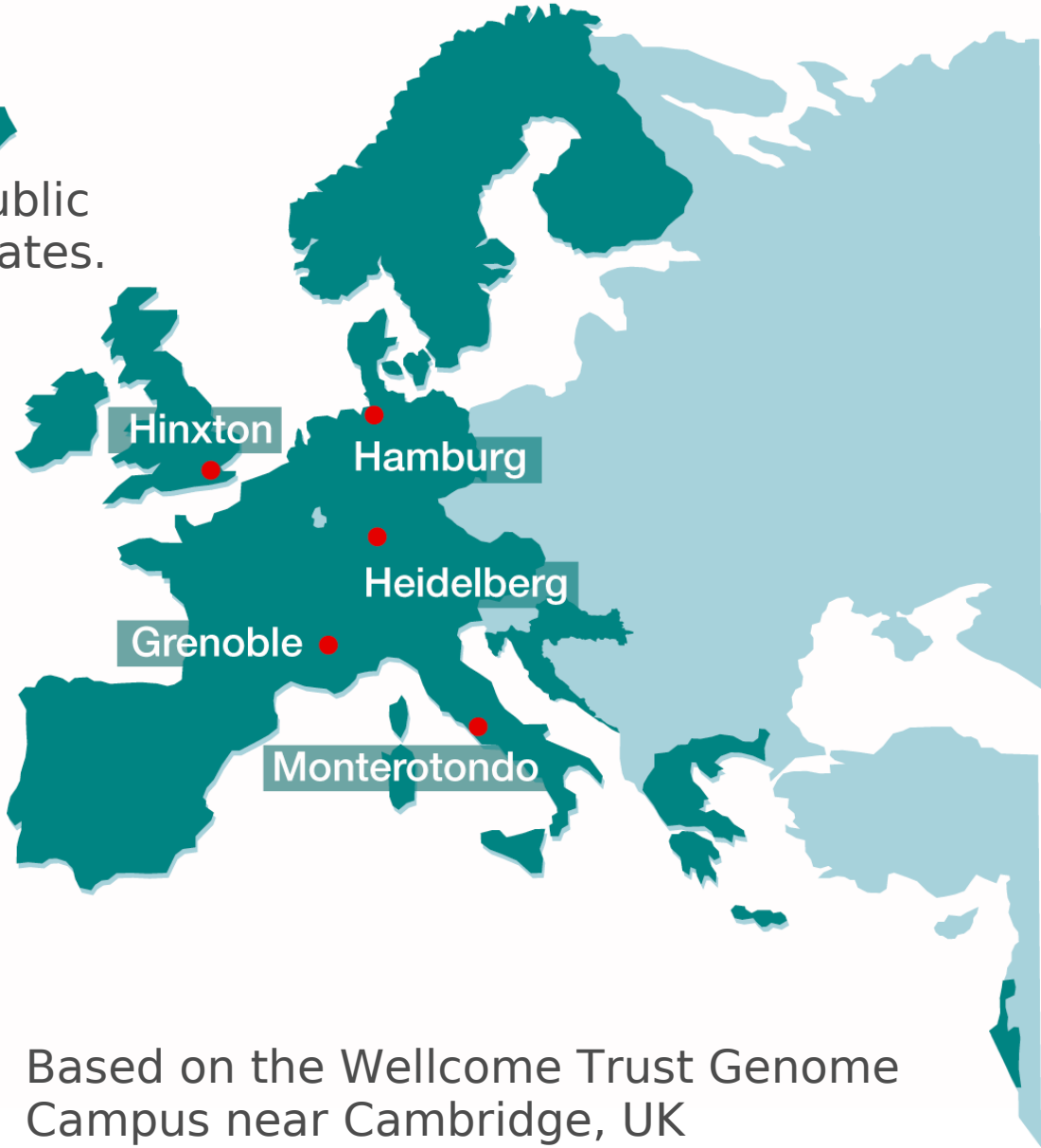
Nicolas Le Novère, EMBL-EBI

EBI is part of EMBL

- Non-profit organization
- Part of the European Molecular Biology Laboratory (EMBL), a basic research institute funded by public research monies from 19 member states.



