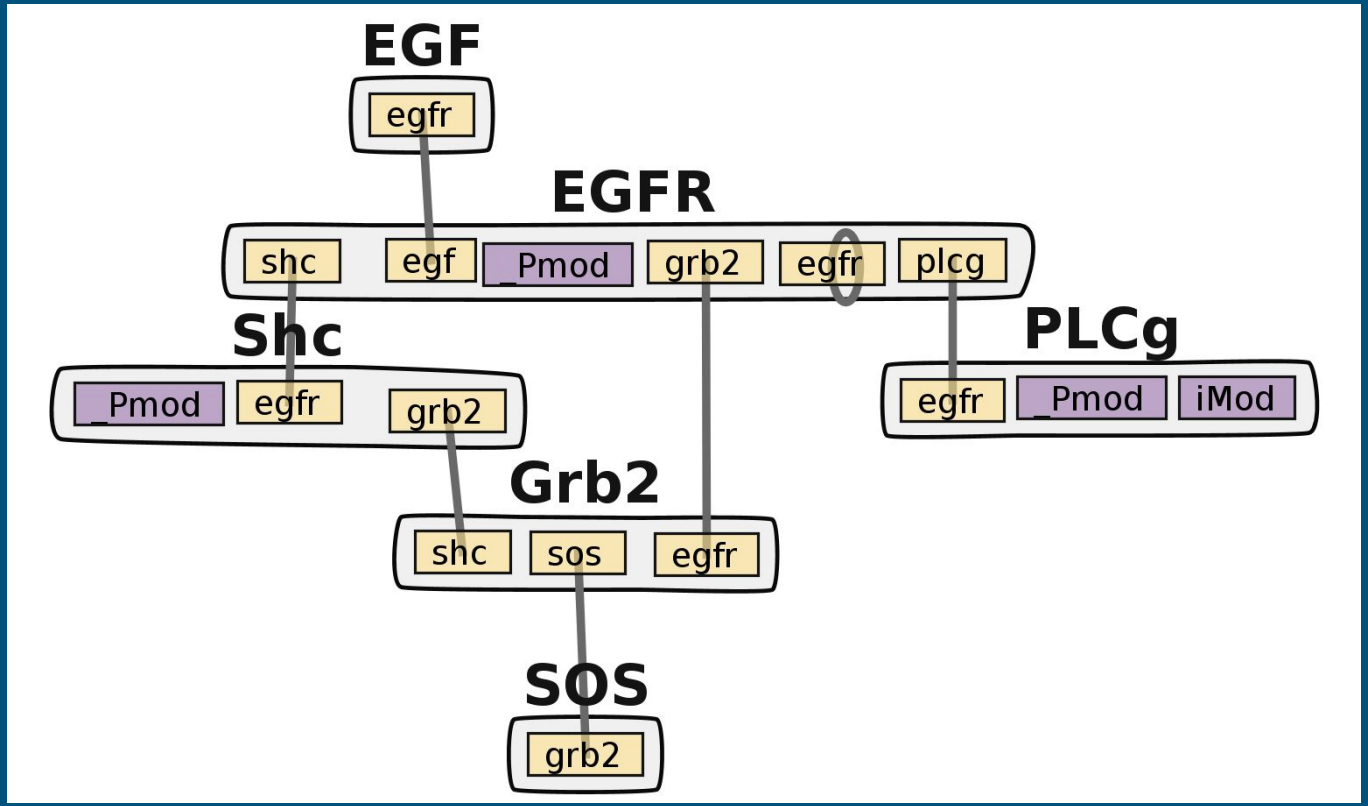


Motivational example

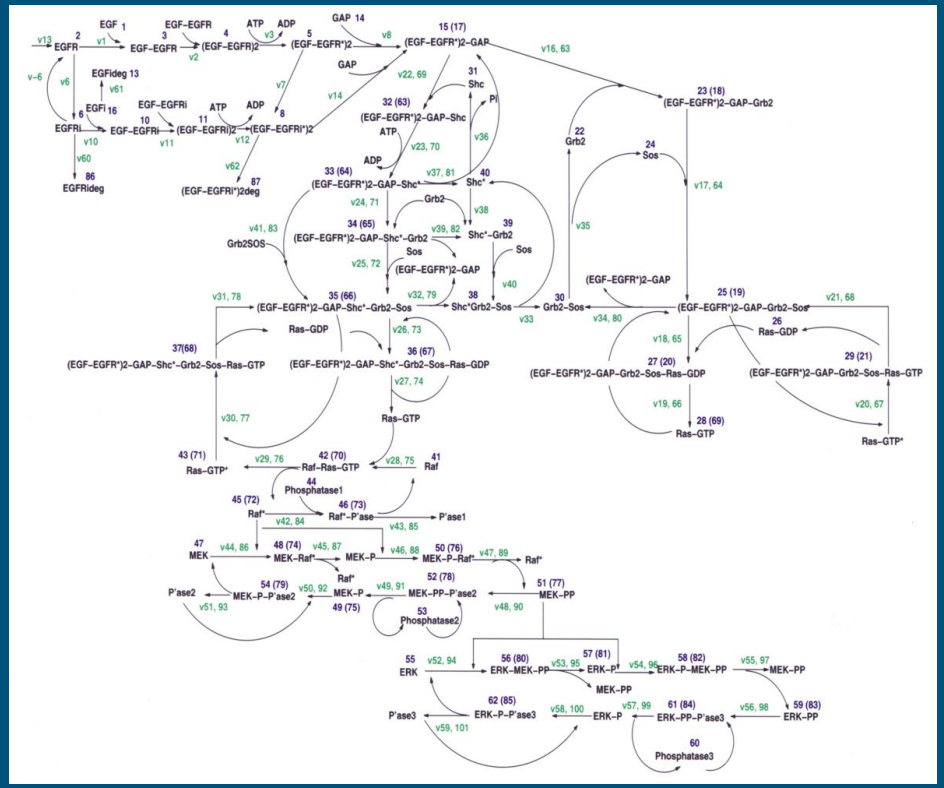
BIOMD48
(Kholodenko B. 1999)

