

# Graphical Notations to Represent System Biology

Stuart Moodie

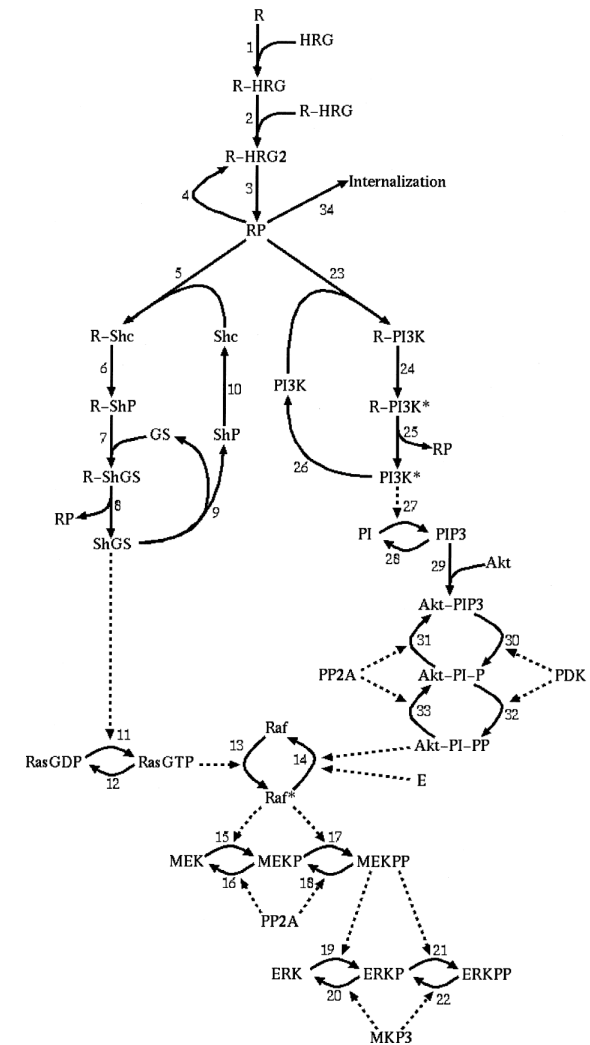
# Overview

- Background on Graphical Notations
- Overview of previous notations
- SBGN

# Graphical Representations - Why?

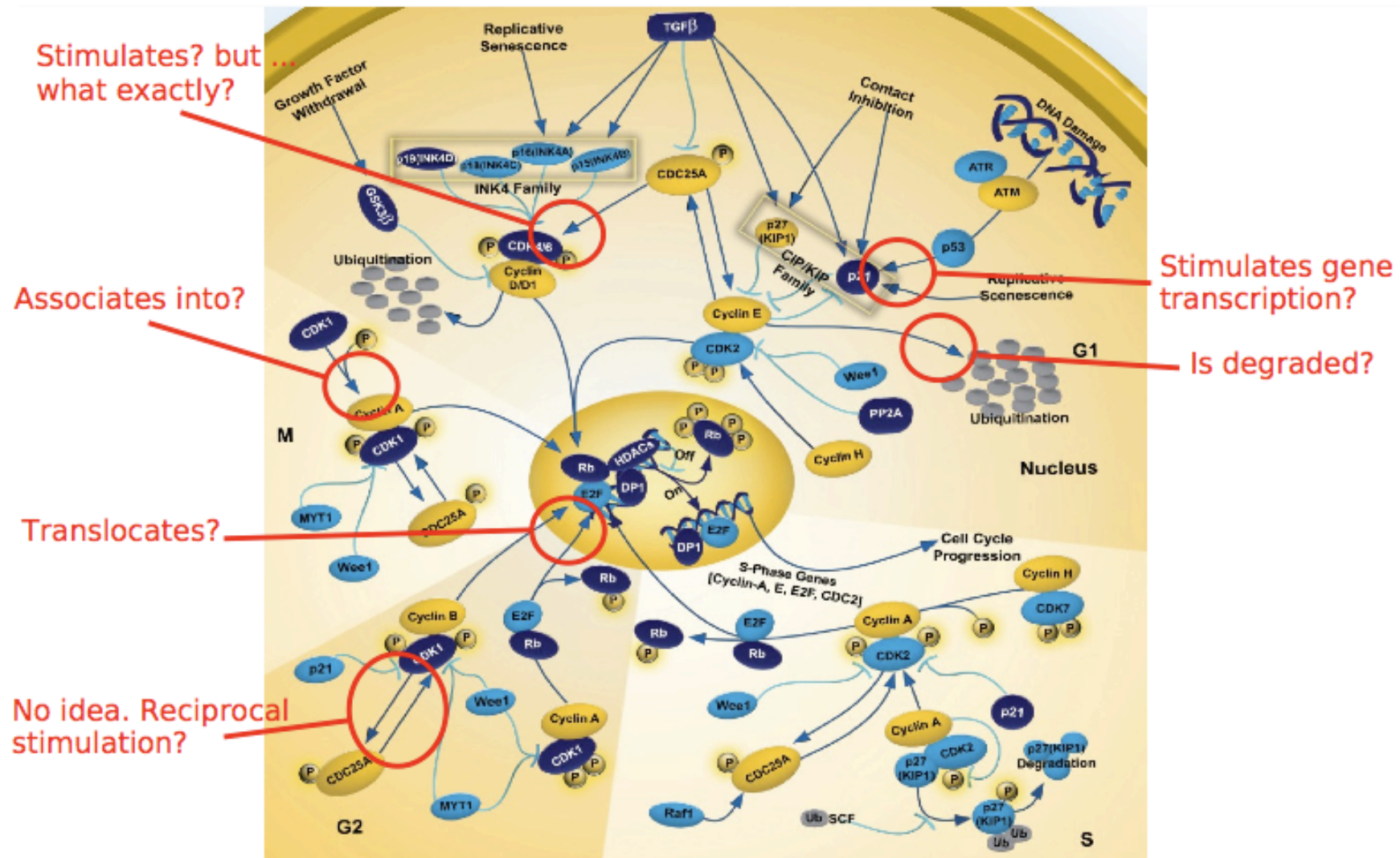
- Summarise information
- Networks easier to understand visually

Reaction number	Rate equation
1	$k_1[R][HRG] - k_{-1}[R-HRG]$
2	$k_2[R-HRG] - k_{-2}[R-HRG2]$
3	$k_3[R-HRG2] - k_{-3}[RP]$
4	$V_4[RP](K_4 + [RP])$
5	$k_5[RP][Shc] - k_{-5}[R-Shc]$
6	$k_6[R-Shc] - k_{-6}[R-ShP]$
7	$k_7[R-ShP][GS] - k_{-7}[R-ShGS]$
8	$k_8[R-ShGS] - k_{-8}[ShGS][RP]$
9	$k_9[ShGS] - k_{-9}[GS][ShP]$
10	$V_{10}[ShP](K_{10} + [ShP])$
11	$k_{11}[ShGS][RasGDP](K_{11} + [RasGDP])$
12	$V_{12}[RasGTP](K_{12} + [RasGTP])$
13	$k_{13}[RasGTP][Raf] - k_{-13}[Raf]$
14	$k_{14}[Raf][Ad-Pi-PP] + [E][Raf][K_{14} + [Raf*]]$
15	$k_{15}[Raf*][MEK] / \left( K_{15} \left( 1 + \frac{[RasGTP]}{K_{15}} \right) + [MEK] \right)$
16	$k_{16}[PP2A][MEKPP] / \left( K_{16} \left( 1 + \frac{[RasGTP]}{K_{16}} + \frac{[MEKPP]}{K_{16}} + \frac{[MEKPP]^2}{K_{16}} \right) + [MEKPP] \right)$
17	$k_{17}[Raf*][MEKPP] / \left( K_{17} \left( 1 + \frac{[RasGTP]}{K_{17}} \right) + [MEKPP] \right)$
18	$k_{18}[PP2A][MEKPP] / \left( K_{18} \left( 1 + \frac{[RasGTP]}{K_{18}} + \frac{[MEKPP]}{K_{18}} + \frac{[MEKPP]^2}{K_{18}} \right) + [MEKPP] \right)$
19	$k_{19}[MEKPP][ERK] / \left( K_{19} \left( 1 + \frac{[RasGTP]}{K_{19}} \right) + [ERK] \right)$
20	$k_{20}[MEKPP][ERKPP] / \left( K_{20} \left( 1 + \frac{[RasGTP]}{K_{20}} \right) + [ERKPP] \right)$
21	$k_{21}[MEKPP][ERKPP] / \left( K_{21} \left( 1 + \frac{[RasGTP]}{K_{21}} \right) + [ERKPP] \right)$
22	$k_{22}[MEKPP][ERKPP] / \left( K_{22} \left( 1 + \frac{[RasGTP]}{K_{22}} \right) + [ERKPP] \right)$
23	$k_{23}[RP][PI3K] - k_{-23}[R-Pi-3K]$
24	$k_{24}[R-Pi-3K] - k_{-24}[R-Pi-3K*]$
25	$k_{25}[R-Pi-3K*] - k_{-25}[RP][Pi-3K*]$
26	$V_{26}[Pi3K*](K_{26} + [Pi-3K*])$
27	$k_{27}[Pi3K*][PI] - k_{-27}[PI]$
28	$V_{28}[Pi3K*][PI](K_{28} + [Pi3K*])$
29	$k_{29}[Pi3K*][Akt] - k_{-29}[Akt-Pi3K]$
30	$V_{30}[Akt-Pi3K] / \left( K_{30} \left( 1 + \frac{[Akt-Pi3K]}{K_{30}} \right) + [Akt-Pi3K] \right)$
31	$k_{31}[PP2A][Akt-Pi3K] / \left( K_{31} \left( 1 + \frac{[Akt-Pi3K]}{K_{31}} + \frac{[Akt-Pi3K]^2}{K_{31}} \right) + [Akt-Pi3K] \right)$
32	$V_{32}[Akt-Pi3K] / \left( K_{32} \left( 1 + \frac{[Akt-Pi3K]}{K_{32}} \right) + [Akt-Pi3K] \right)$
33	$k_{33}[PP2A][Akt-Pi3K] / \left( K_{33} \left( 1 + \frac{[Akt-Pi3K]}{K_{33}} + \frac{[Akt-Pi3K]^2}{K_{33}} \right) + [Akt-Pi3K] \right)$
34	$k_{34}[RP] - k_{-34}[Internalization]$



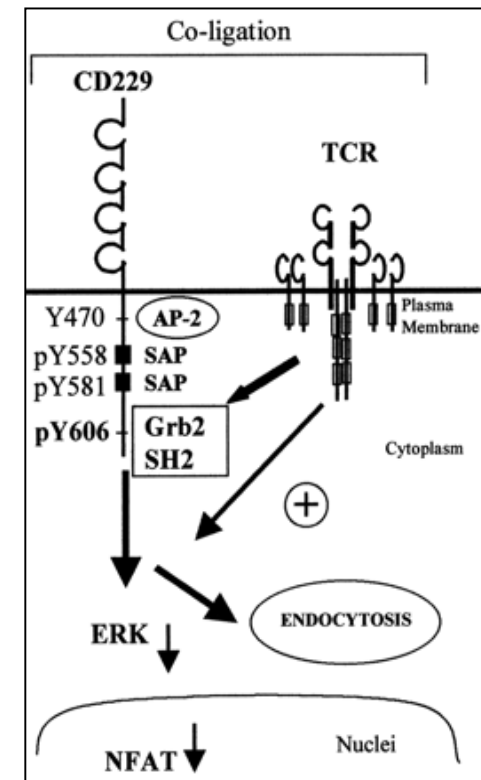
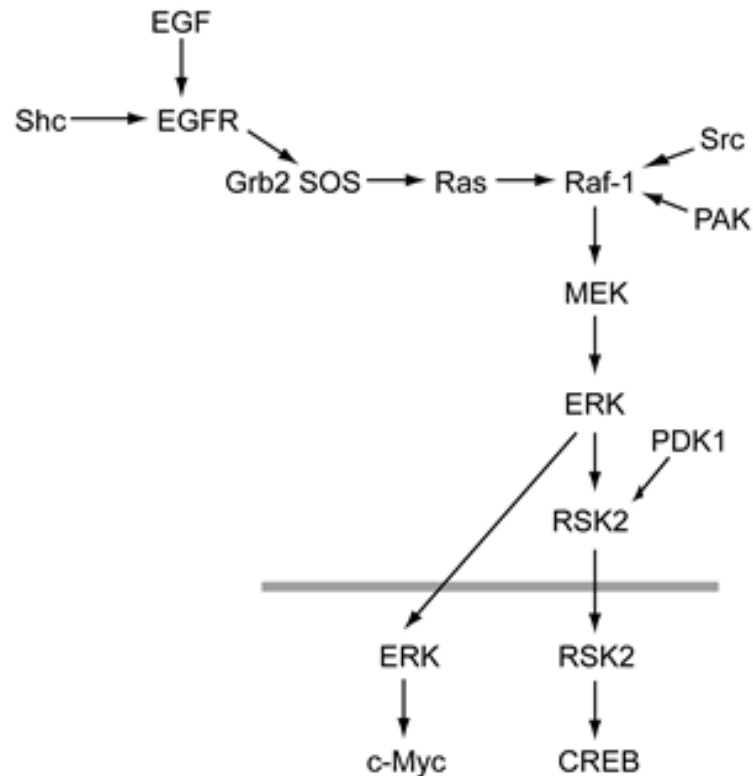
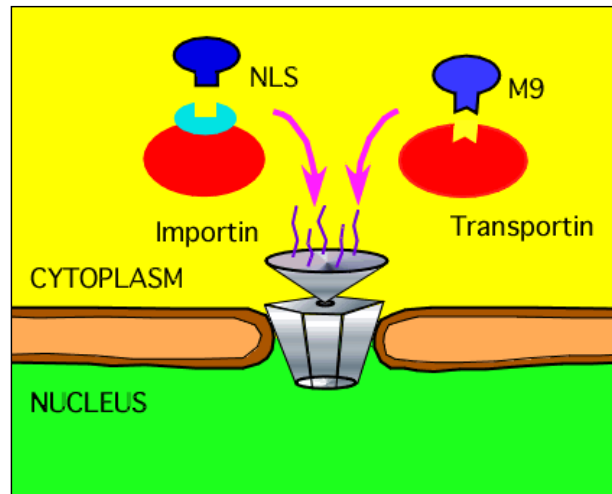
from Hatakeyma *et al.* Biochem J. (2003) 373. 451-463.

# Can this be understood by a Biologist?





# Graphical Representations - Why Not?

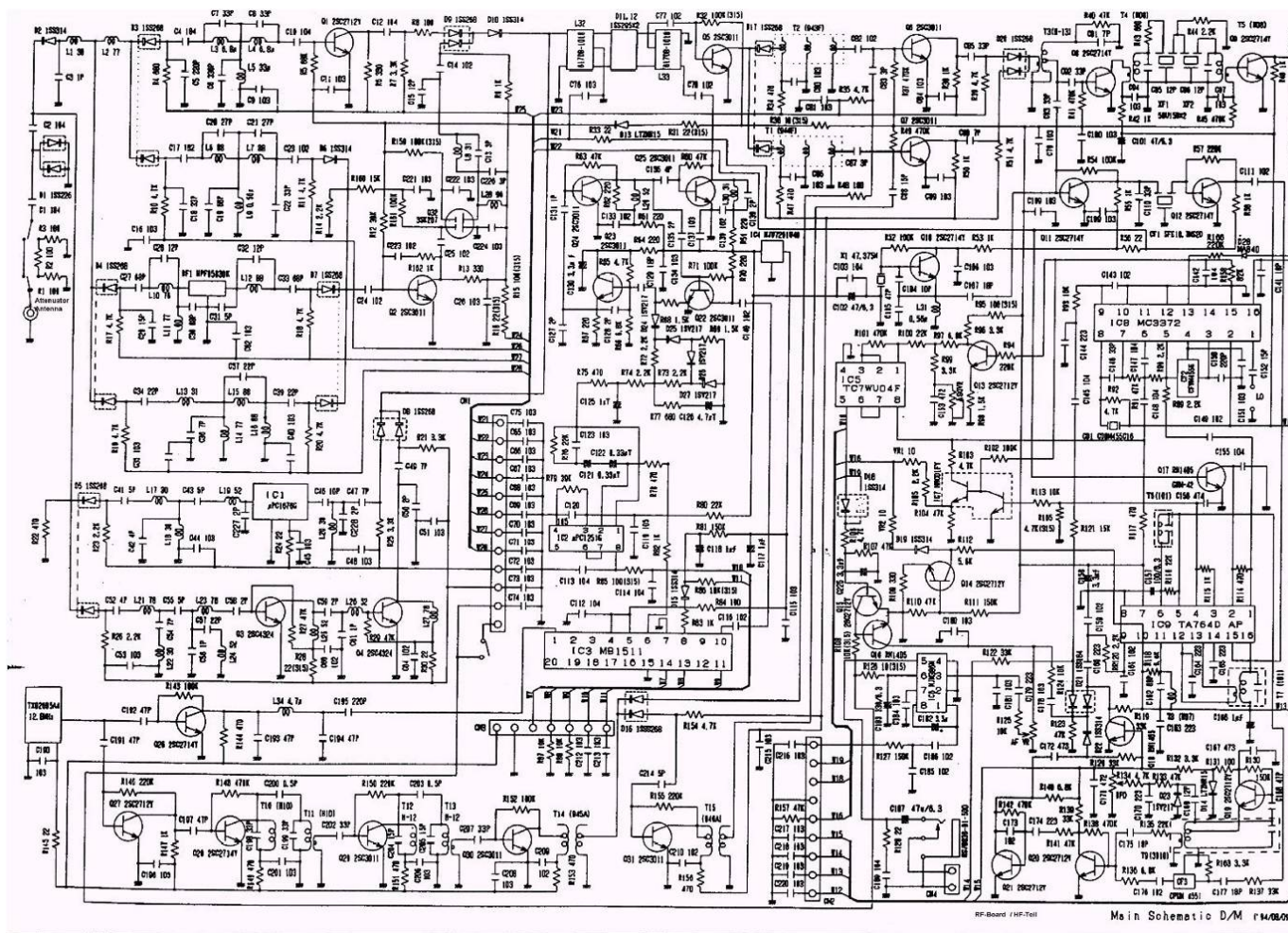


- Ambiguous
- Not amenable to computational analysis

# Graphical Notations: “Formalised” Pictures

- Visual Language
- Rules and guidelines ensure:
  - reader can unambiguously understand what writer meant
  - amenable to computation manipulation/analysis