- Based on the Wellcome Trust Genome Campus near Cambridge, UK

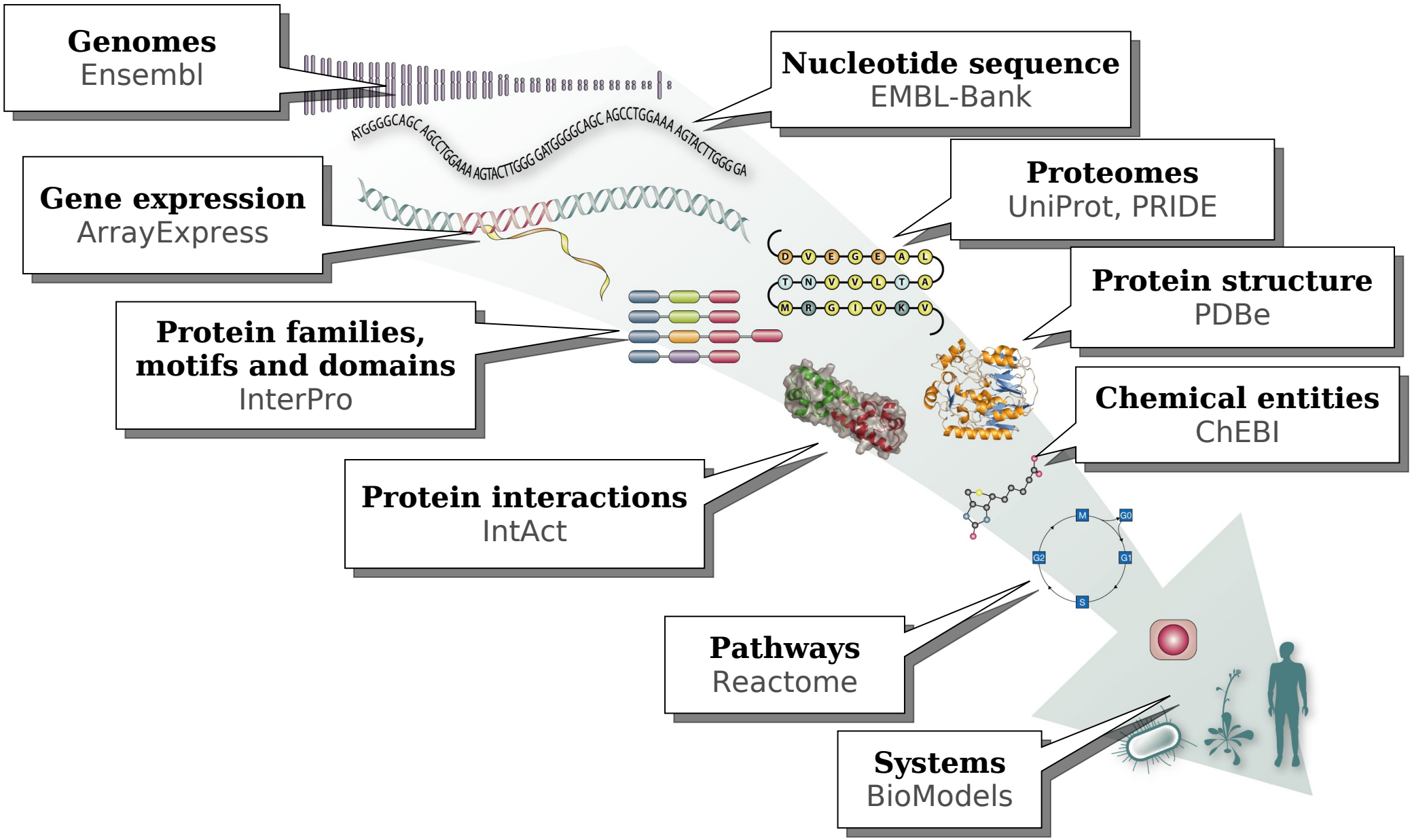
EMBL-EBI's missions

- To provide freely available **data** and bioinformatics **services** to all facets of the scientific community in ways that promote scientific progress
- To contribute to the advancement of biology through basic investigator-driven **research** in bioinformatics
- To provide advanced bioinformatics **training** to scientists at all levels, from PhD students to independent investigators
- To help disseminate cutting-edge technologies to **industry**