Magic!


EGF
ATP
ADP
Pi
RP
Shc
R-Sh
R-ShP
ShP
Grb
R-G
R-Sh-G
Sh-G
SOS
R-Sh-G-S
Sh-G-S
G-S
R-G-S
R-PLP
PLCγP
PLCγ–I
Ras

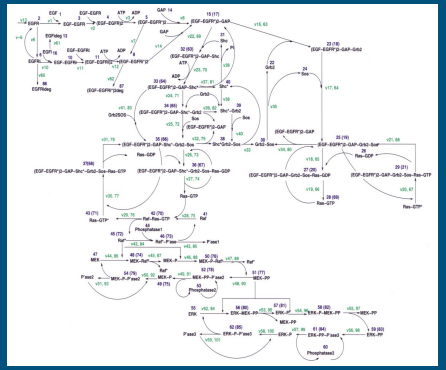

EGF
egfr
EGFR
shc
egf
_Pmod
grb2
egfr
plcg
Shc
_Pmod
egfr
grb2
PLCg
egfr
_Pmod
iMod
Grb2
shc
sos
egfr
SOS
grb2

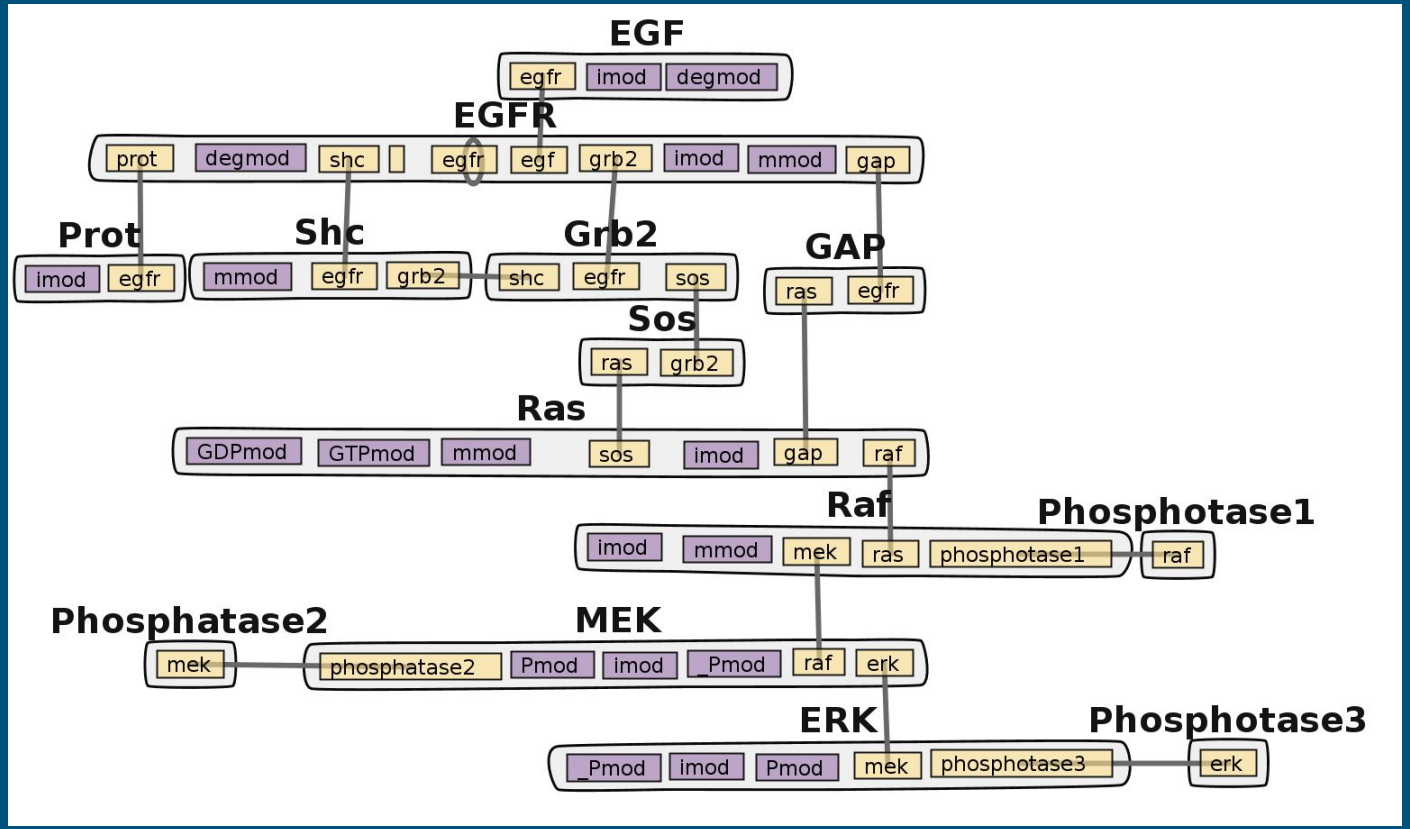
Motivational example (2)



Schoeberl et al.