# Classic Example





# Pathways are not Electronic Circuits

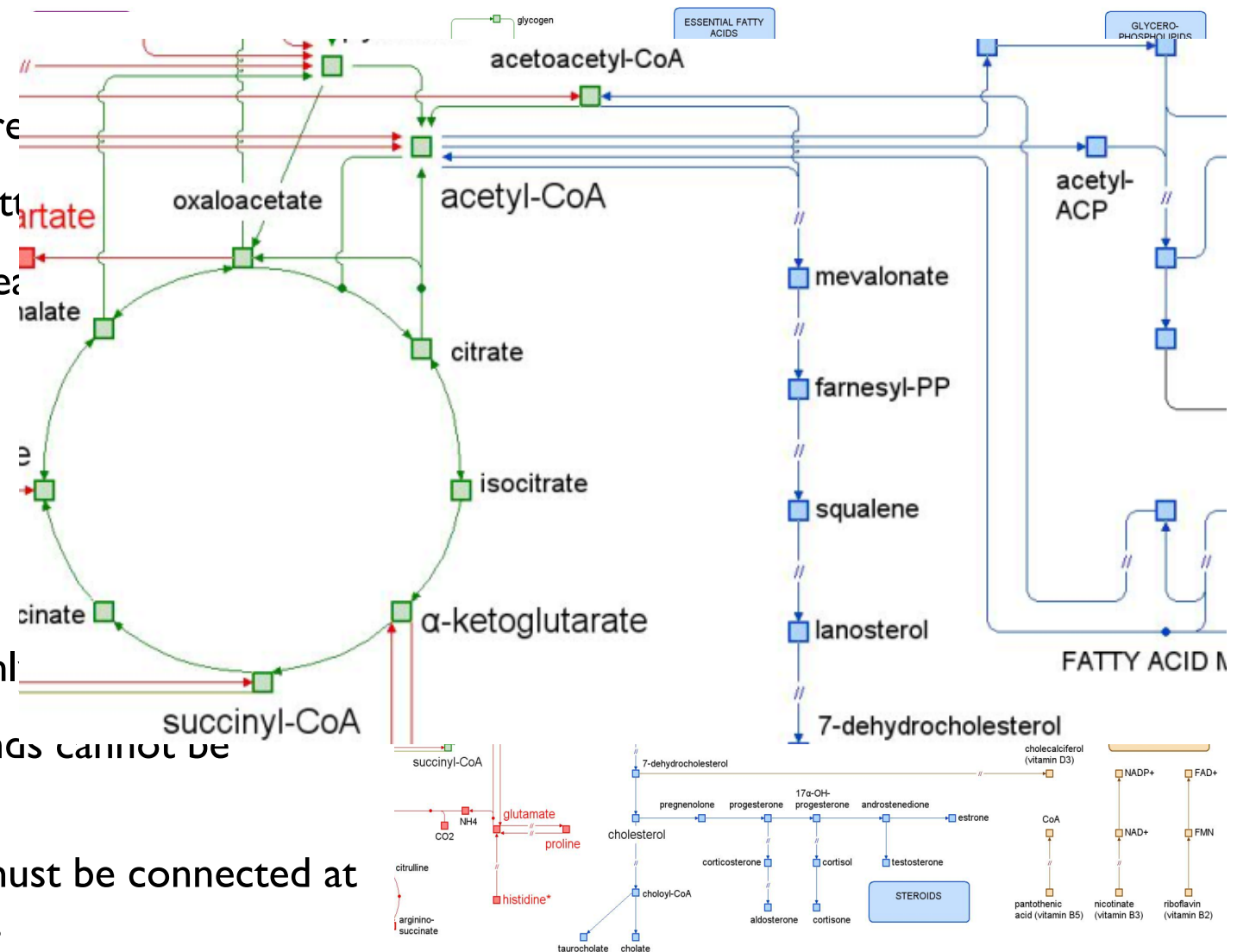
- Circuits designed by humans
  - Follow a defined set of rules
  - Boundaries well defined
- Pathways
  - Don't know all the "rules"
  - Boundaries not clear
  - Knowledge often incomplete

# Graphical Notation

- Has an underlying conceptual abstraction
  - a model of its world
- Visual Language
  - vocabulary (symbols/**glyphs**)
  - **syntax** (basic assembly or glyphs)
  - grammar (rules based on meaning - **semantics**)

# Notation Example

- Abstraction
  - biochemical reactions
  - enzymes omitted
  - net flow of reactions
- Language
  - Symbols
  - Syntax
  - Semantics
    - Names only
    - Compounds cannot be orphaned
    - Arrows must be connected at both ends





# Types of Graphical Notation

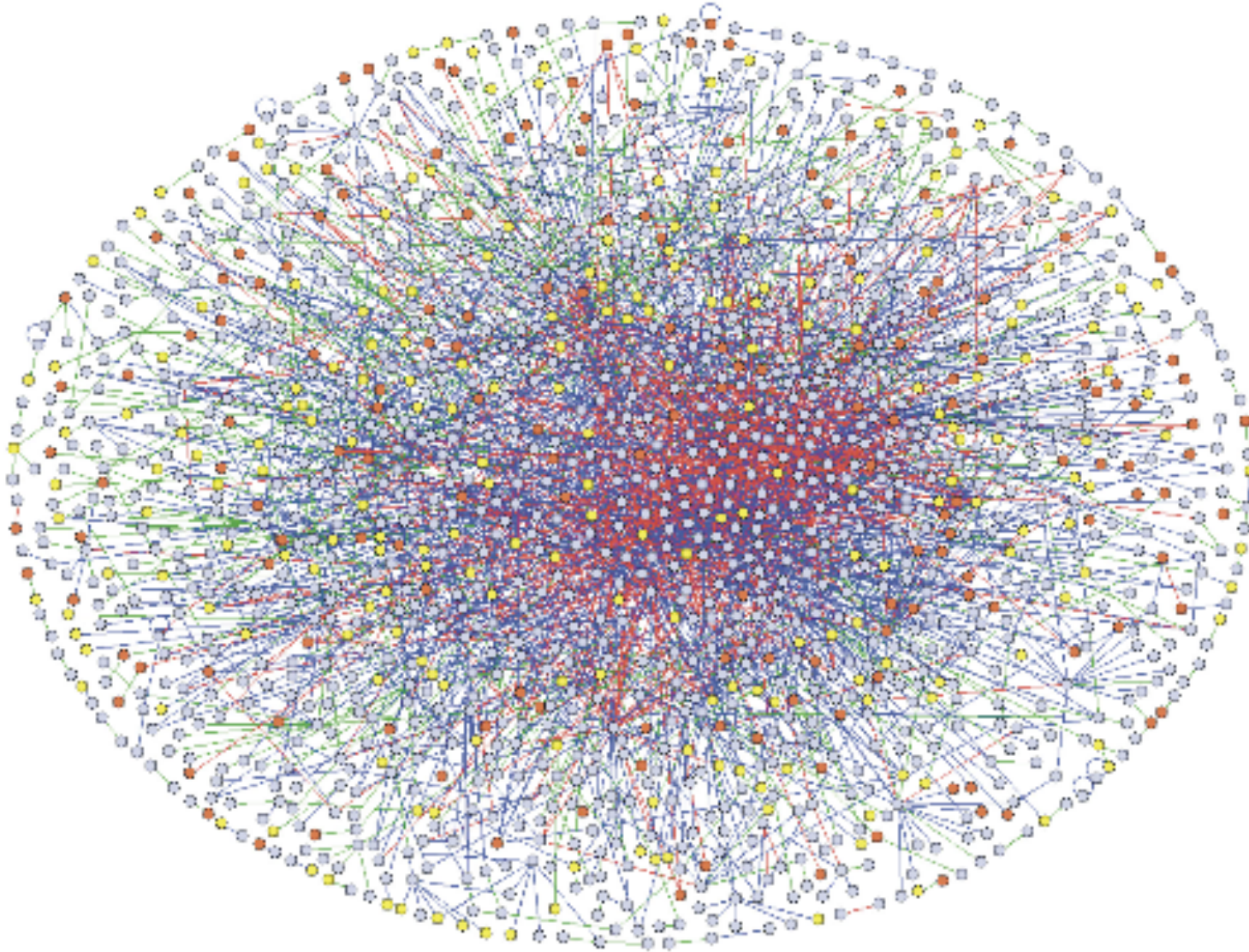
- Entity Relationship
  - Protein Interaction Diagrams
  - MIM (Kohn)
  - SBGN-ER
- State Transition/Process
  - EPN
  - PDN (Kitano)
  - SBGN-PD

# Overview of Pre-SBGN Notation





# Entity Relationship View





# MIMs

- Inspired by circuit diagrams
- Invented by Kurt Kohn, NIH
- First published 1999
- Presented to SBGN by Marit Aladjem
- Slides taken from Aladjem presentation at SBGN-1



Molecular Biology of the Cell  
Vol. 17, 1–13, January 2006

Essay

# Molecular Interaction Maps of Bioregulatory Networks: A General Rubric for Systems Biology

Kurt W. Kohn, Mirit I. Aladjem, John N. Weinstein, and Yves Pommier

Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Submitted September 1, 2005; Accepted October 21, 2005

Monitoring Editor: Gerard Evan

Molecular Systems Biology (2006) doi:10.1088/msb.100088  
© 2006 EMBO and Nature Publishing Group. All rights reserved 1744-4292/06  
www.molecular-systems-biology.com  
Artisanian 53



## REVIEW

## Depicting combinatorial complexity with the molecular interaction map notation

Kurt W. Kohn\*, Mirit I. Aladjem, Sohyoung Kim, John N. Weinstein  
and Yves Pommier

Laboratory of Molecular Pharmacology, Center for Cancer Research, National  
Cancer Institute, NIH, Bethesda, MD, USA

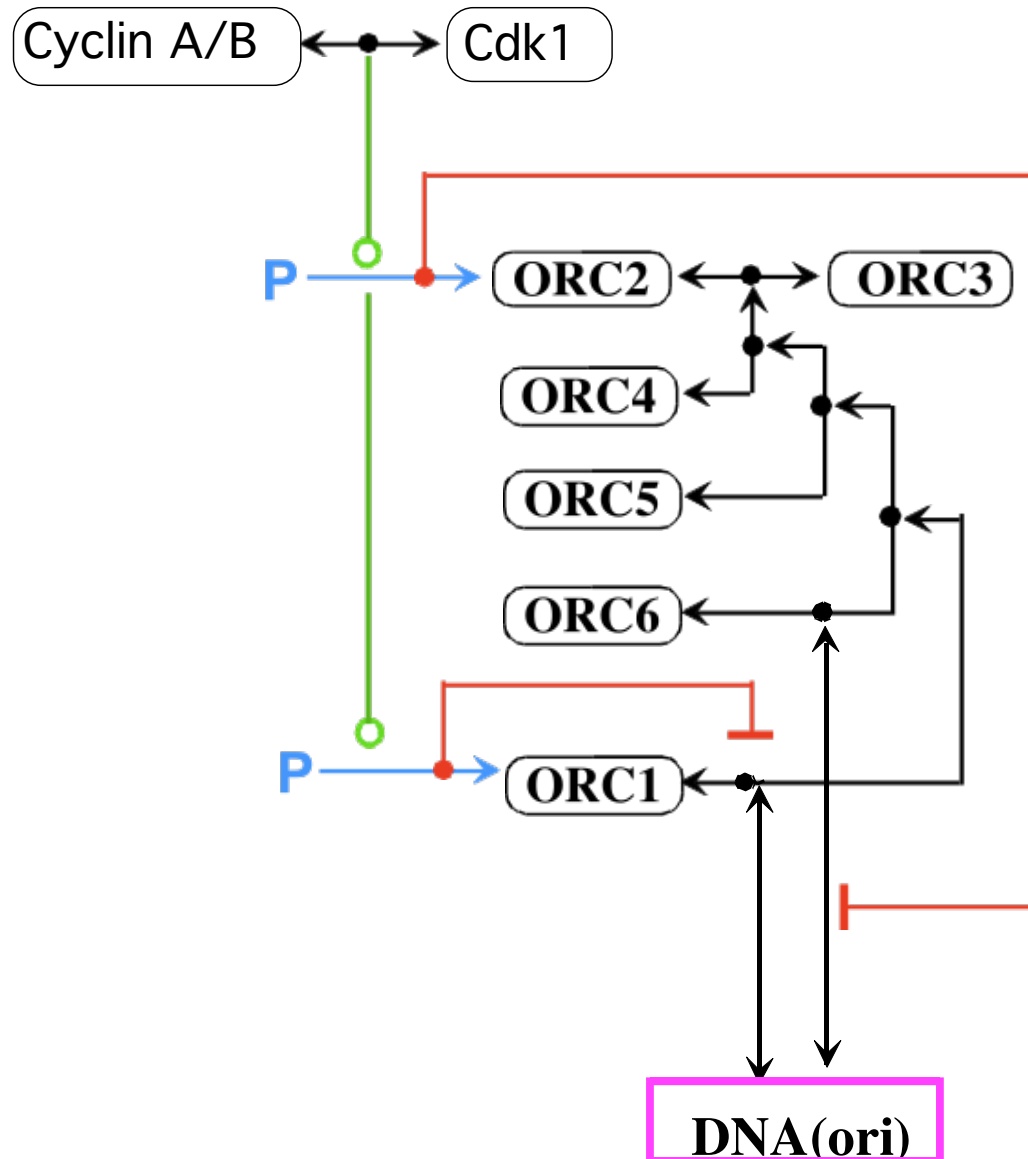
\* Corresponding author. Laboratory of Molecular Pharmacology, Center for  
Cancer Research, National Cancer Institute, Building 37, Room 5085, NIH,  
Bethesda, MD 20892, USA. Tel: +1 301 496 2769; Fax: +1 301 402 0762;  
E-mail: kohnk@dc37a.nc.nih.gov

Received 2.1.05; accepted 2.7.05

Two of the best developed notations are the molecular  
interaction maps (MIMs) that we have described (Kohn,  
1990, 2001; Kohn *et al.*, 2006) and the 'process diagrams'  
described by Kitano *et al.* (Kitano, 2003; Kitano *et al.*, 2005). We  
recently discussed the strengths and weaknesses of the various  
notations that have been proposed (Kohn *et al.*, 2006). These  
include, in addition to the MIM and process diagram notations,  
the computer-aided design (CAD)-like diagrams produced by  
CellDesigner (Funahashi *et al.*, 2003), a software suite called  
CADLIVE (Kurata *et al.*, 2003), the automated diagrams of  
Cook *et al.* (2001), and BIOCARX's connection diagrams

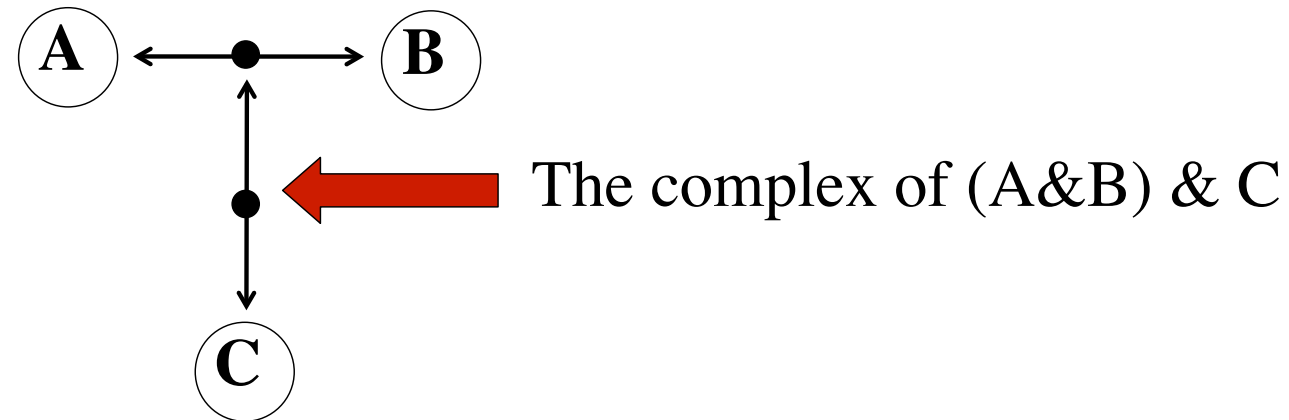


## Assembly of a multimolecular complex: ORC, the origin recognition complex (involved in cell cycle regulation)





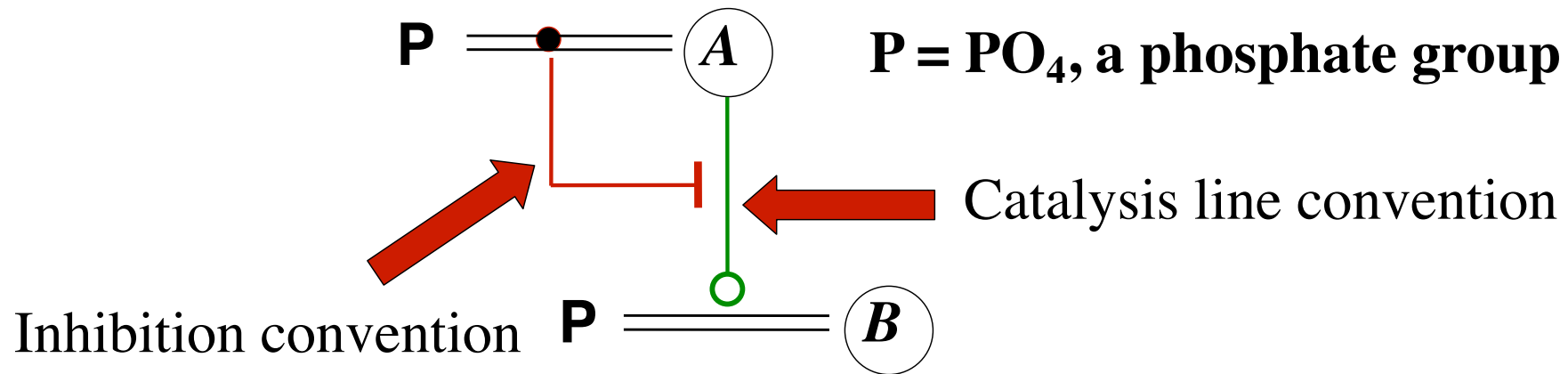
## Multi-protein complexes



Each molecule appears only once per diagram.  
Interaction outcomes - complexes or modified molecules - are depicted as nodes on the interaction lines.



## Covalent modification (e.g., protein phosphorylation)



Inhibitory phosphorylation:

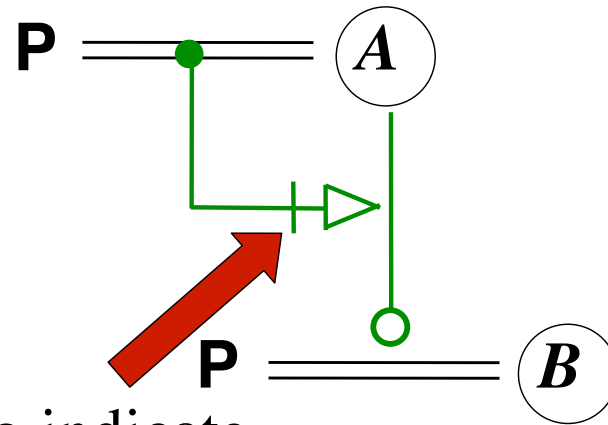
Phosphorylation of *A* blocks the kinase activity of *A*.