Principles of service provision

- **Accessibility** – all data and tools freely available without restriction
- **Compatibility** – we develop and promote the use of standards in bioinformatics
- **Comprehensive data sets** – agreements with other data providers ensure that our resources contain comprehensive and up-to-date data; agreements with publishers ensure that published data are placed in a public repository at the earliest opportunity
- **Portability** – data and software can be downloaded and installed locally
- **Quality** – Our databases are enhanced through annotation and cross-referencing

EBI Data resources: molecules to systems



Standards development

Nucleotide sequence
MIGS

Genome annotation
Gene Ontology

Protein sequence
IPI

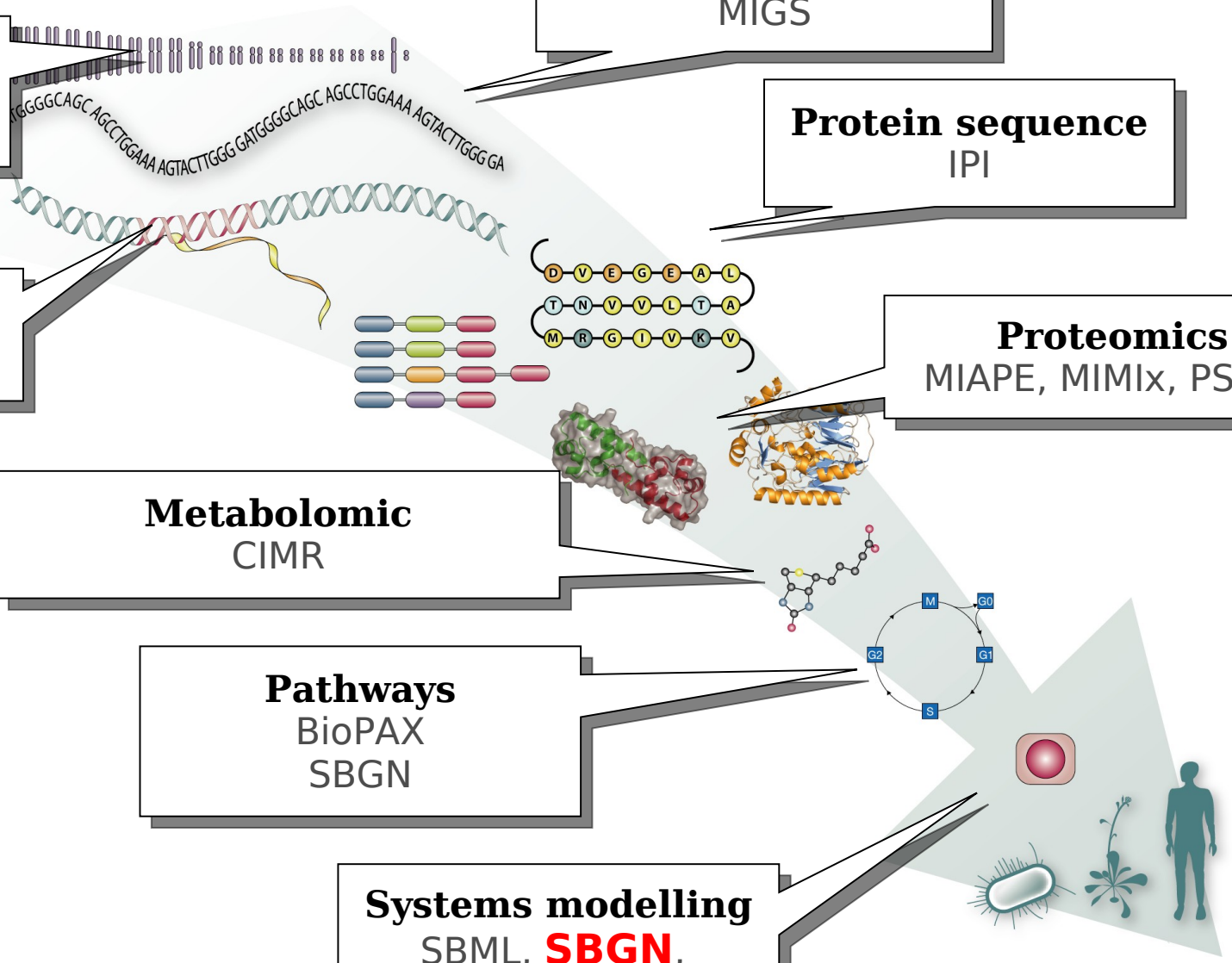
Transcriptomics
MIAME

Proteomics
MIAPE, MIMIx, PSI-MI

Metabolomic
CIMR

Pathways
BioPAX
SBGN

Systems modelling
SBML, **SBGN**,
MIRIAM, SBO



Research groups

Goldman
Evolutionary sequence analysis

Rebholz-Schuhmann
Text mining

Huber
Functional genomics

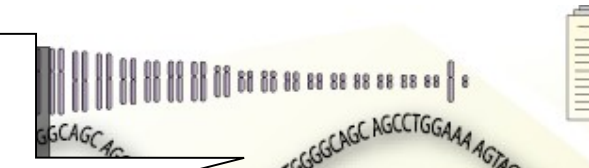
Enright
microRNAs

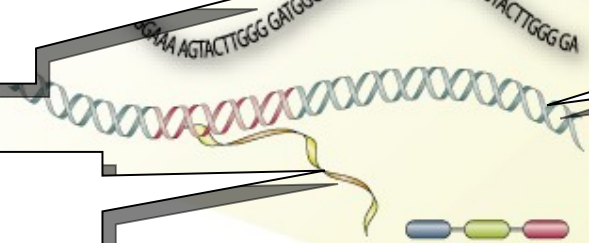
Thornton
Structural bioinformatics

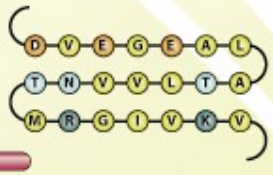
Le Novère
Computational systems neurobiology