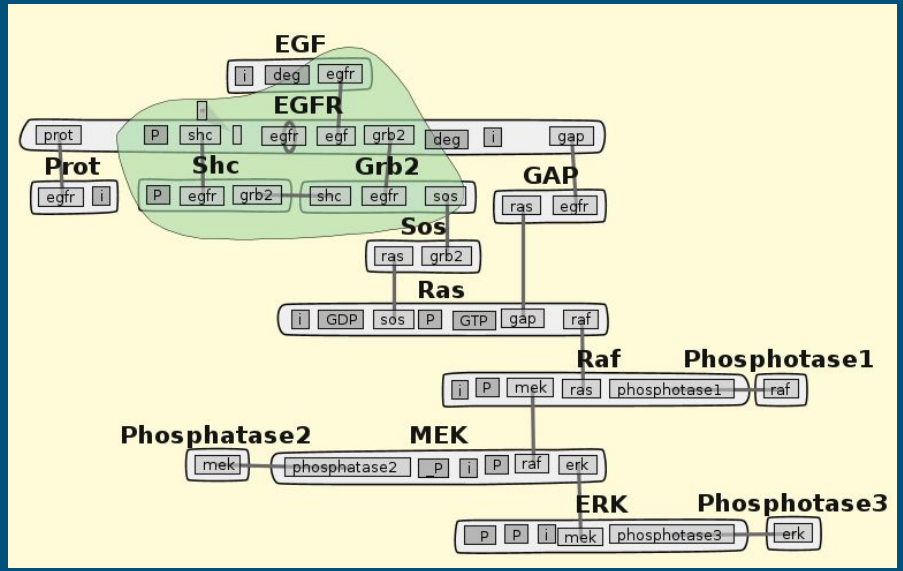
BIOMD19 atomized



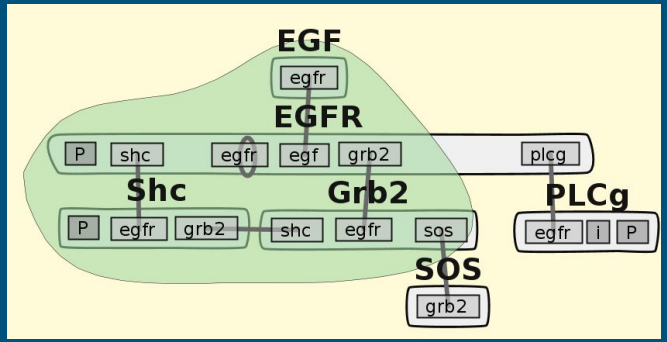


Comparison

BIOMD 19



BIOMD 48



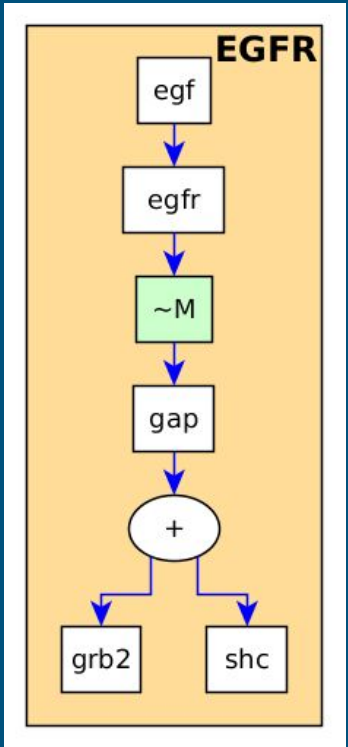
What we learned so far

- Atomization resolves sites of protein-protein interaction and modification
- Rule-based modeling based visualization techniques provide a summary view of the structural information we recovered.
- However structural information alone is not enough to describe how the processes in a system occur.

Can I extract model process information from a Rule-based model?

The State Transition Diagram (STD)

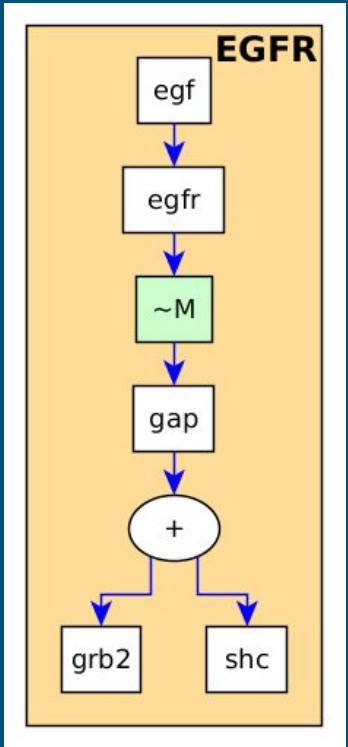
From BMD19



This graph was built by programmatically analyzing the hierarchy of bond formation and state modification inside a model

Nodes represent different states inside a molecule. Edges represent an activation/deactivation sequence.