Activating phosphorylation:

The phosphorylated form of kinase *A* is the active form, phosphorylates *B*



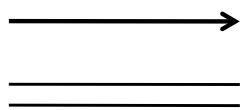
Bar added to indicate  
obligatory requirement



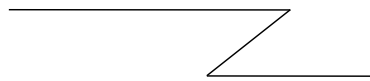
## Reactions



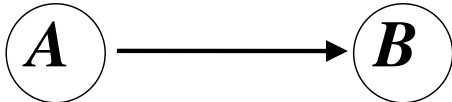
Binding (non-covalent)



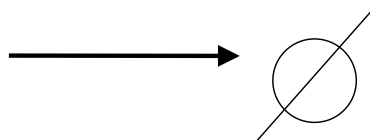
Covalent Modification  
(e.g. phosphorylation)



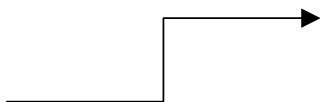
Bond cleavage  
(e.g. Phosphatase)



Stoichiometric  
Conversion (A to B)



Degradation



Transcription/translation



Dimerization



Transport

## Contingencies



Catalysis



Stimulation



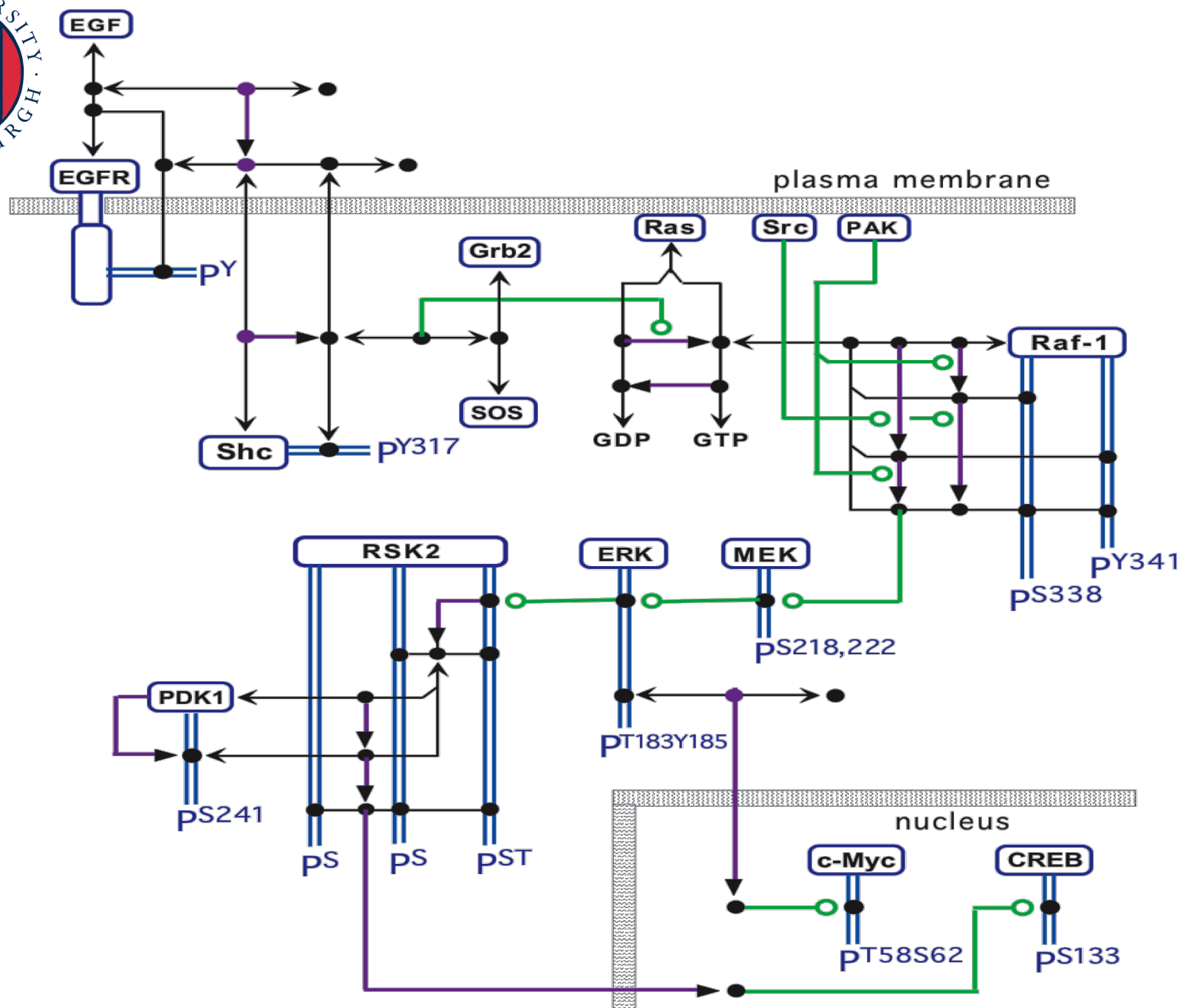
Stimulation  
required



Inhibition

*Reactions operate on molecular species; contingencies operate on reactions, or on other contingencies; reaction outcomes (nodes) are treated as molecular species.*







# MIMs

- Benefits
  - Compact
  - Established
  - Useful to biologists
- Drawbacks
  - Learning curve
  - No tool support

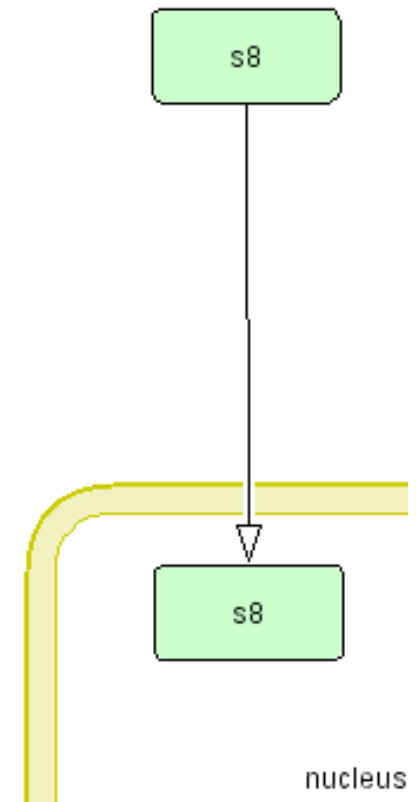
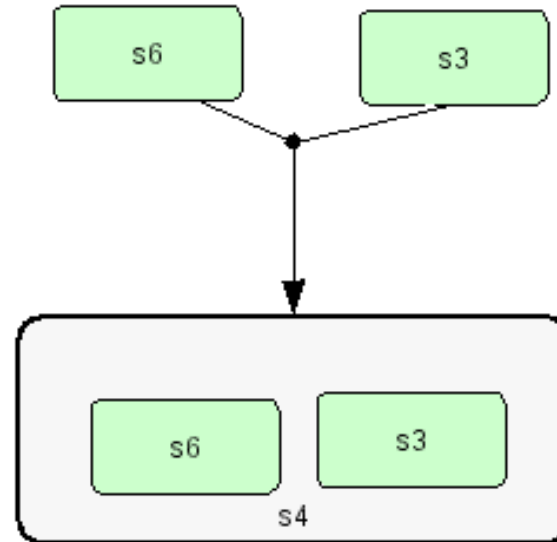
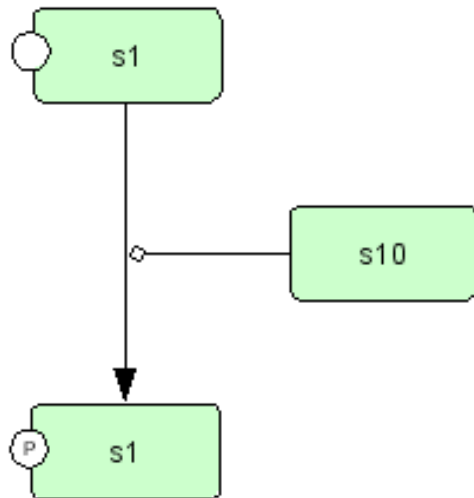


# State Transition Diagrams

- PDN & EPN



# State Transition/Process Diagrams

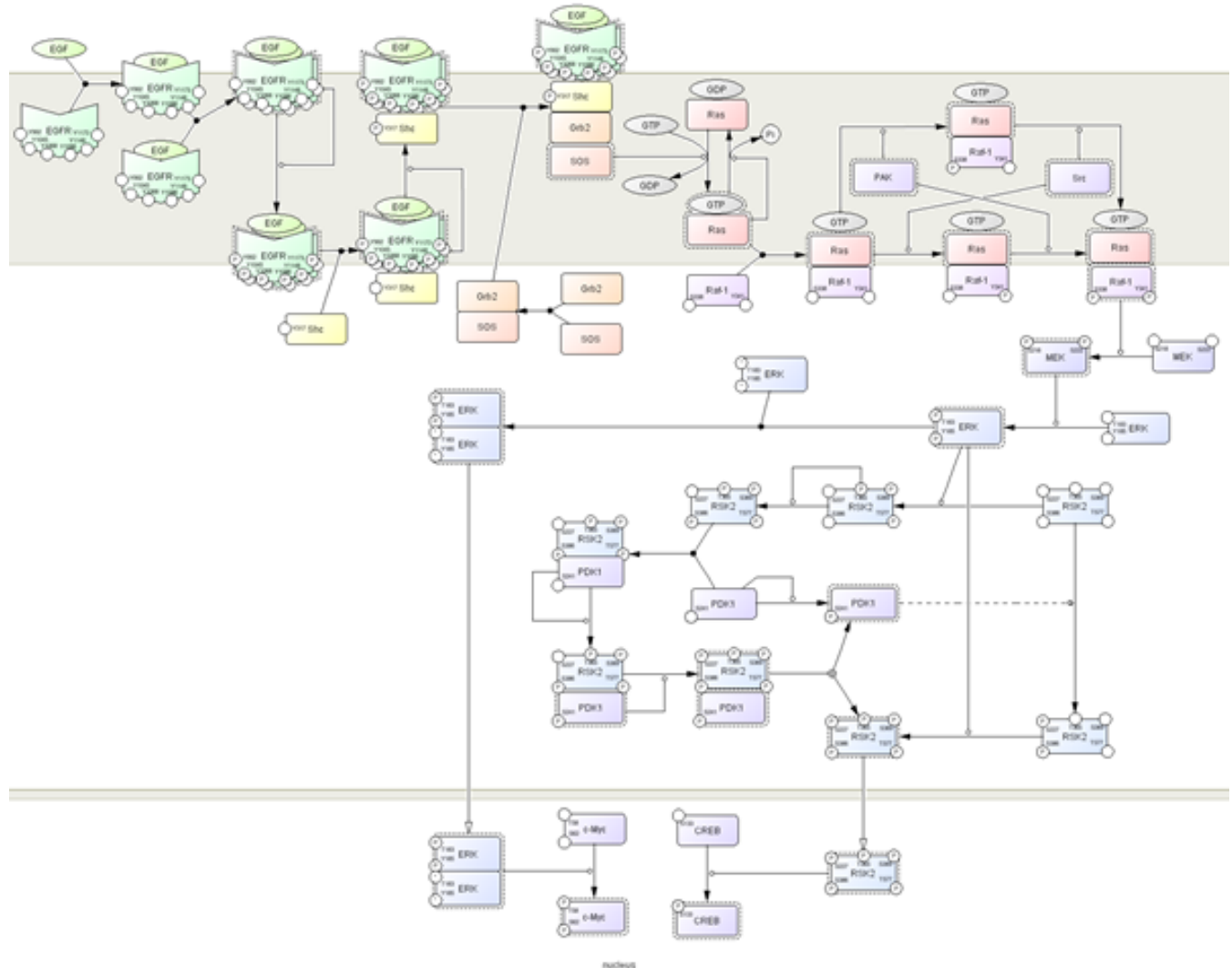
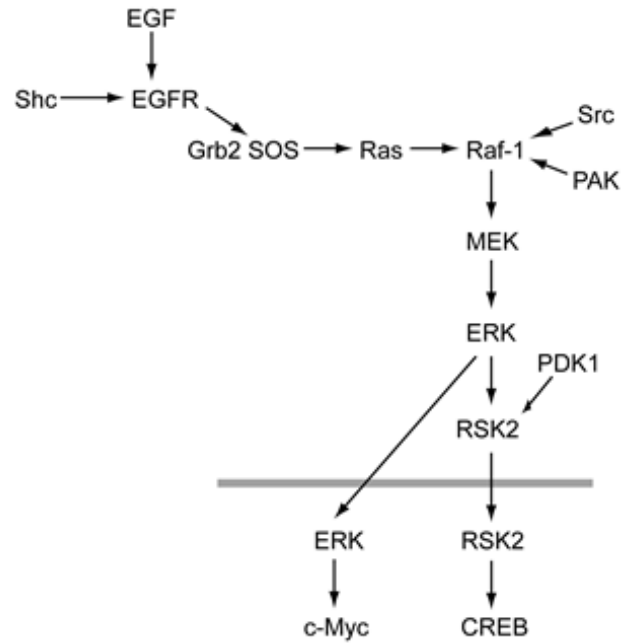




- Slides taken from Kitano presentation at SBGN-1

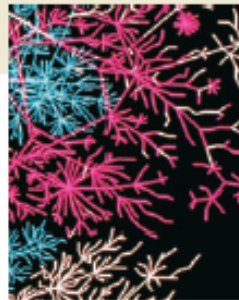


# PDN (Kitano Notation)





<http://www.nature.com/nbt/journal/v23/n8/abs/nbt1111.html>



computational  
BIOLOGY

PERSPECTIVE

## Using process diagrams for the graphical representation of biological networks

Hiroaki Kitano<sup>1–4</sup>, Akira Funahashi<sup>1,3,4</sup>, Yukiko Matsuoka<sup>1,3</sup> & Kanae Oda<sup>1,4</sup>

With the increased interest in understanding biological networks, such as protein-protein interaction networks and gene regulatory networks, methods for representing and communicating such networks in both human- and machine-readable form have become increasingly important. Although there has been significant progress in machine-readable representation of networks, as exemplified by the Systems Biology Mark-up Language (SBML) (<http://www.sbml.org>) issues in human-readable representation have been largely ignored. This article discusses human-readable diagrammatic representations and

when representing interactions within larger networks. Therefore, there is a need for diagrams that contain unambiguous process information in the symbols used and that can be transferred to standard machine-readable codes such as SBML for computational analysis<sup>1</sup>.

Circuit schematic diagrams used in electronics are ideal examples of a graphical diagram. Engineers can reproduce the circuits drawn in the schematic diagrams without substantial additional information, because the diagrams are unambiguously defined, contain sufficient information and are based on well-accepted standards.

Kurt Kohn was the first to produce canonical representations for

## State Node Symbols

Protein	
Receptor	
Ion Channel (Closed)	
Ion Channel (Open)	
Truncated Protein	
Gene	
RNA	
Anti-sense RNA	
Ion	
Simple Molecule	
Unknown	
Phenotype	
Homodimer / N-mer with N stacked symbols	
Active Protein	

## Arc Symbols (Transit Node and Edges)

State transition	
Known transition omitted	
Unknown transition	
Bidirectional transition	
Translocation	
Association	
Dissociation	
Truncation	
Promote transition	
Inhibit transition	
Add reactant	
Add product	
AND	
OR	

## Reduced Notation Symbols

### Class-I Reduced Notation

Degradation	
Transcription	
Translation	
Module	

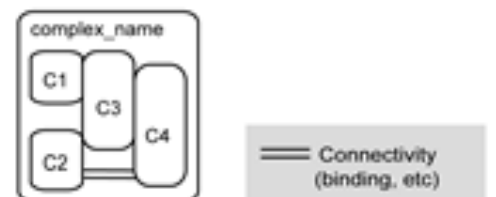
### Class-II Reduced Notation (Viewer Only)

Activation/Inhibition/Modification	
------------------------------------	--

### Node Structure

Residue modification		<ul style="list-style-type: none"> <li> phosphorylated</li> <li> acetylated</li> <li> ubiquitinated</li> <li> methylated</li> <li> hydroxylated</li> </ul>
	<ul style="list-style-type: none"> <li> empty</li> <li> don't care</li> <li> unknown</li> </ul>	

### Complex State Node



### Promotor and coding structure for gene

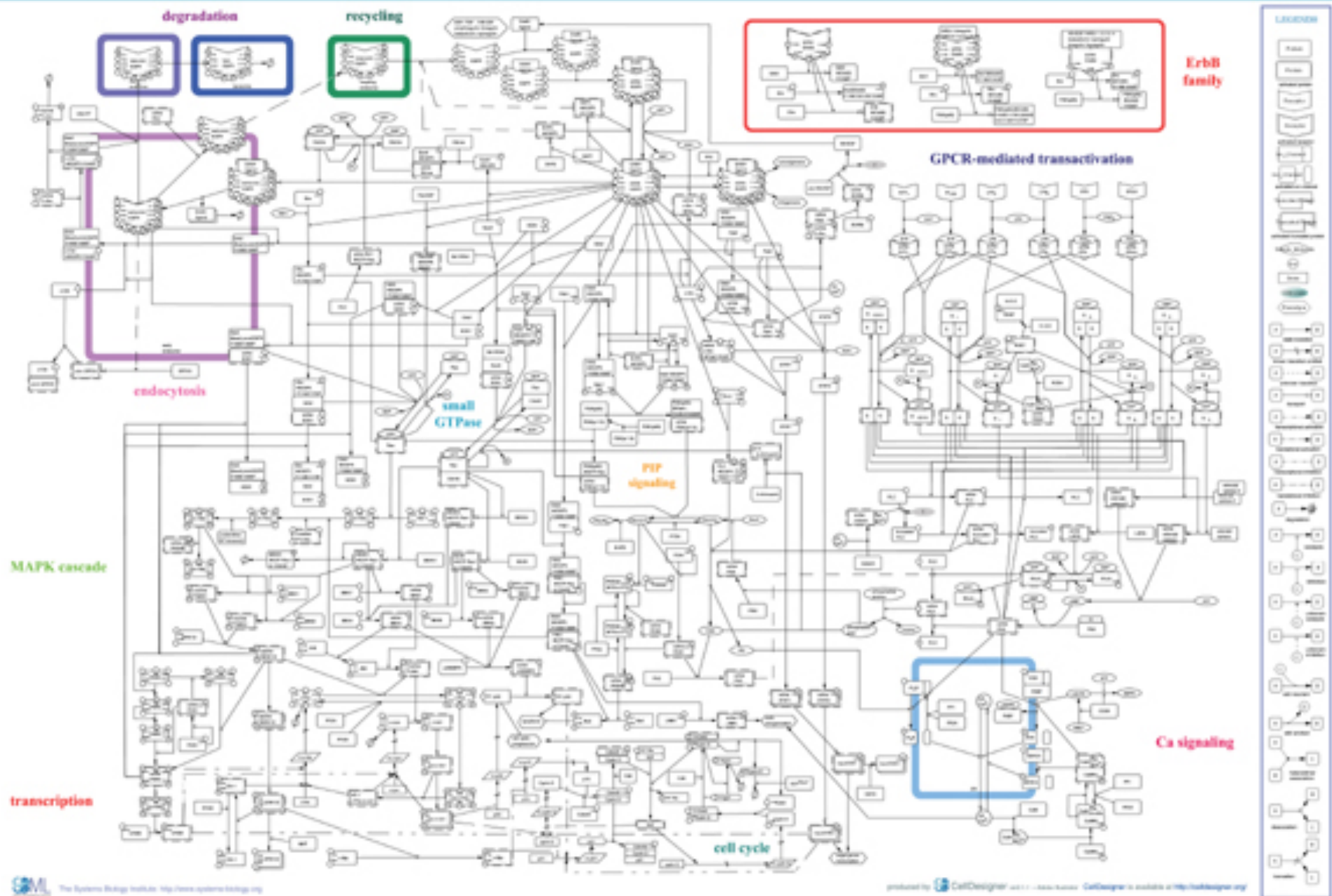


### exon structure for RNA





# EGF Receptor Cascade



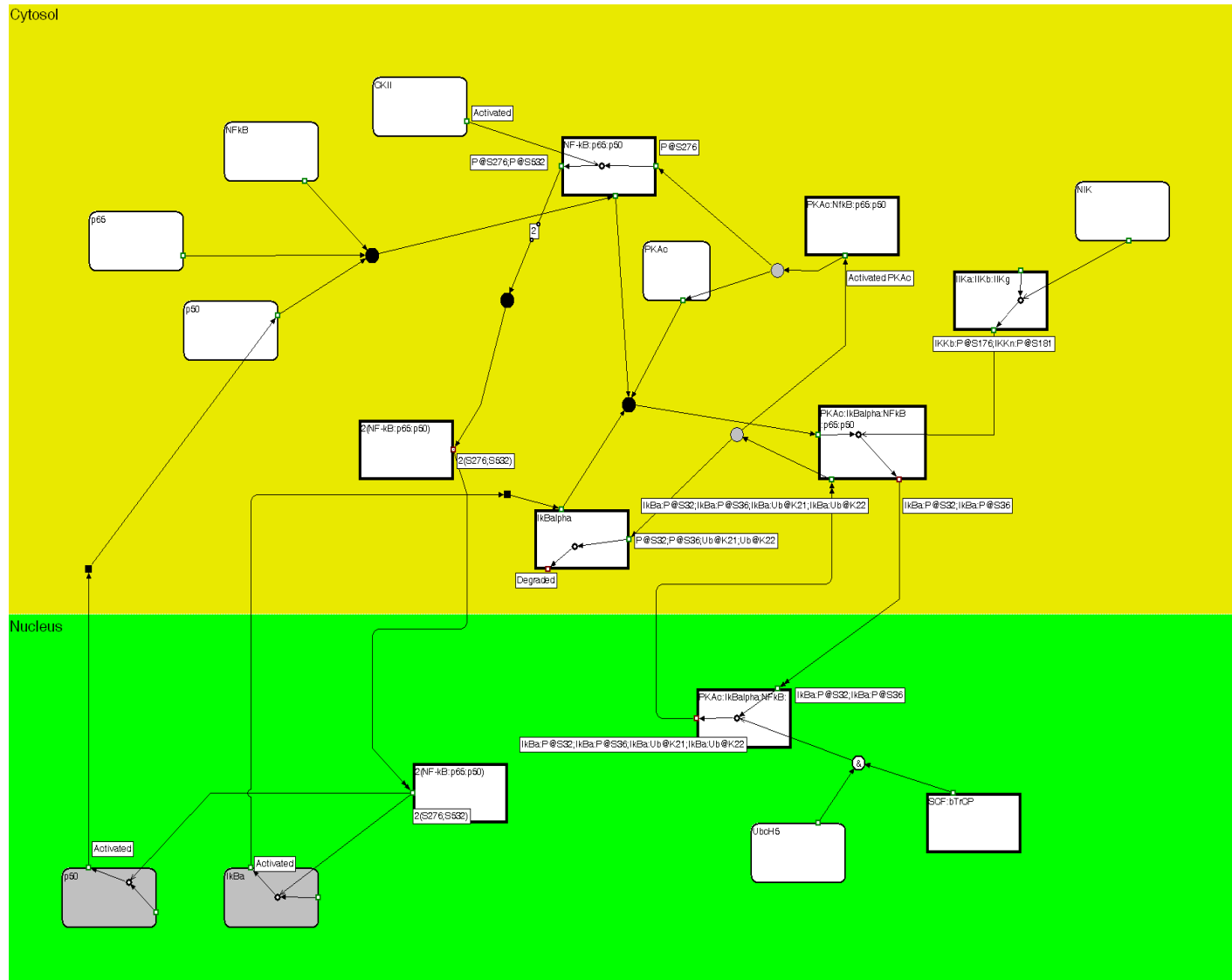
Oda, et al. Molecular Systems Biology, 2005