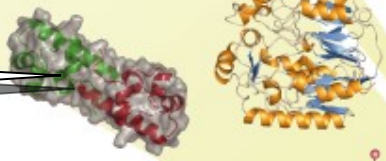
Luscombe
Regulatory networks

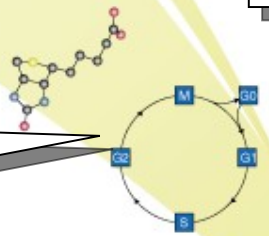
Bertone
Differentiation and development









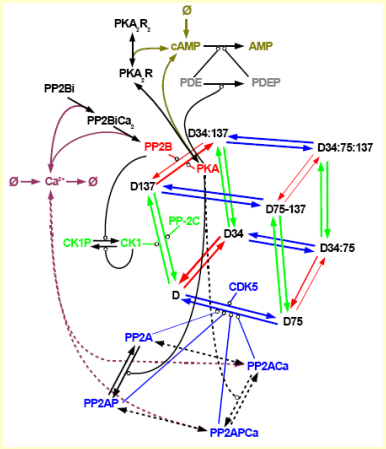


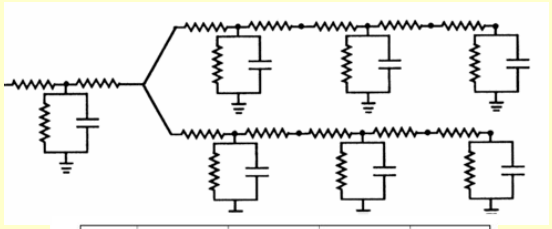
EBI usage

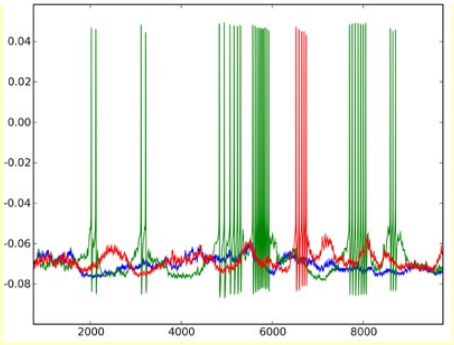
- >3 million web requests per day
- >300 000 unique hosts served per month
- ~5 million files downloaded per month
- ~173 TB data downloaded in 2008 (excluding data sync)
- Millions of cross-references
- 5 petabytes of storage
- > 1000 cores in clusters
- 1 million compute jobs per day