How to read this diagram?



- EGF binding enables EGFR dimerization
- which enables modification (phosphorylation)
- which enables GAP binding
- which enables either Grb2 or Shc binding to EGFR (but not both)

So once we atomize a model...

- We can visualize it using contact maps to understand what is in it (structural information)

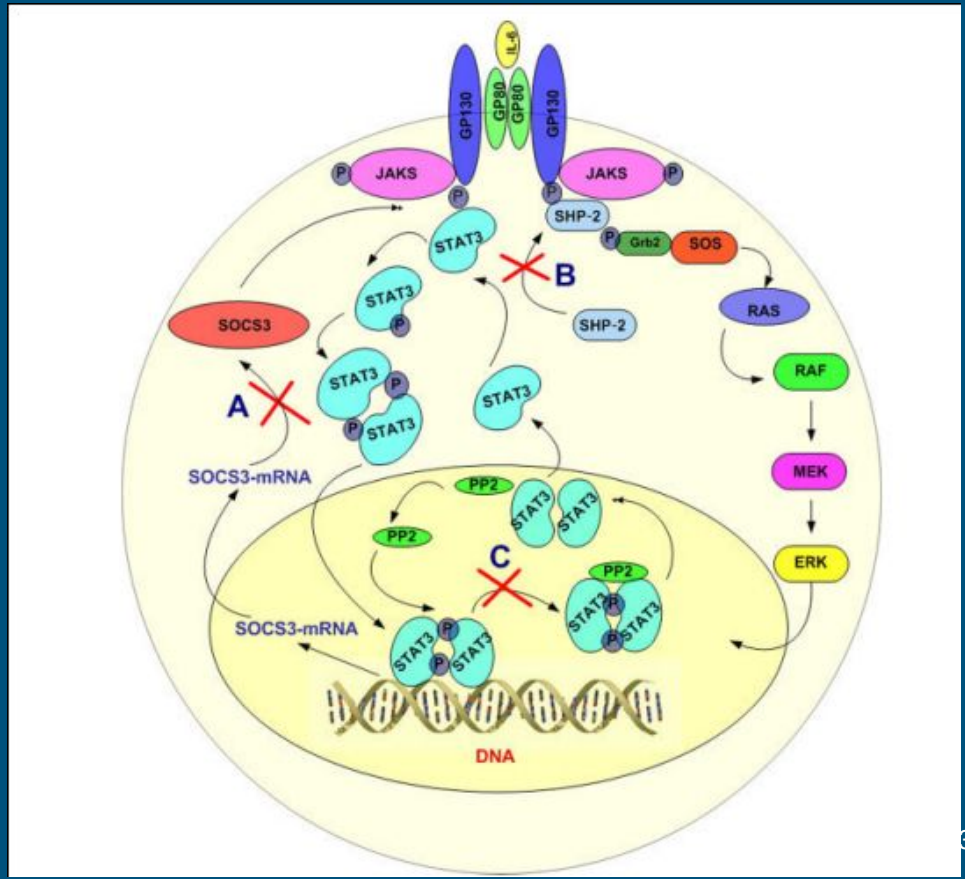
- We can visualize it using STD's to understand how things happen inside it (context and process information).