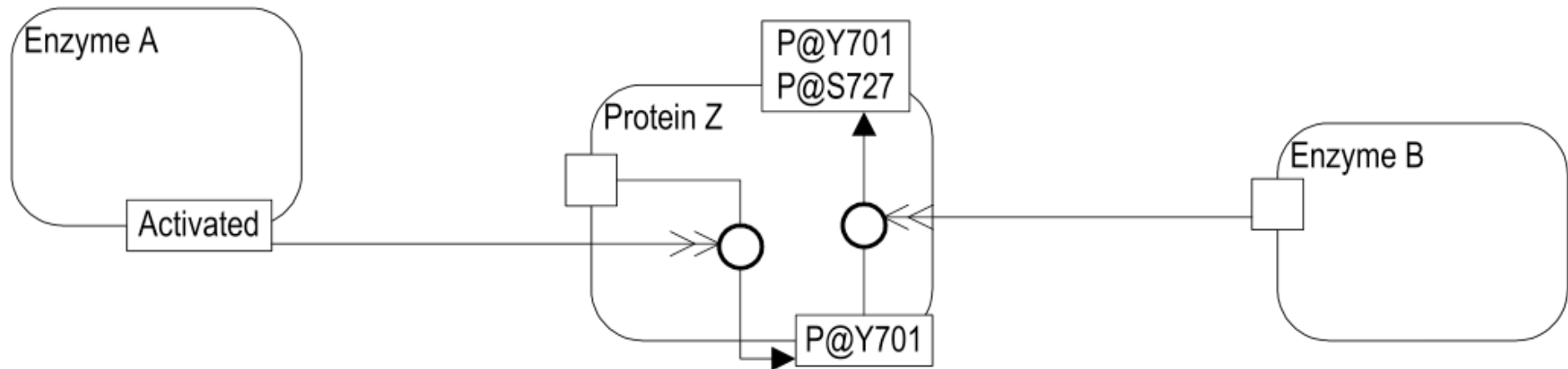
# PDN (Kitano) Notation

- Benefits
  - Simple – easy to learn
  - Maps easily to SBML
  - Established
  - Tool support (CellDesigner, EPE)
- Drawbacks
  - Verbose (takes up a lot of space)
  - Need to know pathway in detail

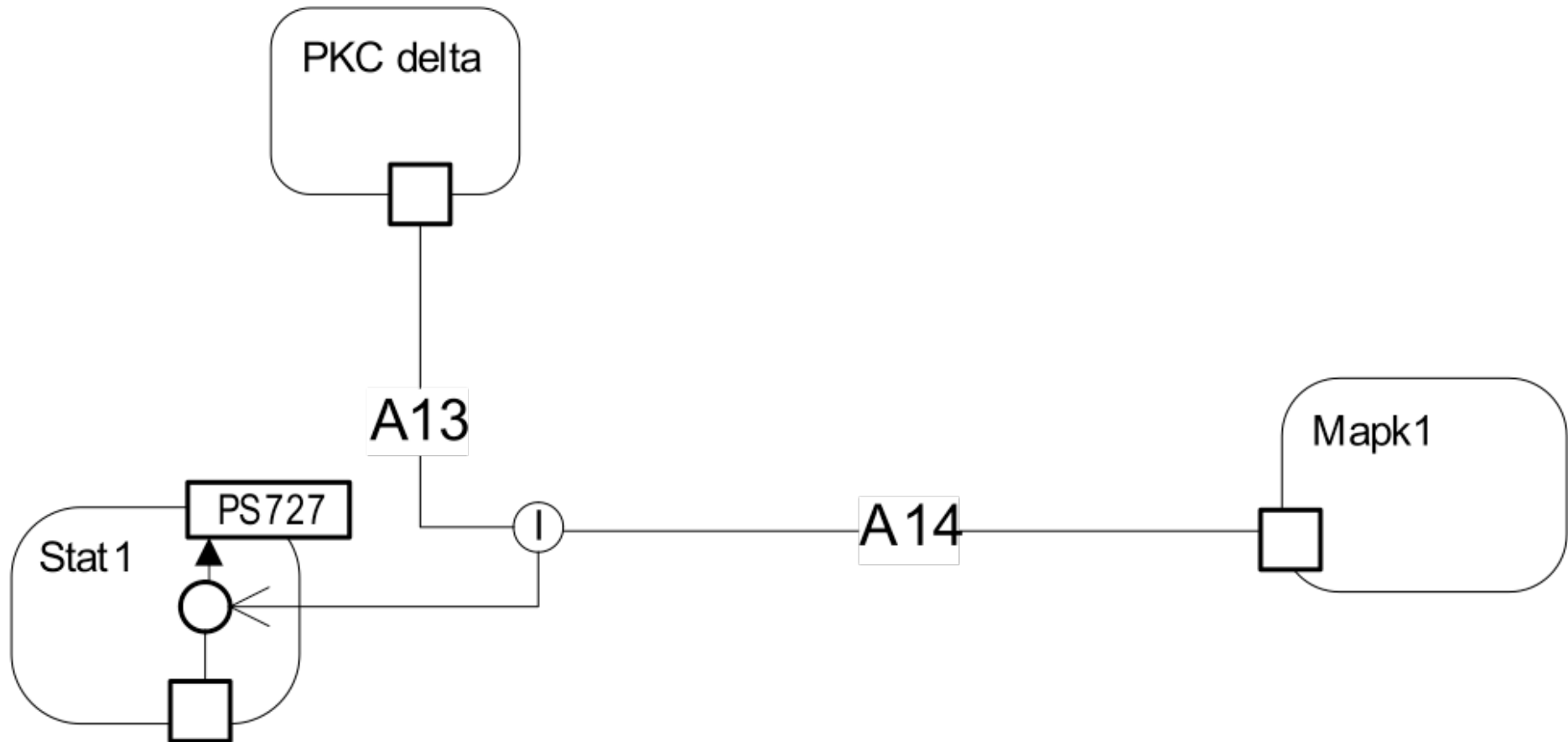
# Edinburgh Pathway Notation (EPN)



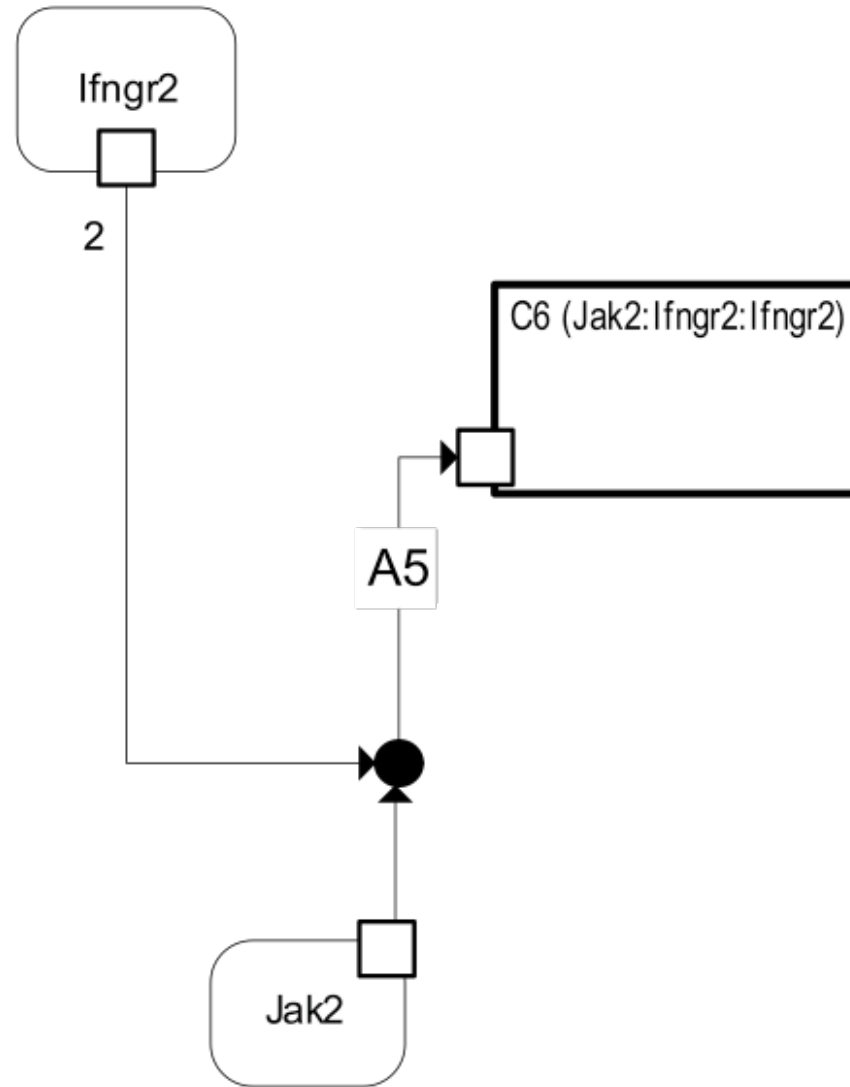
# State Transition/ Activation



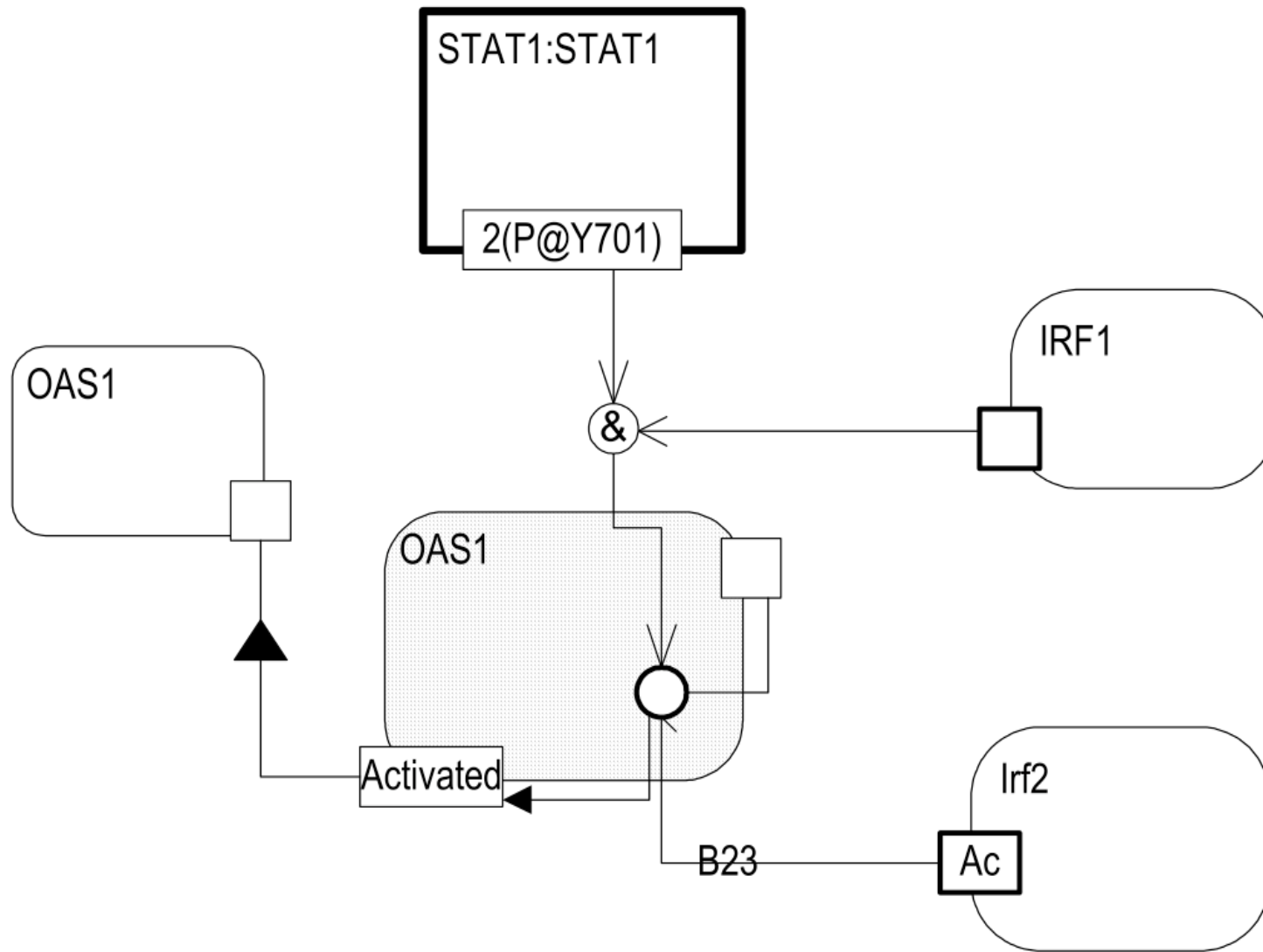
# Logic Gates



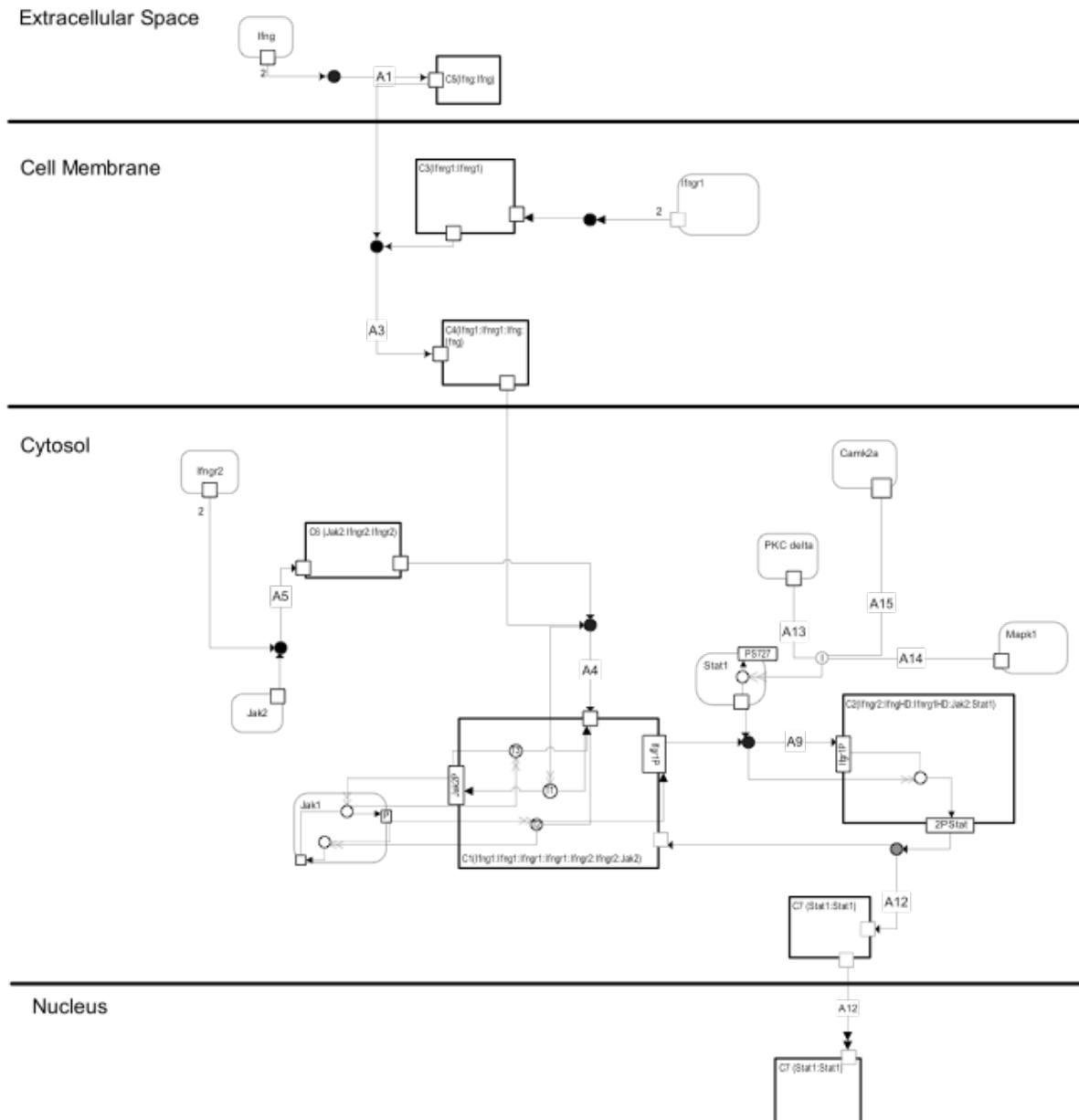
# Complex Formation



# Gene Regulation



# Localisation



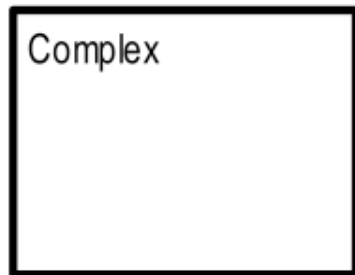
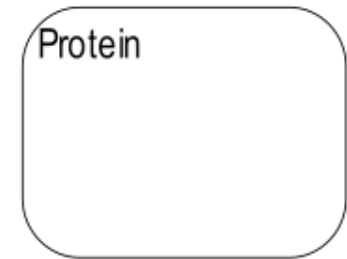
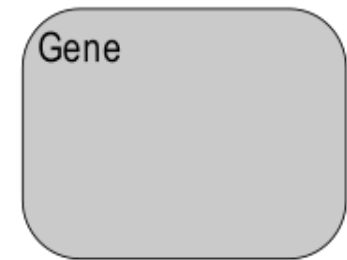
Outside



Inside



# Symbols



State

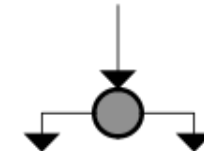
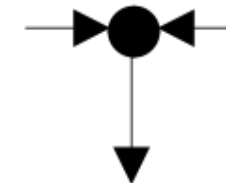
State  
Transition