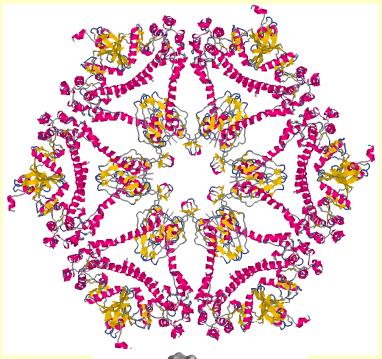
Group themes and projects

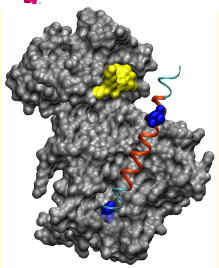
Computational Neurobiology

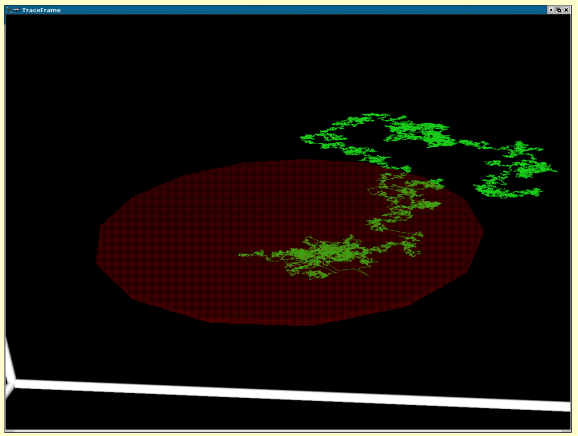








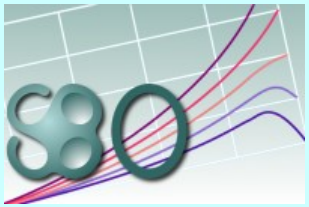


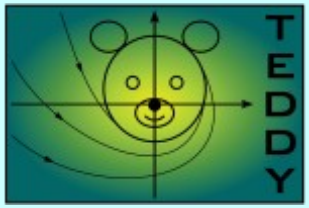


Systems Biology

SBML

SBGN





KISAO

e-MeP

Ligand-Gated Ion Channel database