Comparing model information

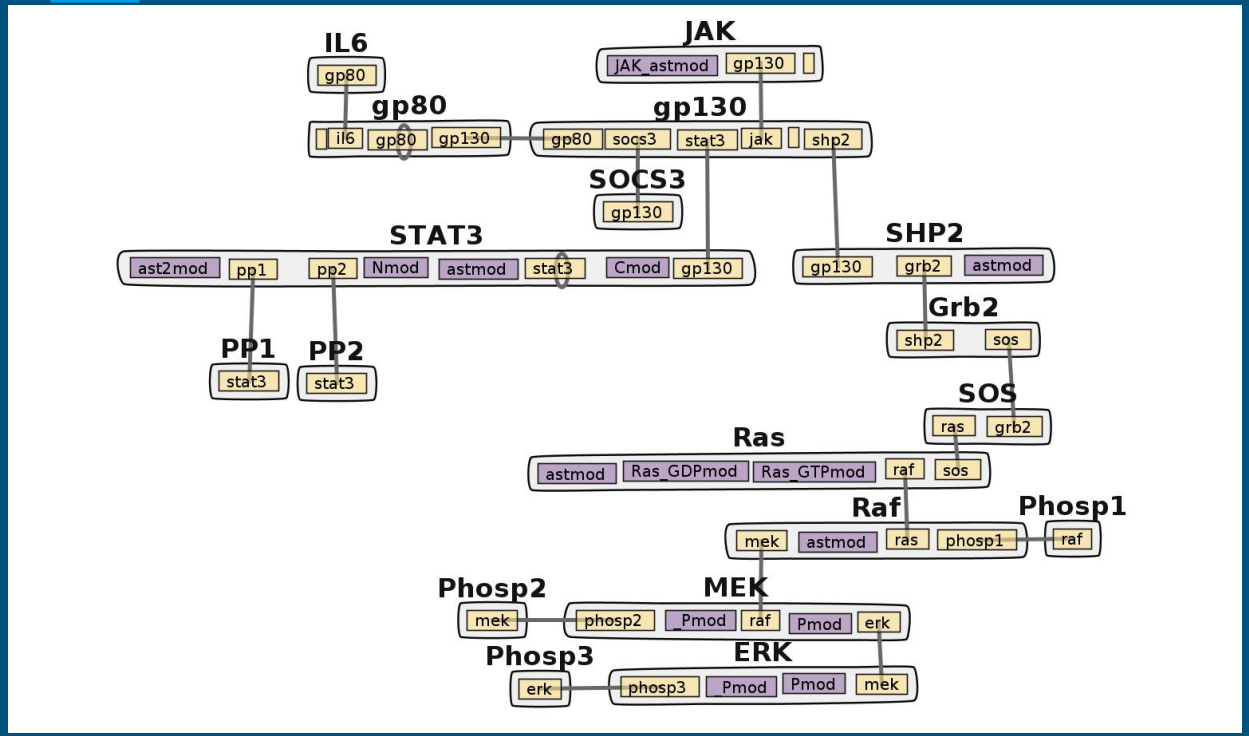
BIOMD 151

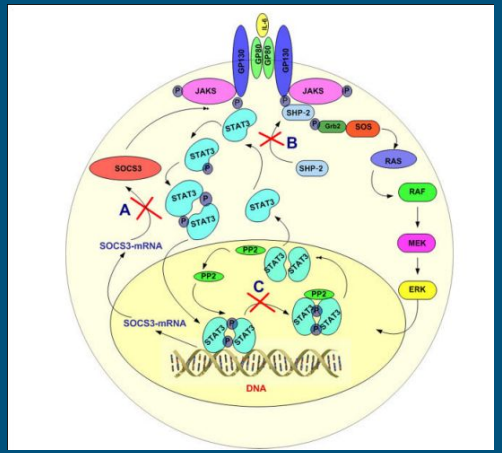
Modeling Regulatory Mechanisms in IL-6 Signal Transduction in Hepatocytes

Singh A et al.



Contact map

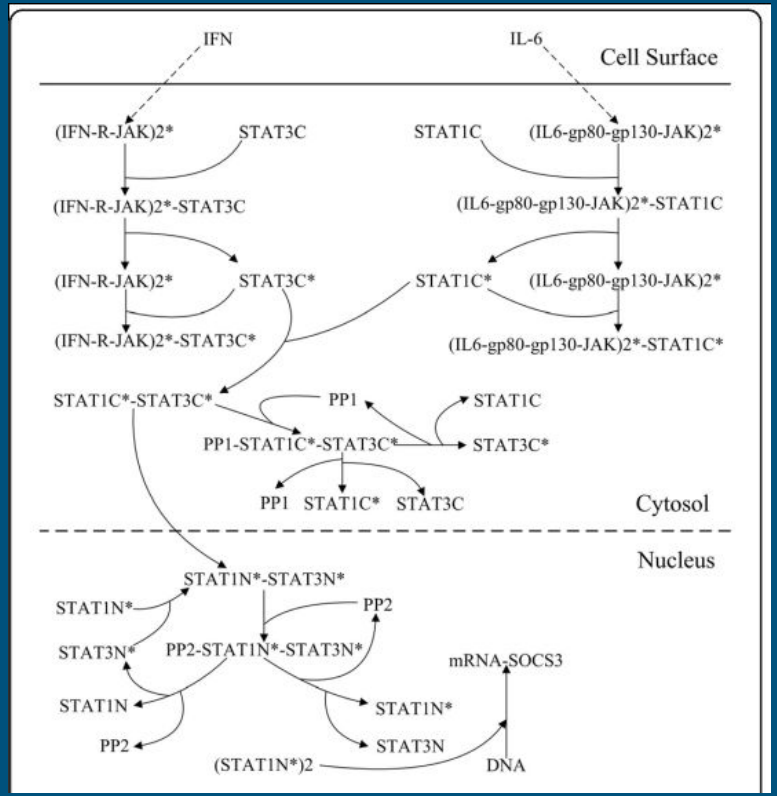

IL6
gp80
JAK
JAK_astmod
gp130
gp80
il6
gp80
gp130
gp130
gp80
socs3
stat3
jak
shp2
SOCS3
gp130
STAT3
ast2mod
pp1
pp2
Nmod
astmod
stat3
Cmod
gp130
SHP2
gp130
grb2
astmod
Grb2
shp2
sos
PP1
stat3
PP2
stat3
SOS
ras
grb2
Ras
astmod
Ras_GDPmod
Ras_GTPmod
raf
sos
Raf
Phosp1
mek
astmod
ras
phosp1
raf
Phosp2
MEK
mek
phosp2
_Pmod
raf
Pmod
erk
Phosp3
ERK
erk
phosp3
Pmod
Pmod
mek