OR

AND

XOR

NOT



Activation

Inhibition

Gene  
Expression

Complex  
Formation

Dissociation

Translocation

# Innovations

- Permitted arbitrary “logical states”
- Comprehensive use of logic gates
- Workable description of gene regulation
- Use of state description language: P@S727

Journal of Integrative Bioinformatics 2006

<http://journal.imbio.de/>

## **A Graphical Notation to Describe the Logical Interactions of Biological Pathways**

**Stuart L. Moodie<sup>1,2\*</sup>, Anatoly Sorokin<sup>2</sup>, Igor Goryanin<sup>2</sup>, Peter Ghazal<sup>1</sup>**

<sup>1</sup>Scottish Centre for Genomic Technology and Informatics, University of Edinburgh Medical School, Chancellor's Building, Little France Crescent, Edinburgh EH16 4SB, UK

<sup>2</sup>Computational Systems Biology Group, University of Edinburgh, Appleton Tower, Crichton Street, Edinburgh EH8 9LE, UK



[www.sbgn.org](http://www.sbgn.org)

**SYSTEMS BIOLOGY GRAPHICAL  
NOTATION**

# SBGN ([www.sbgn.org](http://www.sbgn.org))

- Aims
  - Agree a standard set of graphical notations for systems biology
  - Encourage tool support for SBGN
  - Encourage community usage
- Initiated by:
  - Kitano, SBI
  - Le Novère, EBI

# SBGN Governance

- 5 Editors who write the specs and coordinate
  - Nicolas Le Novere (EBI)
  - Huaiyu Mi (SRI)
  - Stuart Moodie (UofE)
  - Falk Schreiber (MLU Halle-Wittenberg)
  - Emek Demir (MSKCC)
- 5 Member Scientific Committee

Annual Hackathon (Spring 2010 in Germany)

Annual Forum Meeting (Edinburgh Oct 2010)

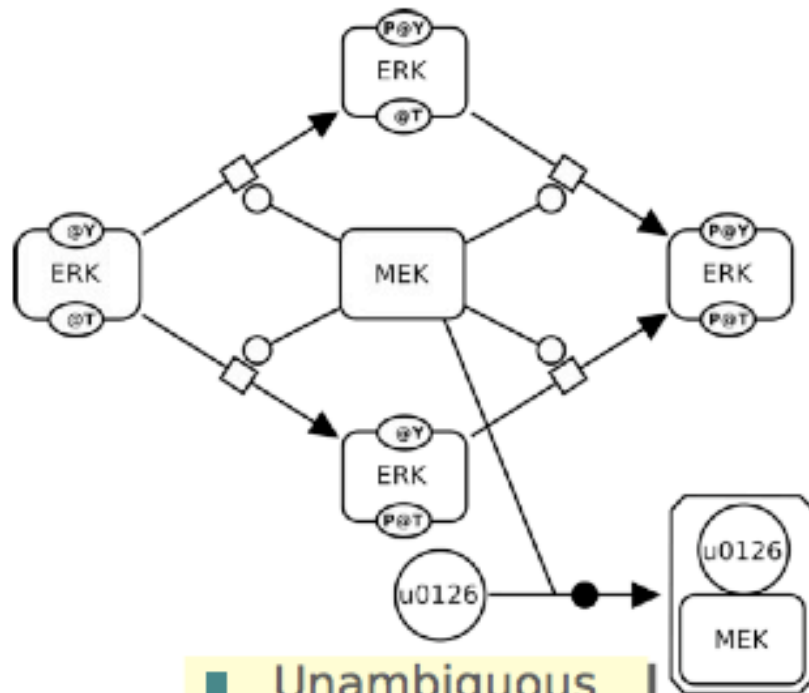
# A Systems Biology Graphical Notation

## Desirable Properties

1. Unambiguous
2. Parsimonious (NL)
3. Computationally tractable
4. Can be hand drawn
5. Follows accepted conventions where-ever possible
6. Can be printed and reproduced in black and white
7. Allows for incomplete information

# SBGN 3 Languages

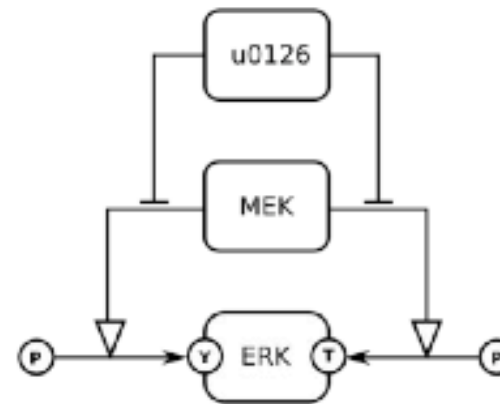
Process diagrams



- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships

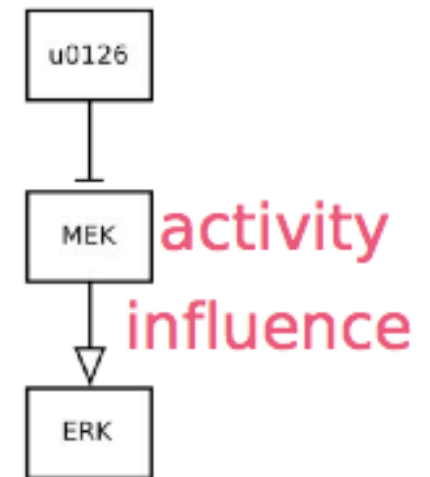
diagrams



- Unambiguous
- Mechanistic
- Non-sequential

Activity Flow

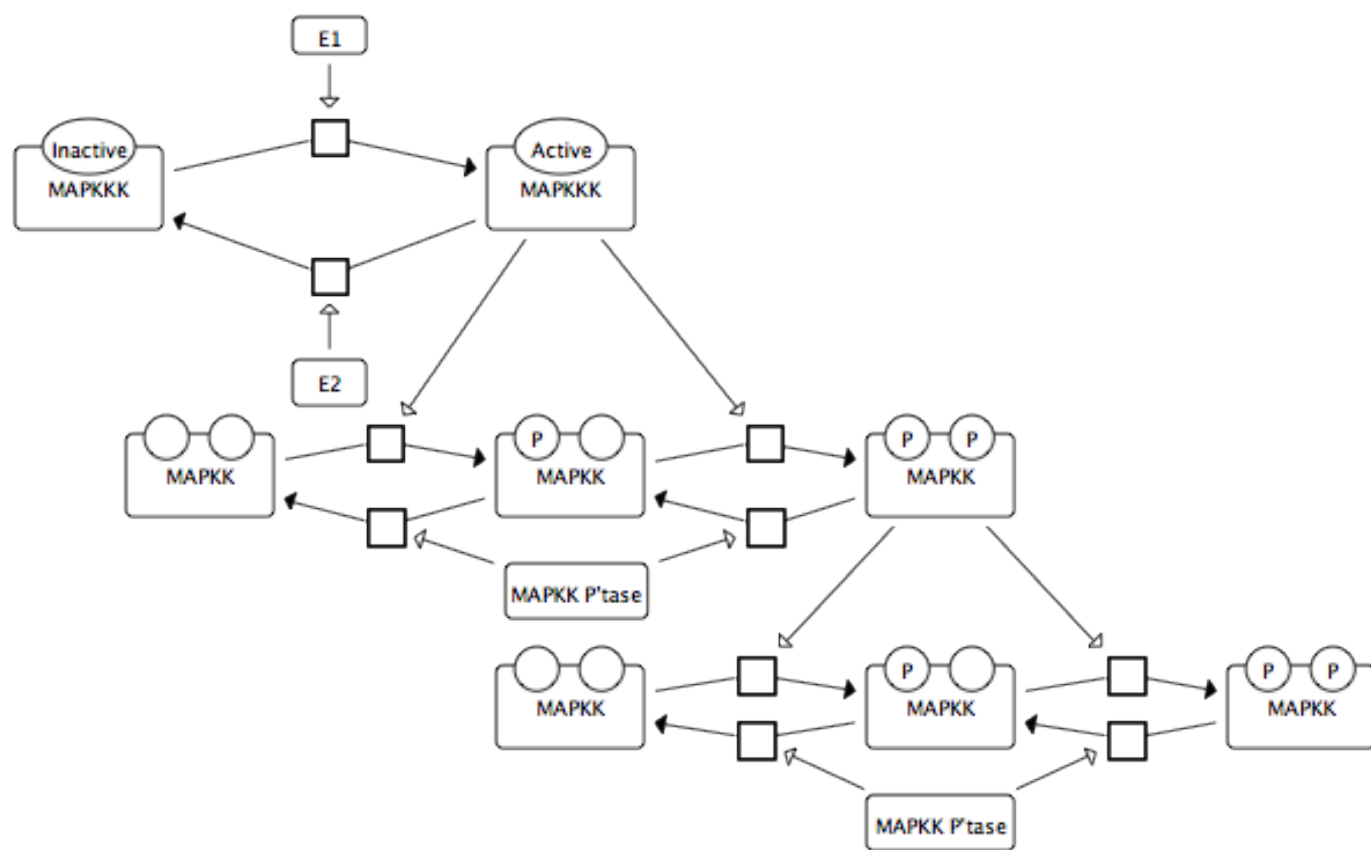
diagrams



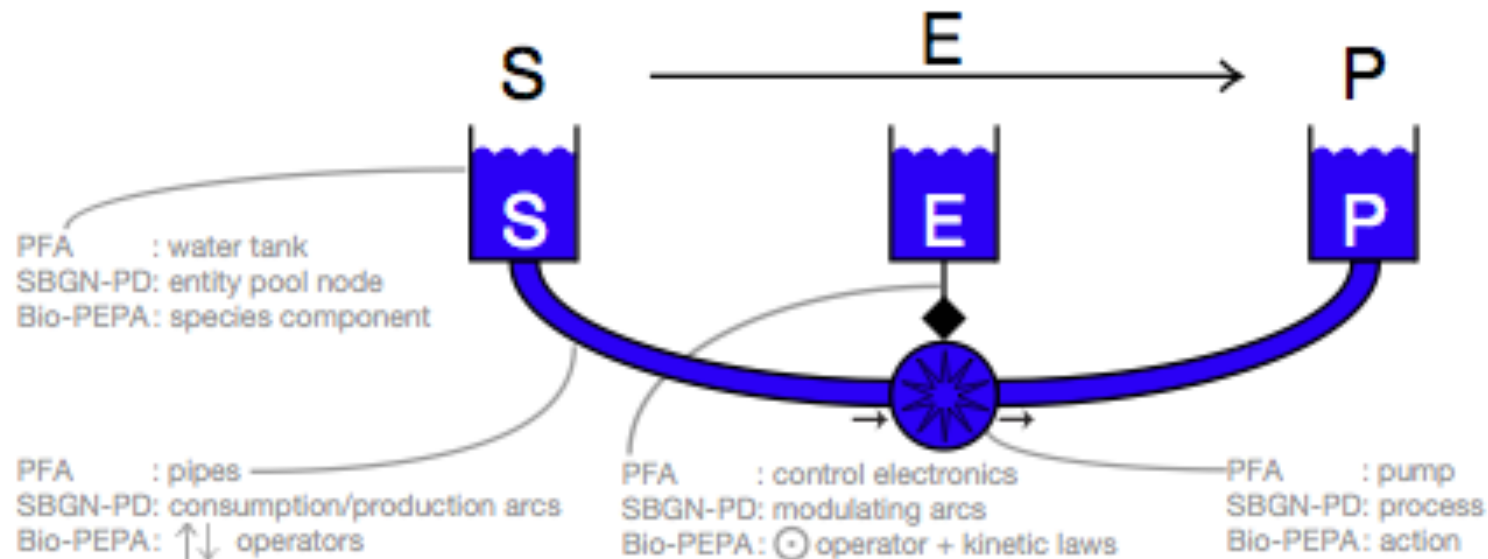
- Ambiguous
- Conceptual
- Sequential

# SBGN: Process Description

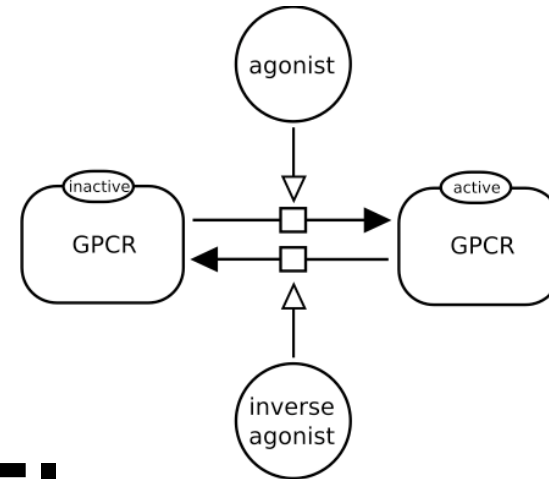
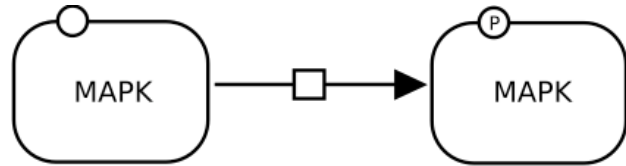




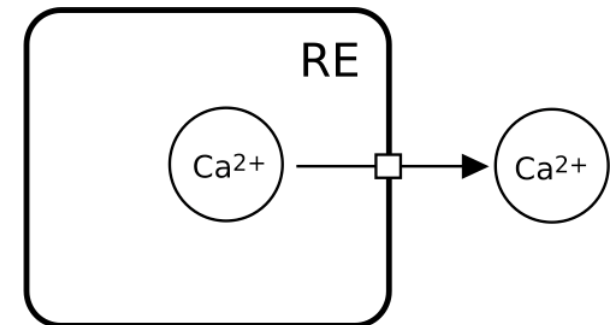
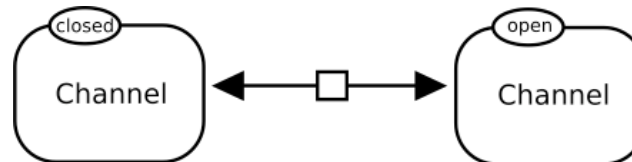
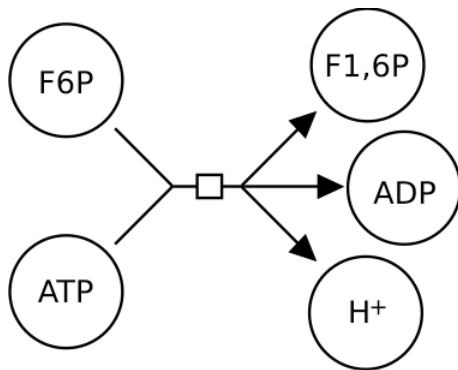
# Process-flow abstraction

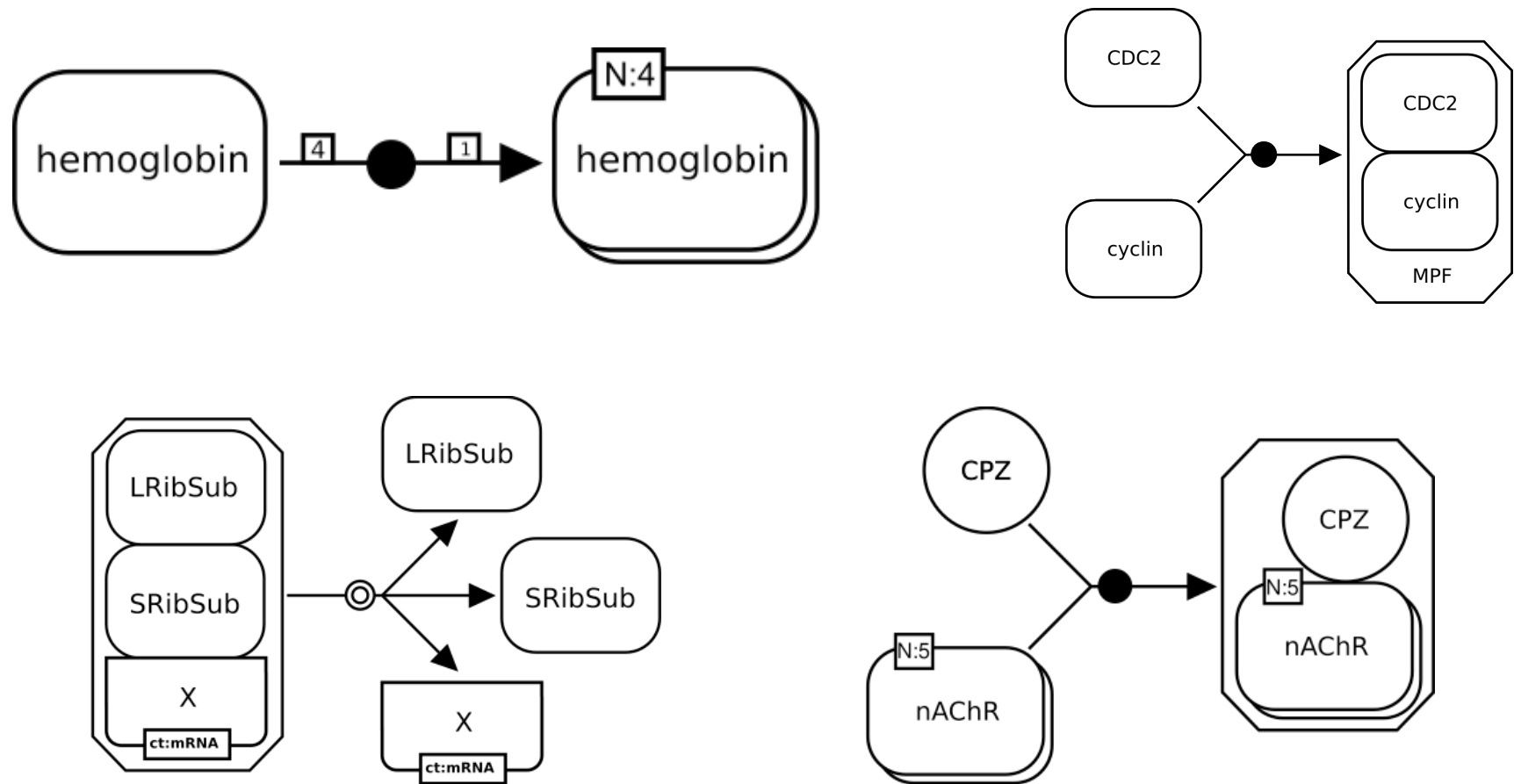


from Loewe, Moodie & Hilston CompMod 2009.