MIRIAM

MIASE

BioModels

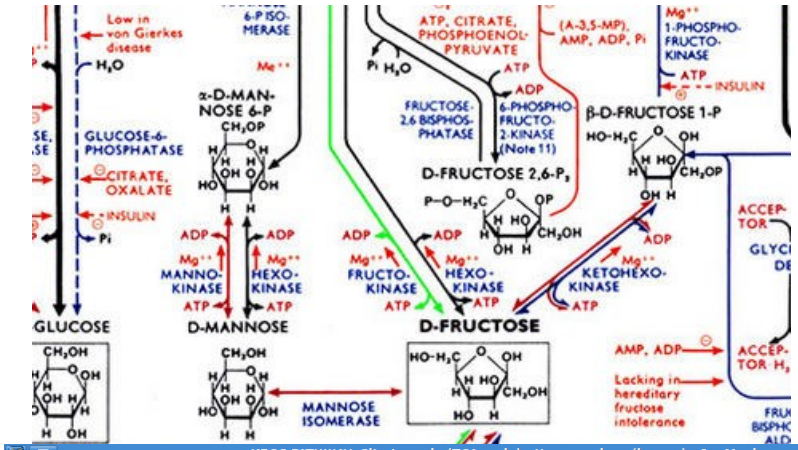
http://www.sbgn.org/

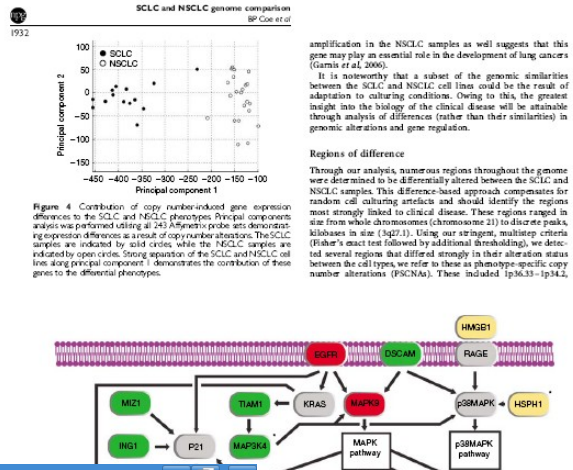
The Systems Biology Graphical Notation

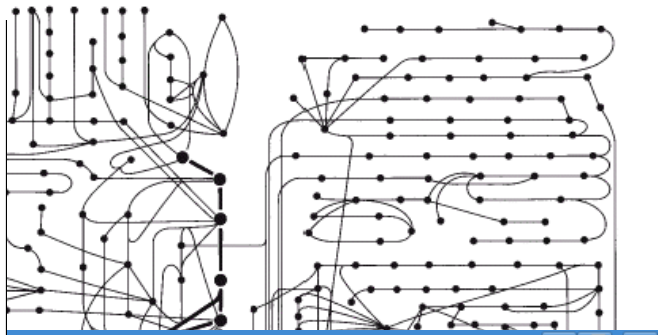
Nicolas Le Novère, EMBL-EBI

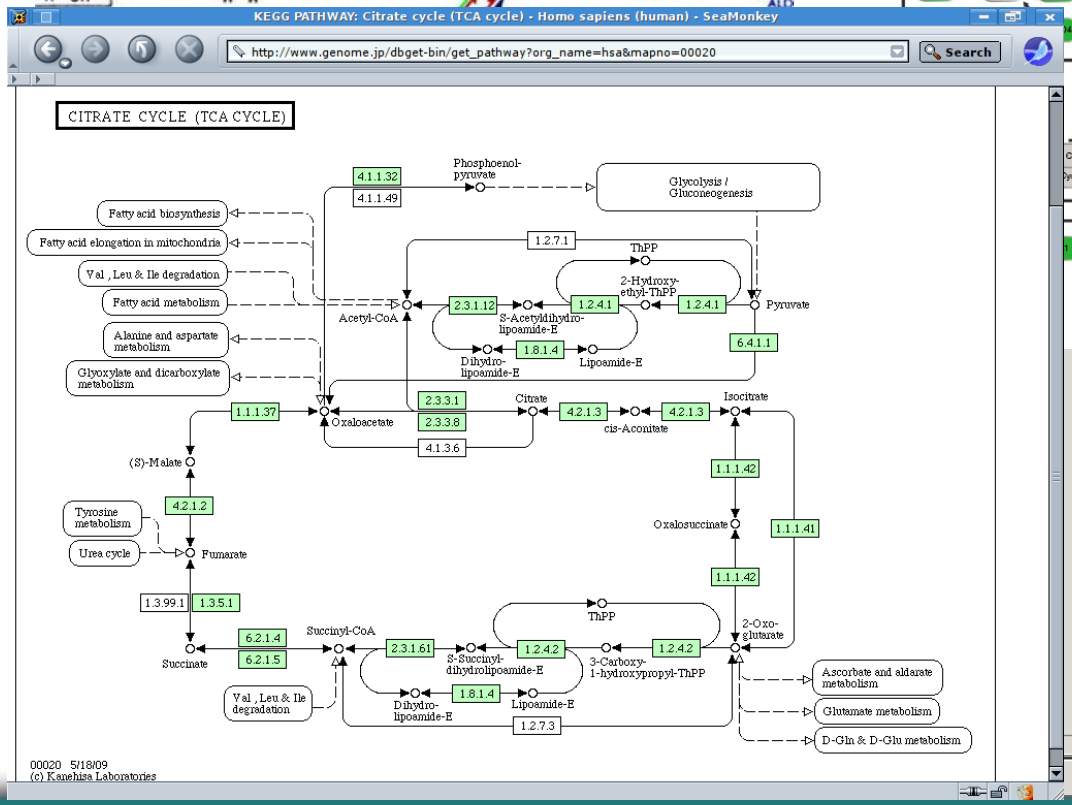
(on the behalf of SBGN editors, authors and contributors)