JAKS
SHP-2
STAT3
SOCS3
SOS
RAS
RAF
MEK
ERK
A
B
C
SOCS3-mRNA
PP2
DNA

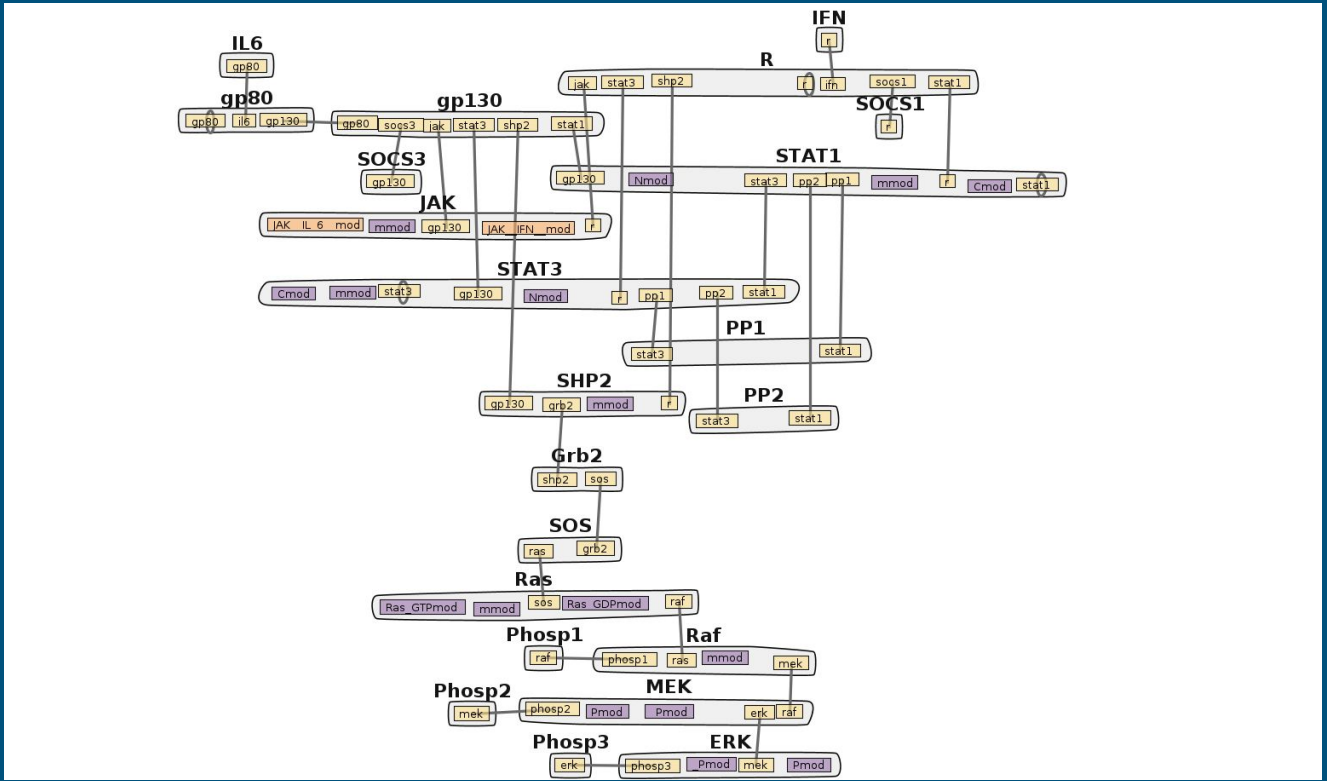
BIOMD 543

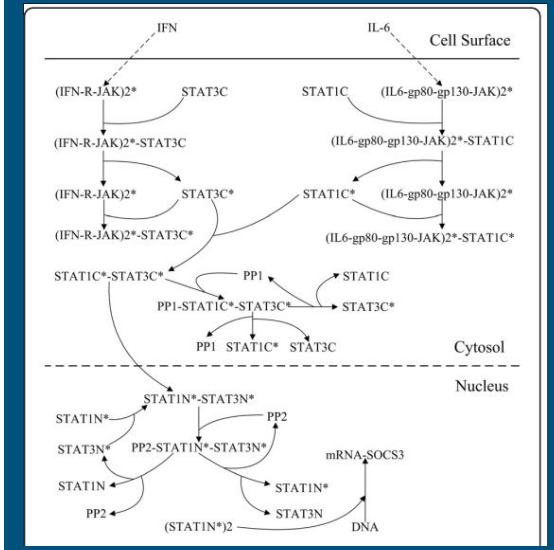
Elucidating the crosstalk mechanism between IFN-gamma and IL-6 via mathematical modelling

Qi YF et al.