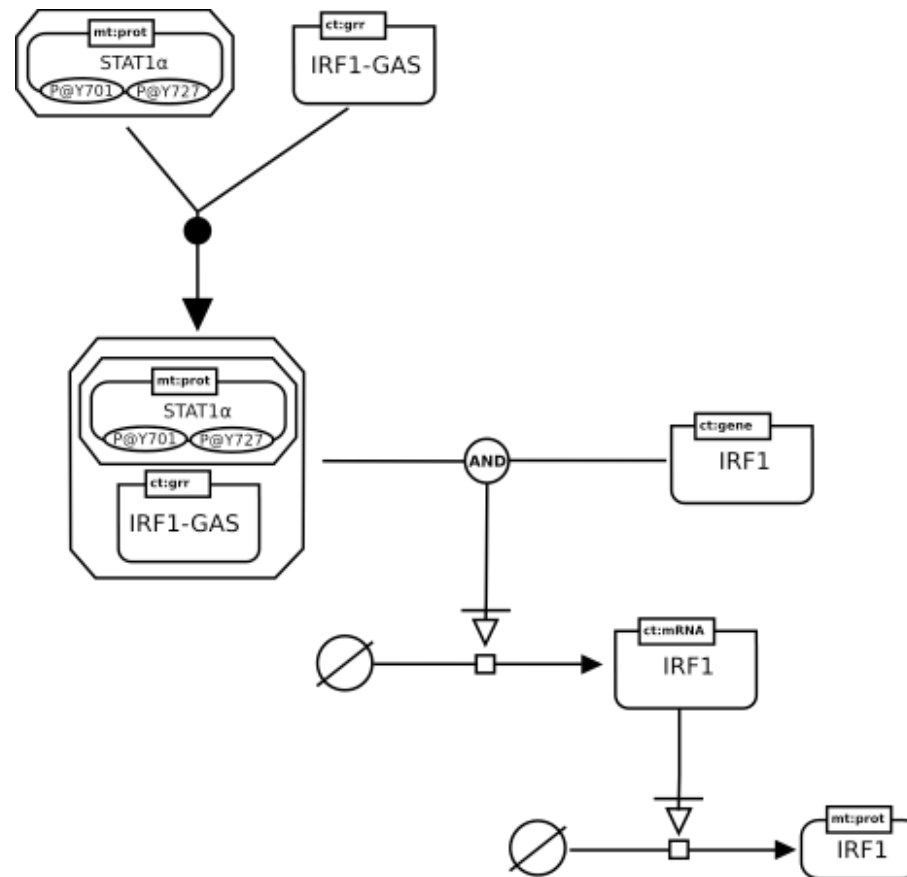


# Process Flow

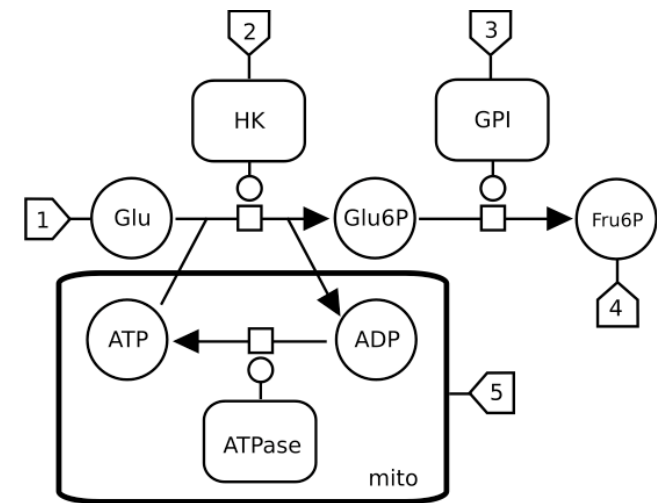
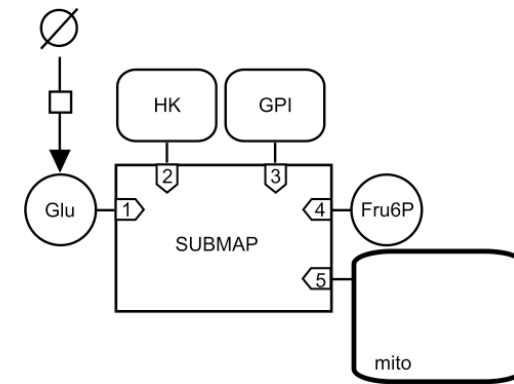
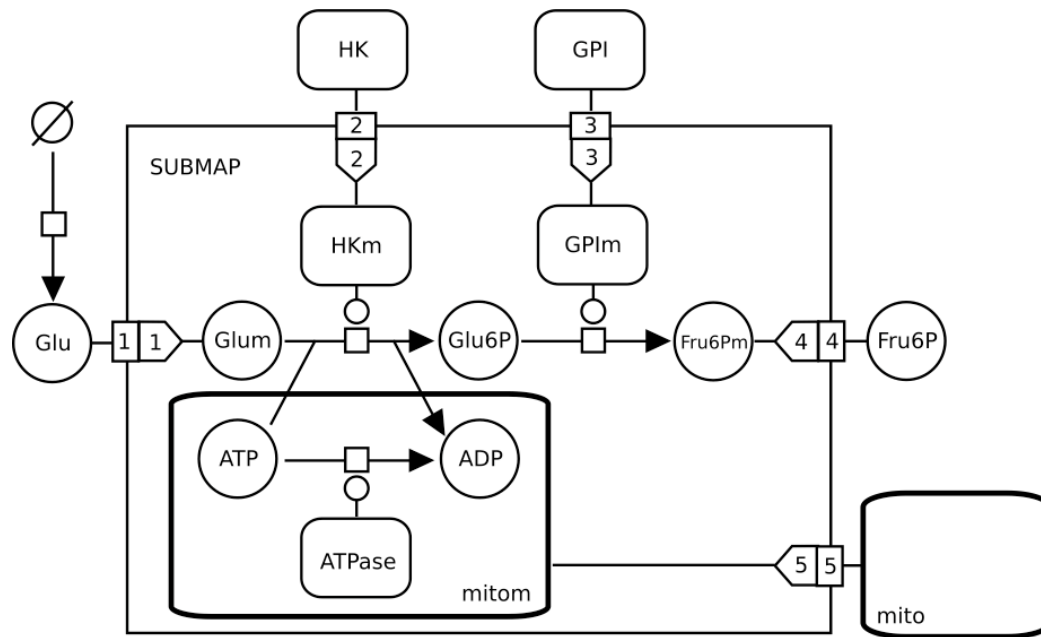




# Complex Formation

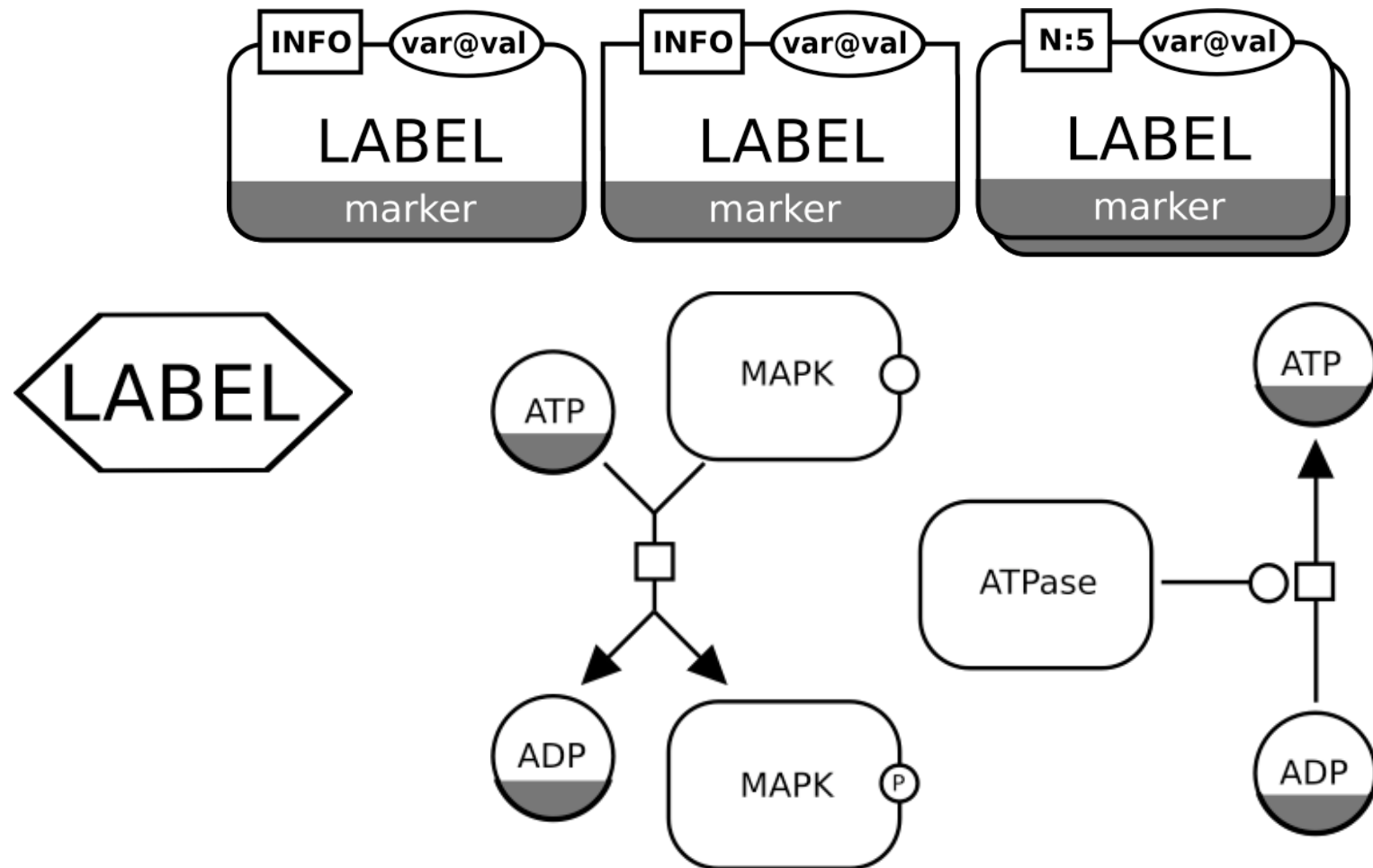


Summarising information: Logic gates



Reducing complexity: Sub-Maps

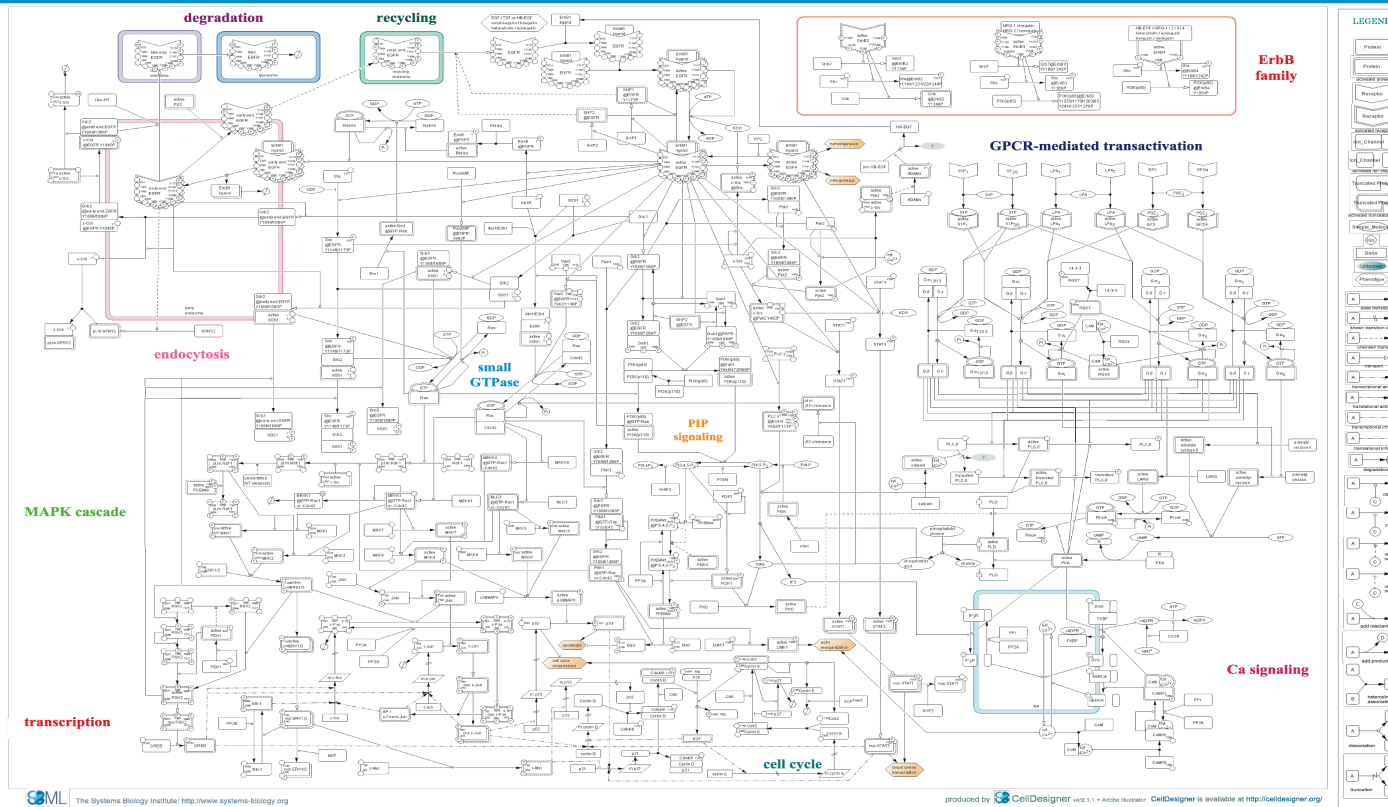
# Redundant EPNS: Cloning



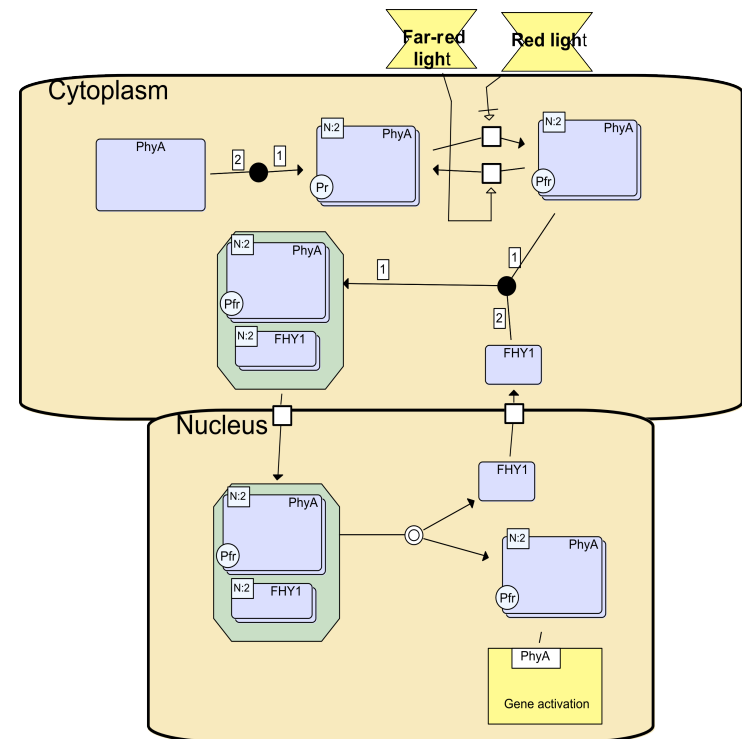
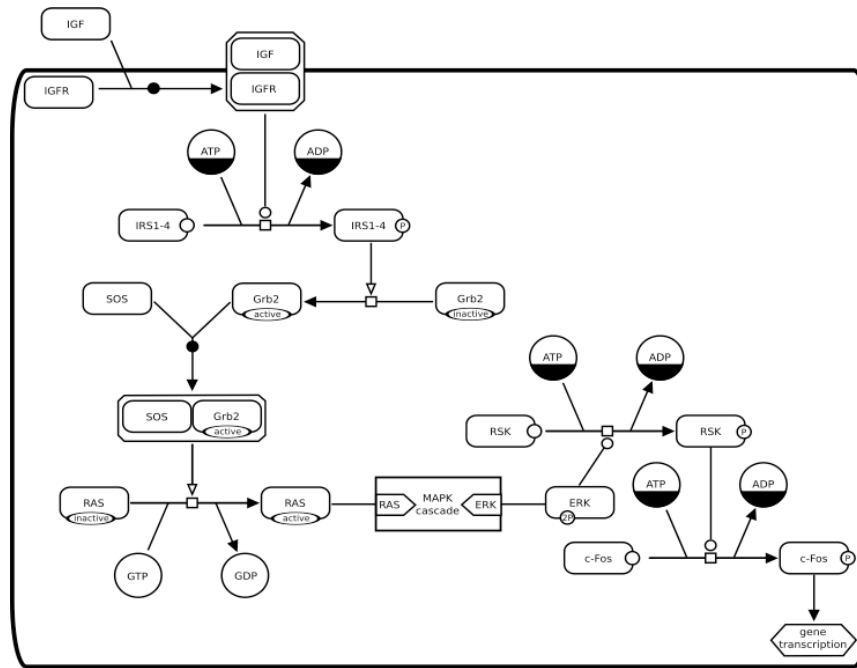
# Why Cloning is Necessary

## Epidermal Growth Factor Receptor Pathway Map

Kanoe Oda (1,2), Yukiko Matsuoka (3), Hiroaki Kitano (1,2,4)

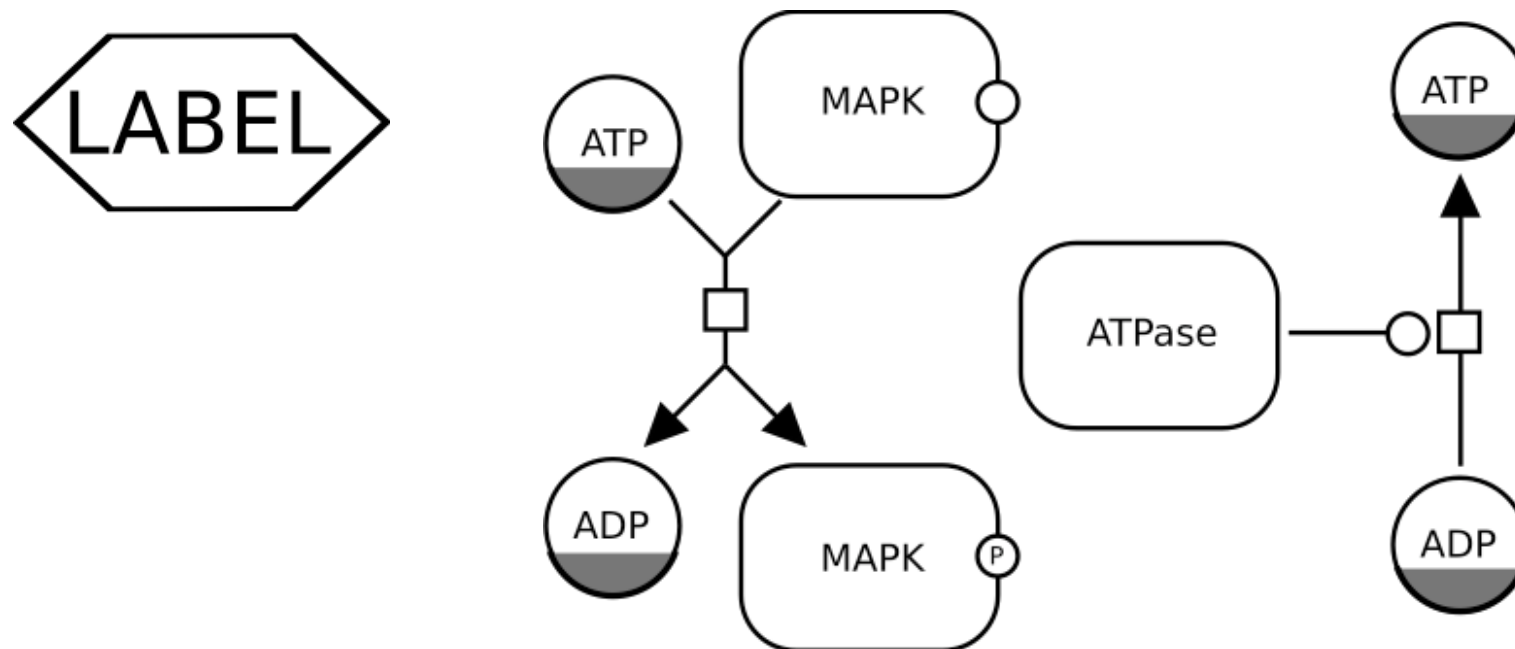
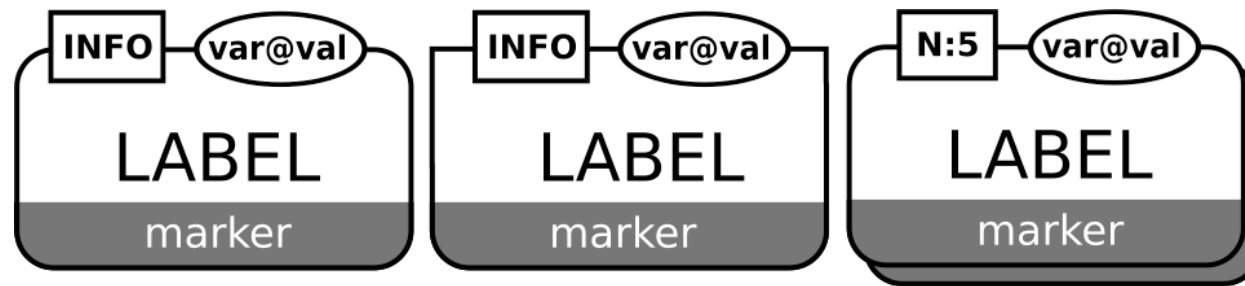


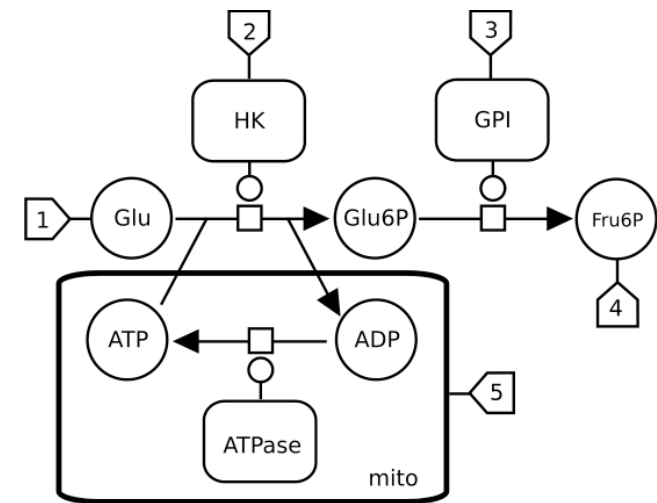
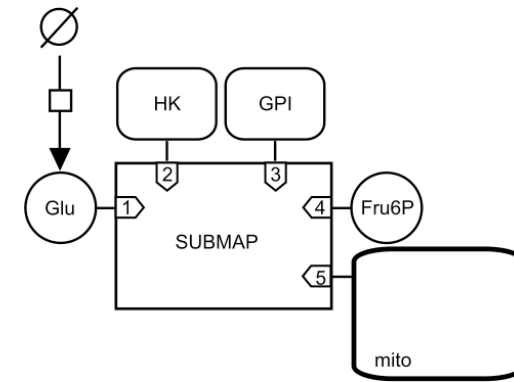
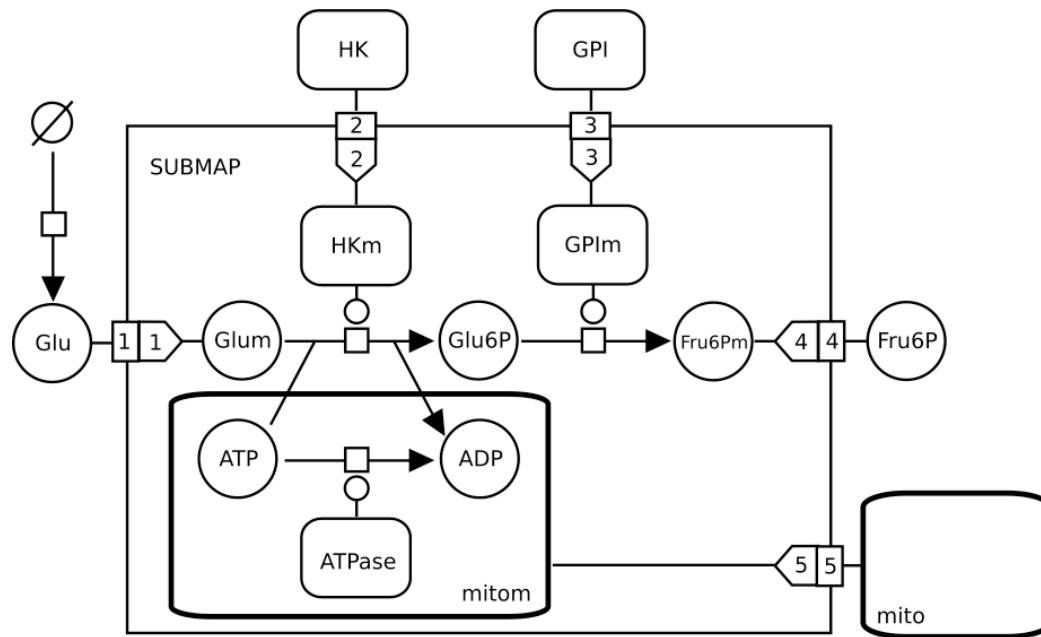




Relating to the “Outside World”:  
Perturbation/Observable

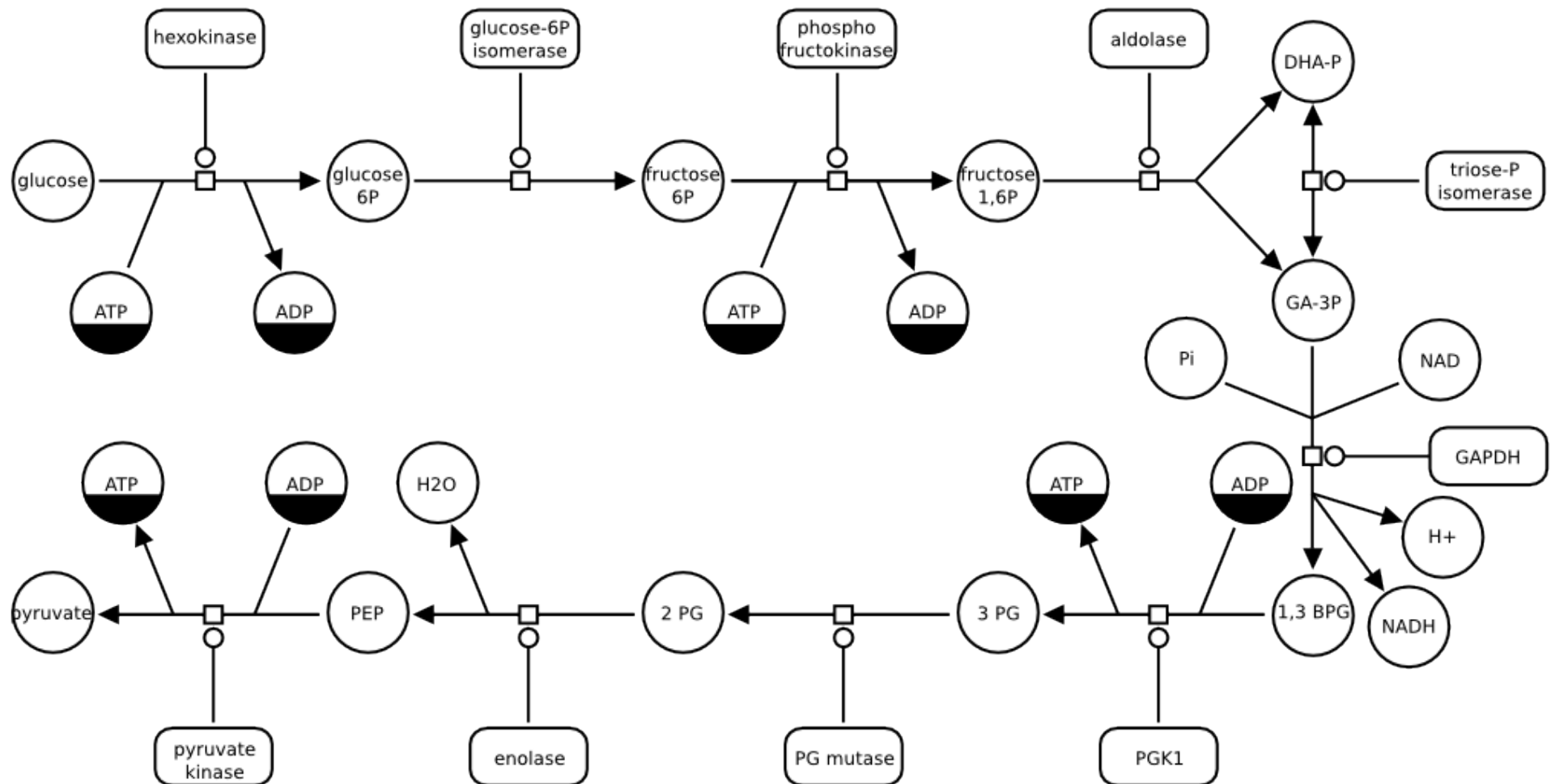
# Redundant EPNS: Cloning



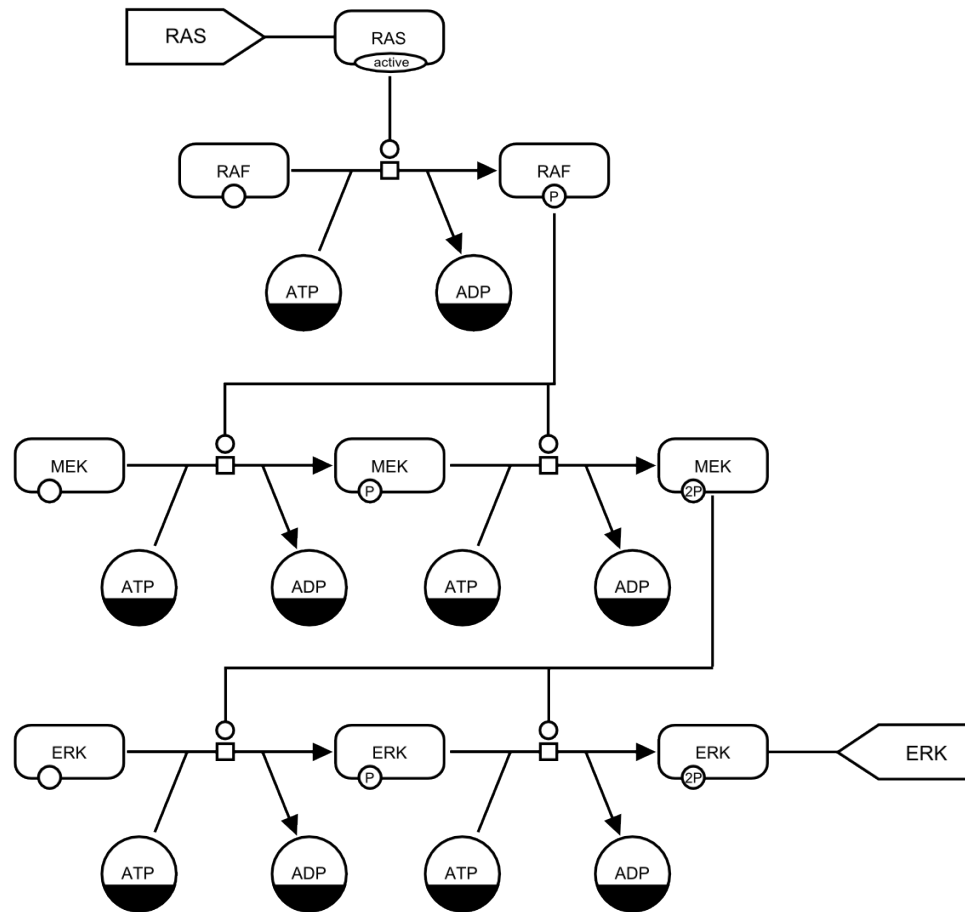


# Reducing complexity: Sub-Maps

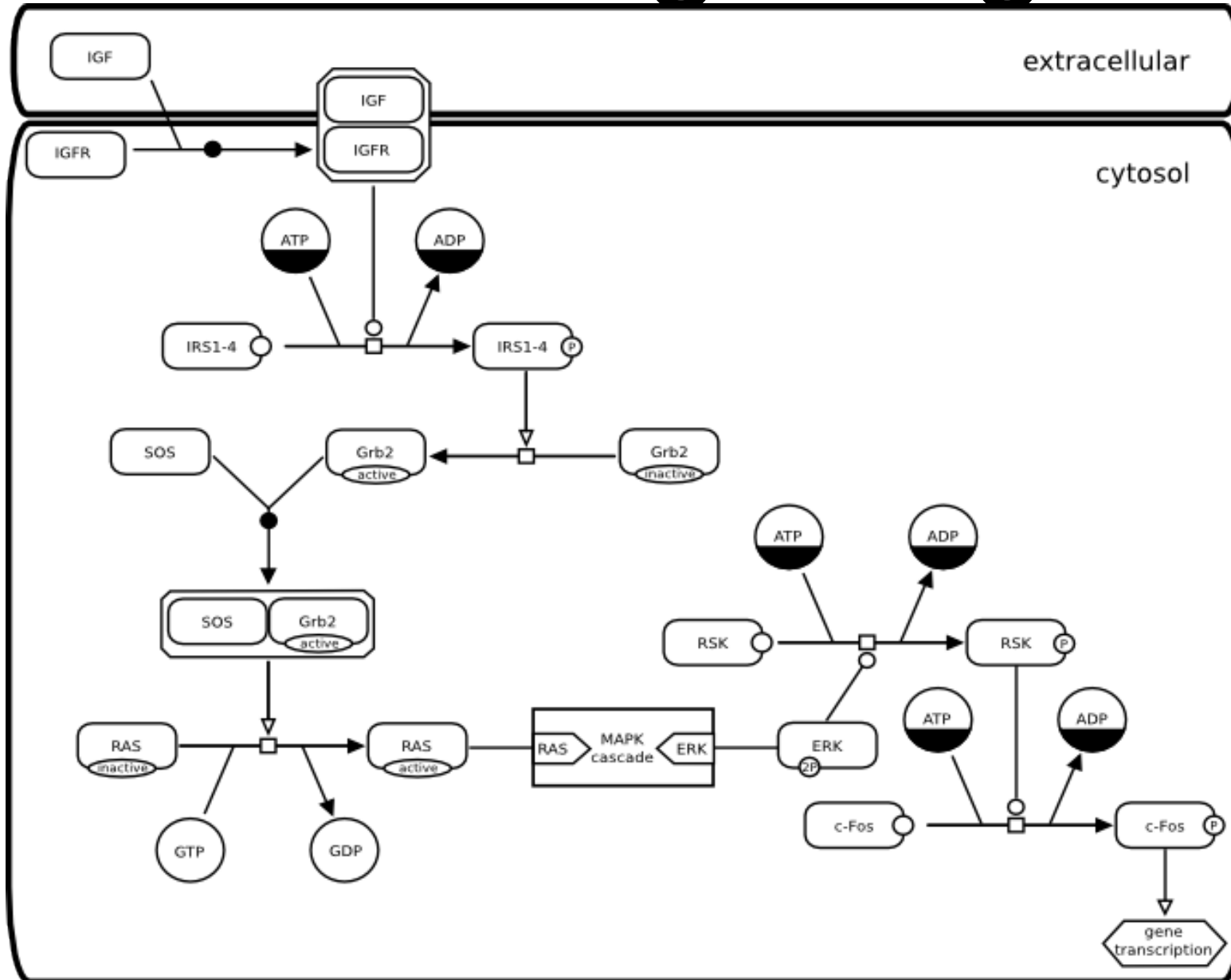
# Process Diagram Examples



# MAPK

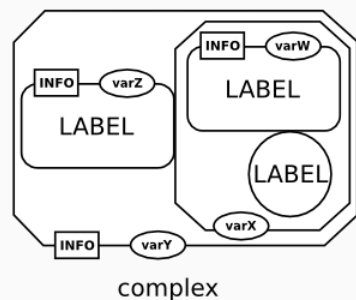
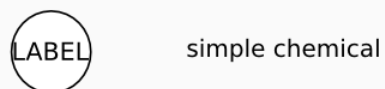
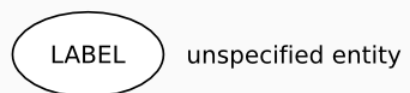


# Insulin Signalling

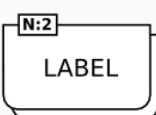
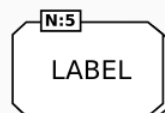
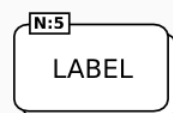
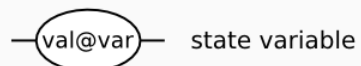
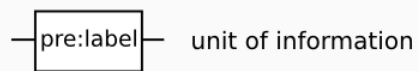


# SYSTEMS BIOLOGY GRAPHICAL NOTATION REFERENCE CARD (PD Level 1 R1.1)

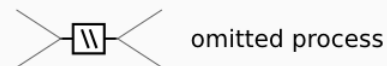
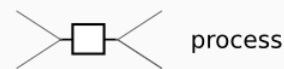
## Entity Pool Nodes



## Auxiliary Units



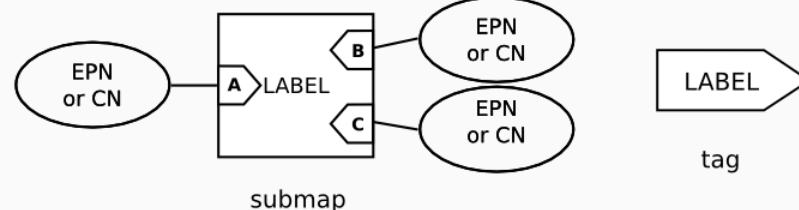
## Process Nodes



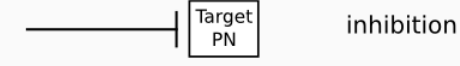
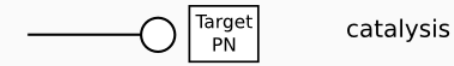
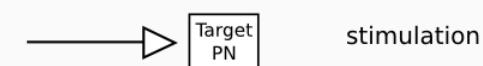
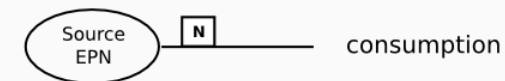
## Compartment Nodes



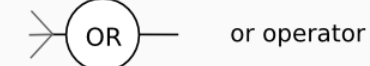
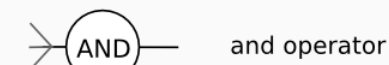
## Reference Nodes