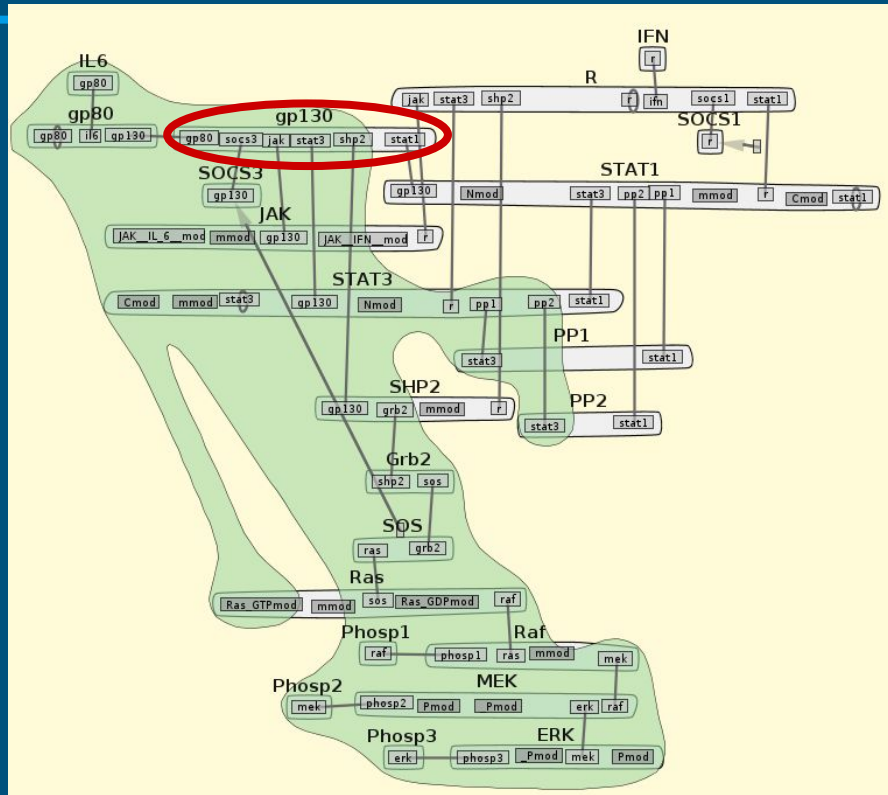
BIOMD 543


IL6
gp80
IFN
R
gp80
gp130
SOCS1
SOCS3
STAT1
JAK
STAT3
PP1
SHP2
PP2
Grb2
SOS
Ras
Raf
Phosp1
MEK
Phosp2
ERK
Phosp3

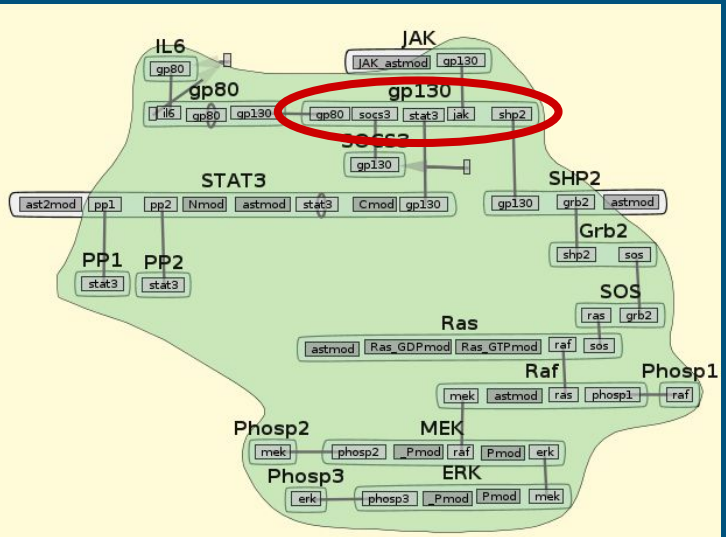

IFN
IL-6
Cell Surface
(IFN-R-JAK)2*
STAT3C
STAT1C
(IL6-gp80-gp130-JAK)2*
(IFN-R-JAK)2*-STAT3C
(IL6-gp80-gp130-JAK)2*-STAT1C
(IFN-R-JAK)2*
STAT3C*
STAT1C*
(IL6-gp80-gp130-JAK)2*
(IFN-R-JAK)2*-STAT3C*
(IL6-gp80-gp130-JAK)2*-STAT1C*
STAT1C*-STAT3C*
PP1
STAT1C
PP1-STAT1C*-STAT3C*
STAT3C*
PP1 STAT1C* STAT3C
Cytosol
Nucleus
STAT1N*-STAT3N*
STAT1N*
PP2
STAT3N*
PP2-STAT1N*-STAT3N*
mRNA-SOCS3
STAT1N
STAT1N*
PP2
STAT3N
(STAT1N*)2
DNA

Structure comparison (MOSBIE)