## Arcs



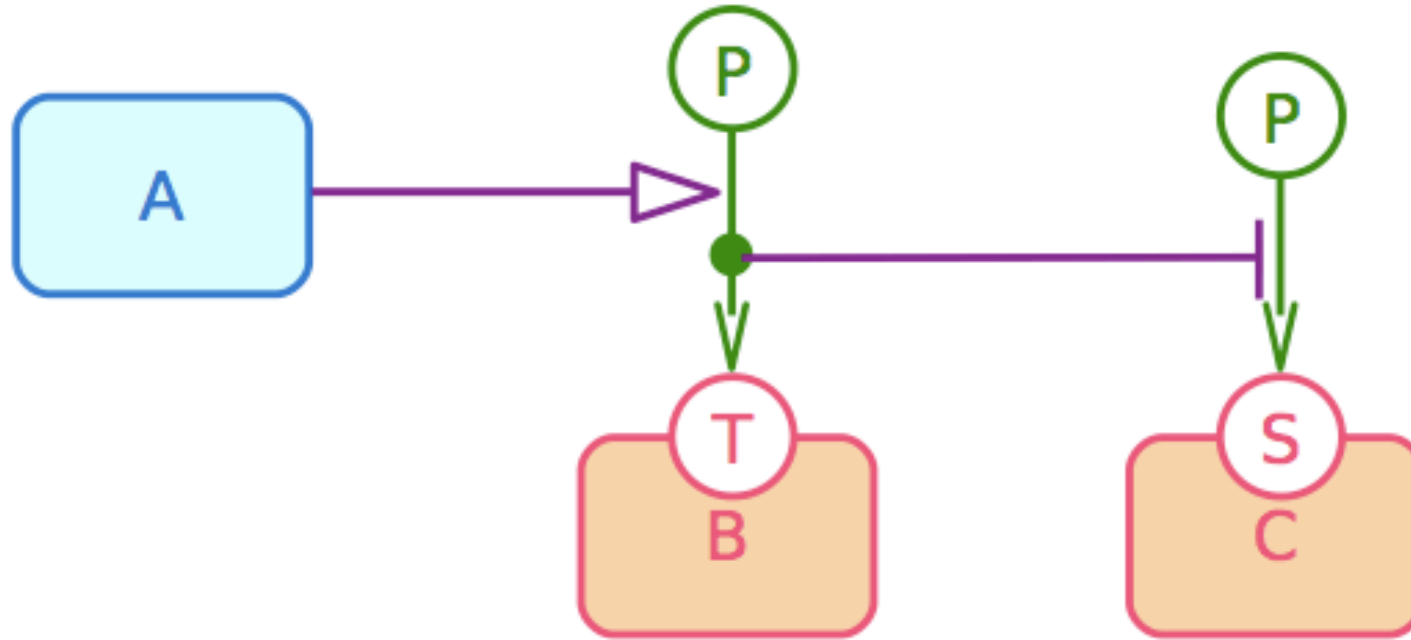
## Logical Operators



# SBGN-Entity Relationship Diagram



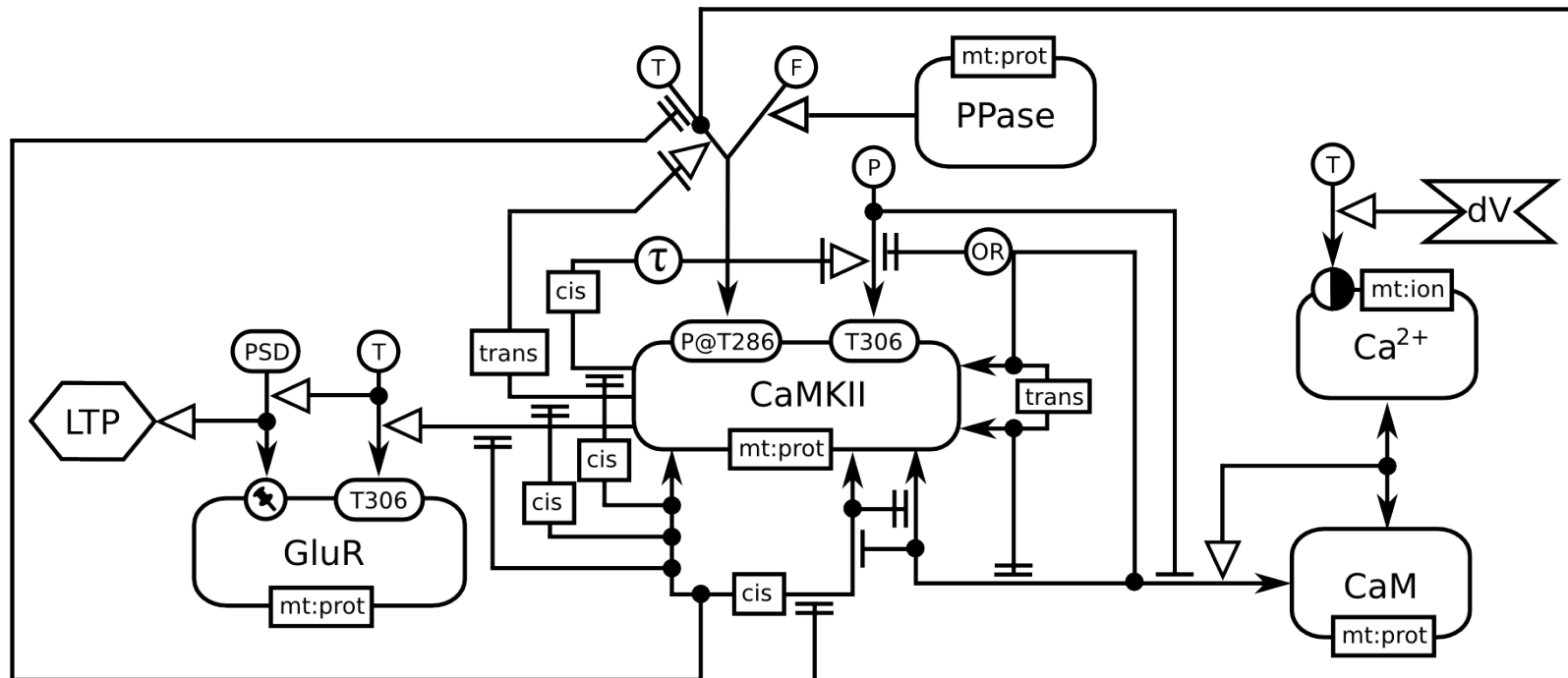
# A Simple Example



If **A** exists, the assignment of the value **P** to the state variable **T** of **B** is increased

**A** stimulates the phosphorylation of **B** on the threonine

# ER Example

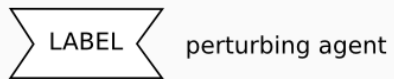
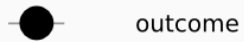


# SYSTEMS BIOLOGY GRAPHICAL NOTATION ENTITY RELATIONSHIP REFERENCE CARD

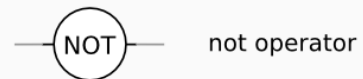
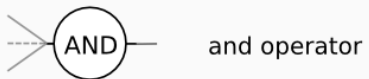
## Entity Nodes

## Relationship Nodes

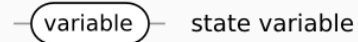
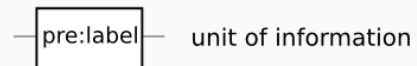
### Interactors



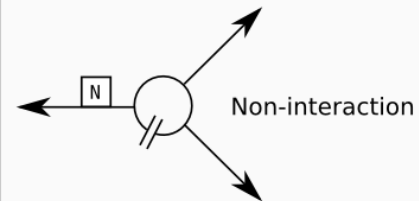
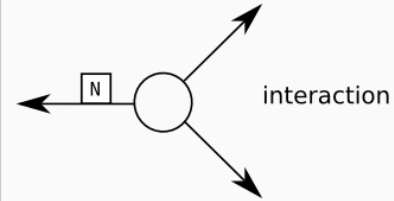
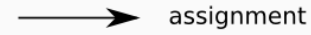
### Logical Operators



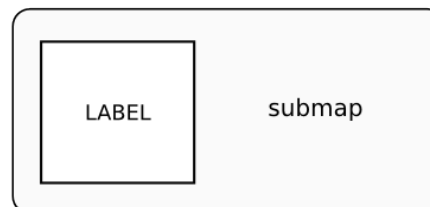
### Auxiliary units



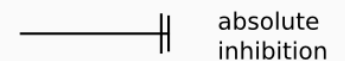
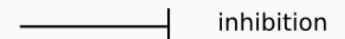
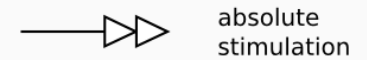
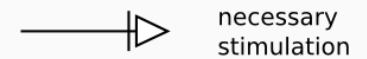
### Statements



## Containers



### Influence



## A Mutant-p53/Smad Complex Opposes p63 to Empower TGF $\beta$ -Induced Metastasis

Maddalena Adorno,<sup>1,8</sup> Michelangelo Cordenonsi,<sup>1,8</sup> Marco Montagner,<sup>1</sup> Sirio Dupont,<sup>1</sup> Christine Wong,<sup>2</sup> Byron Hann,<sup>2</sup> Aldo Solari,<sup>3</sup> Sara Bobisse,<sup>4</sup> Maria Beatrice Rondina,<sup>4</sup> Vincenza Guzzardo,<sup>5</sup> Anna R. Parenti,<sup>5</sup> Antonio Rosato,<sup>4,6</sup> Silvio Biciardi,<sup>6,7</sup> Allan Balmain,<sup>2</sup> and Stefano Piccolo<sup>1,\*</sup>

<sup>1</sup>Department of Histology, Microbiology and Medical Biotechnologies, University of Padua School of Medicine, viale Colombo 3, 35100 Padua, Italy

<sup>2</sup>Cancer Research Institute, University of California, San Francisco, 2340 Sutter Street, San Francisco, CA 94115, USA

<sup>3</sup>Department of Chemical Engineering Processes, University of Padua, via F. Marzolo 9, 35131, Padua, Italy

<sup>4</sup>Department of Oncology and Surgical Sciences and Istituto Oncologico Veneto, University of Padua, via Gattamelata 64, 35126 Padua, Italy

<sup>5</sup>Department of Medical Diagnostic Science and Special Therapies, Section of Pathology, University of Padua, viale Gabelli 2, 35126 Padua, Italy

<sup>6</sup>Istituto Oncologico Veneto, via Gattamelata 64, 35126 Padua, Italy

<sup>7</sup>Department of Biomedical Sciences, University of Modena and Reggio Emilia, via G. Campi 287, 41100, Modena, Italy

<sup>8</sup>These authors contributed equally to the work

\*Correspondence: piccolo@icv.bio.unipd.it

DOI: 10.1016/j.cell.2009.01.039

### SUMMARY

TGF $\beta$  ligands act as tumor suppressors in early stage

tumors but are paradoxically diverted into potent

pro-metastatic factors in advanced cancers. The

One of the most frequent genetic lesions in human tumors is mutation of the p53 tumor suppressor, which acts as transcription factor to promote cytoskeleton, apoptosis and genome integrity. More than 50% of p53 alterations are missense mutations that lead to the synthesis of a stable but transcriptionally

inactive protein. In this study, we show that mutant p53 (mut-p53) and p63, a p53 family member, form a complex that

opposes p63 to empower TGF $\beta$ -induced metastasis. We

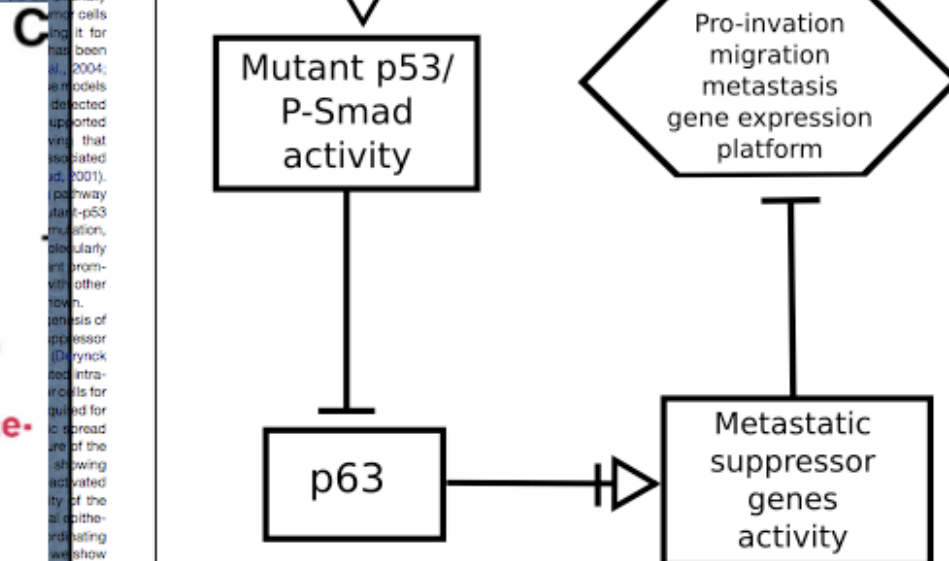
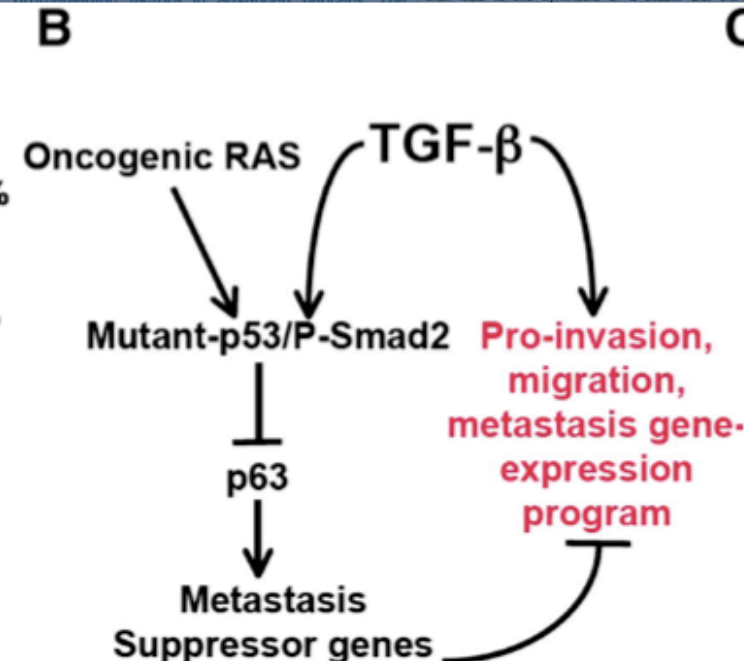
show that the mutant p53/p63 complex acts as a transcriptional repressor of the p63 target genes, including

pro-metastatic genes. Our results suggest that the mutant p53/p63 complex is a key player in the

diversion of TGF $\beta$  signaling from tumor suppression to metastasis promotion. This mechanism may

represent a common pathway for the progression of many human tumors.

**Key Words:** TGF $\beta$ ; p53; p63; metastasis; tumor suppressor; transcription factor



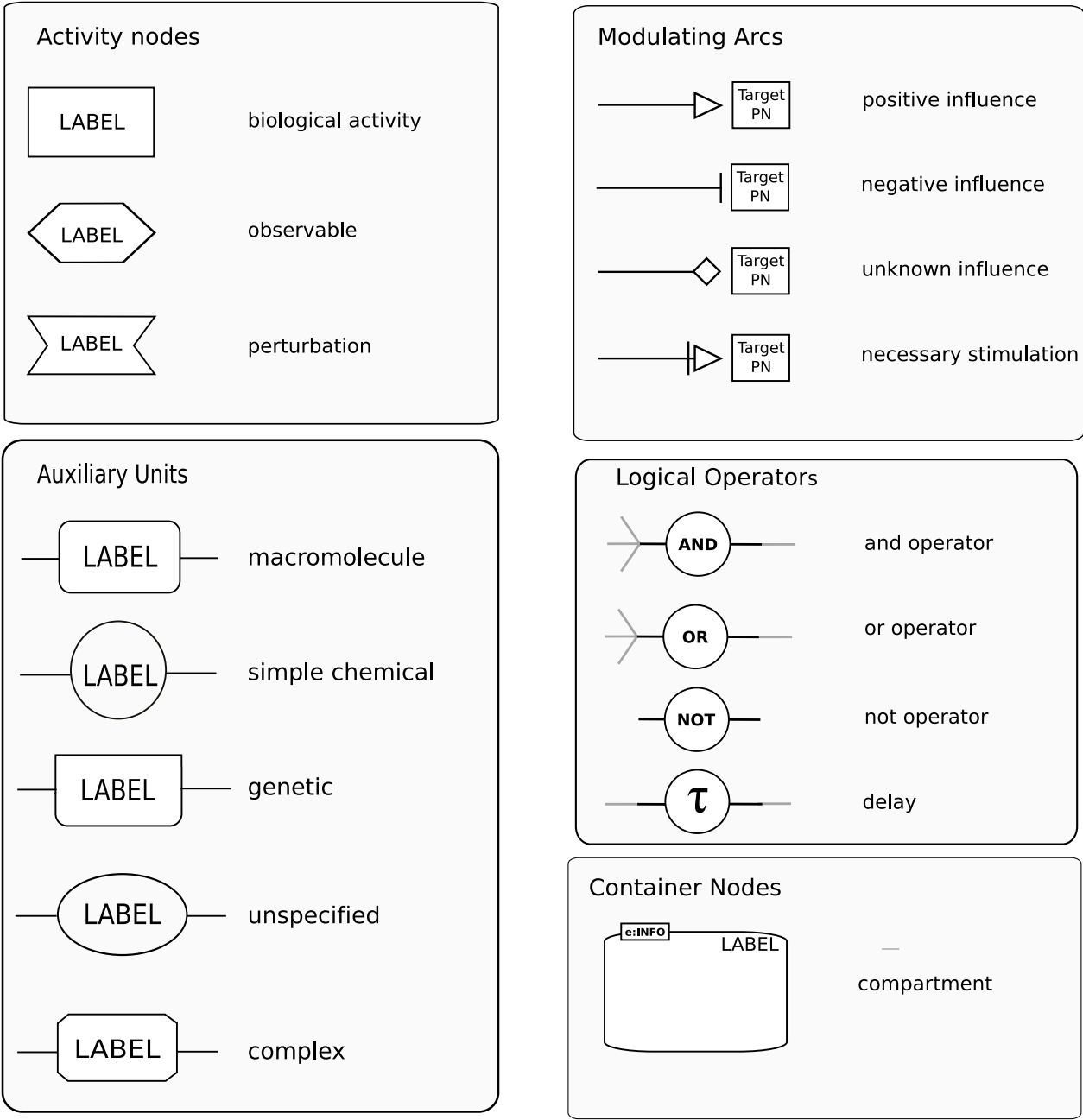
# Activity Node (AN)

## -Biological activity



- Each node represents an activity, but not the entity.
- Multiple ANs can be used to represent activities from one entity, e.g., receptor protein kinase, and ligand gated ion channel.
- One AN can be used to represent activities from a group of entities (e.g., a complex).

SYSTEMS BIOLOGY GRAPHICAL NOTATION  
ACTIVITY FLOW DIAGRAM REFERENCE CARD



# Current Status

- 5 workshops since Feb 2006
- Specification for Process Notation Level I:
  - Released Level I in August at ICSB 2008
- ER Level I Specification
  - Sep 2009
- AF Level I Specification
  - Sep 2009

## References

General overview of the standard:

Nat Biotechnol. 2009 27(8):735-41 (<http://www.nature.com/nbt/journal/v27/n8/abs/nbt.1558.html>)

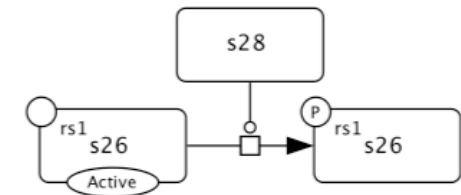
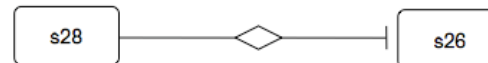
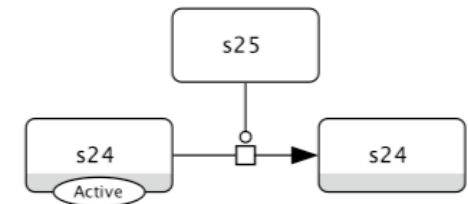
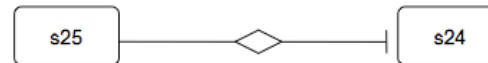
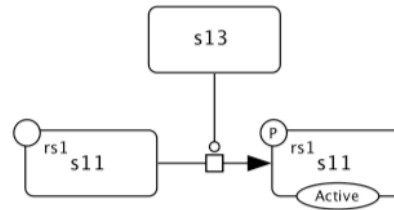
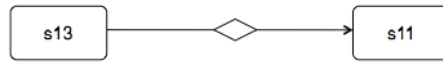
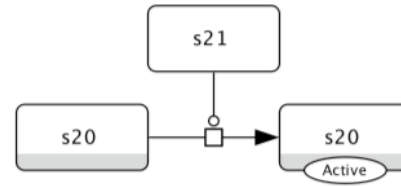
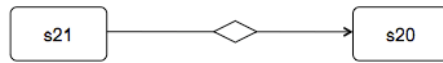
Specification documents:

- PD: doi:10.1038/npre.2009.3721.1 (<http://precedings.nature.com/documents/3721/version/1>)
- ER: doi:10.1038/npre.2009.3719.1 (<http://precedings.nature.com/documents/3719/version/1>)
- AF: doi:10.1038/npre.2009.3724.1 (<http://precedings.nature.com/documents/3724/version/1>)

# Activity Flow

- Describes changes in “activity”
- Good providing overview
- “Lossy”





# Example