BMD 543

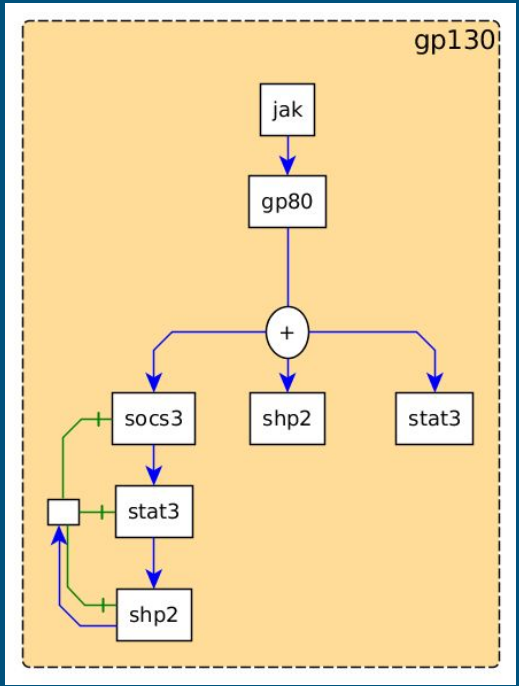


BMD 151

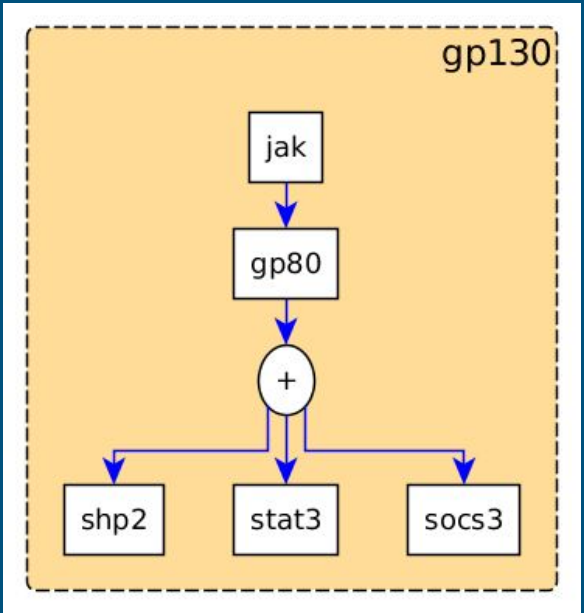


Comparing from a process centric view

BIOMD151



BIOMD543



Comparison of atomized models using STD's

- Atomization resolves sites of protein-protein interaction and modification
- STD's represent the causal relationships between sites
- STD's enable visual comparison of models' processes.
- **All the information required to programmatically generate contact maps and STD's is contained inside an RBM description.**

Overall conclusions

- A full model specification should include intra-model species relationships and help towards the definition of inter-model species relationships
- Rule-based modeling facilitates this by encoding structural and process information through the use of a graph representation and graph operations
- Atomizer leverages the power of RBM’s for reaction-network models by recovering structural and process information from the original system

Now you can atomize too!

http://ratomizer.appspot.com/translate

```
BNG2.pl <sbmlfile.xml>
```

```
https://github.com/RuleWorld/atomizer
```

Acknowledgments

Faeder Lab:

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

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

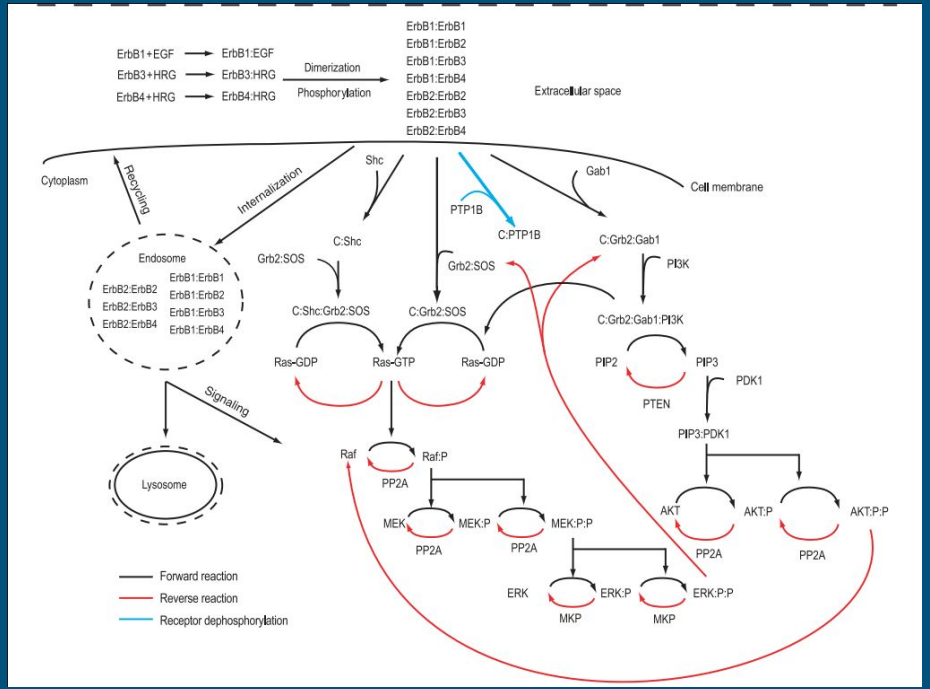
Thank you!

Bonus slides!

Using atomizer to find model quirks: BMD255

Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data.

Chen WW et al.



Using atomizer to find model quirks: BMD255

This model presents several atomization challenges:

- Not all complexes are derived from basic reactions
- Several, not immediately compatible ways of producing complexes.
- Scarce annotation information
- Some misspellings

Not all complexes are derived from basic reactions

Example:

- `EGF:ErbB1#P` is not individually defined anywhere in the reaction system.

Seemingly conflicting definitions

```
v861 2(EGF:ErbB1)#P + ATP -> 2(EGF:ErbB1_h:ATP)-FullActive k123 kd123

v857 2(EGF:ErbB1_h:ATP) + ATP -> 2(EGF:ErbB1_h:ATP)-FullActive k122
kd122

    v852 EGF:ErbB1_h:ATP + EGF:ErbB1_h:ATP -> 2(EGF:ErbB1_h:ATP) k2 kd2
```

`2(EGF:ErbB1_h:ATP)-FullActive`

Has two production pathways!

'ATP', 'EGF:ErbB1#P', 'EGF:ErbB1#P'
'ATP', 'ATP', 'ATP', 'EGF', 'EGF', 'ErbB1_h', 'ErbB1_h'

Misspelling Example

```
v17: ErbB4 +  EGF_ErbB1_Inh  -> EGF_ErbB1_Inh_ErB4  k2b
kd2b
```

How to solve them?

All of these are identified through Atomizer, which can also help in resolving them