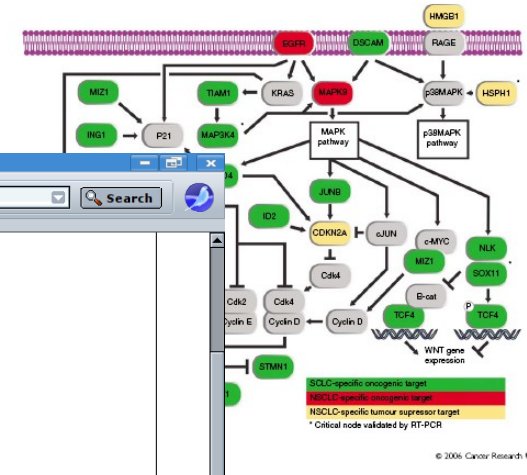
Graphs are everywhere in biology











Ambiguity of usual representations

Ambiguity of usual representations

is transformed into

translocates (X "=" Y)

is degraded into

associates into

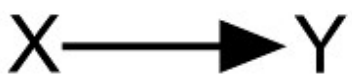
dissociates into

stimulates the activity of

stimulates the expression of

catalyses the formation of

Ambiguity of usual representations



X inhibits Y

is transformed into

translocates (X "=" Y)

is degraded into

associates into

dissociates into

stimulates the activity of

stimulates the expression of

catalyses the formation of

Ambiguity of usual representations

is transformed into

translocates (X "=" Y)

is degraded into

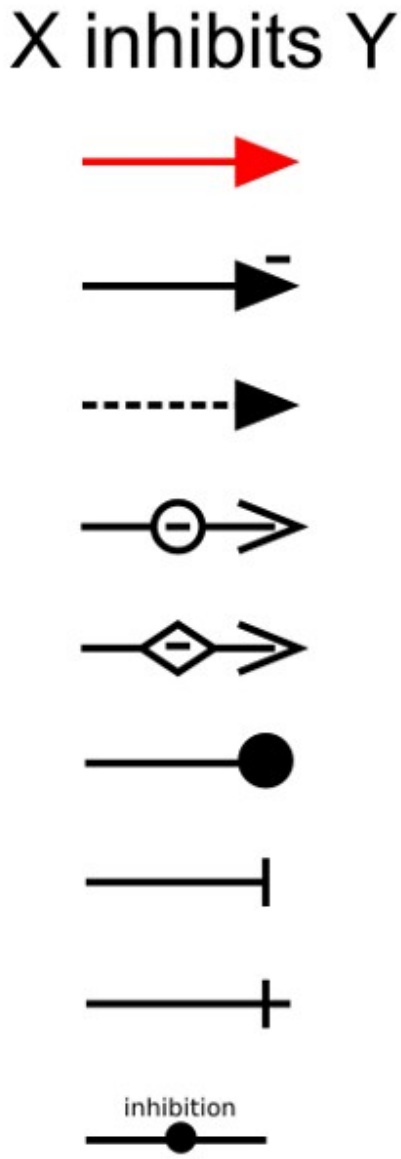
associates into

dissociates into

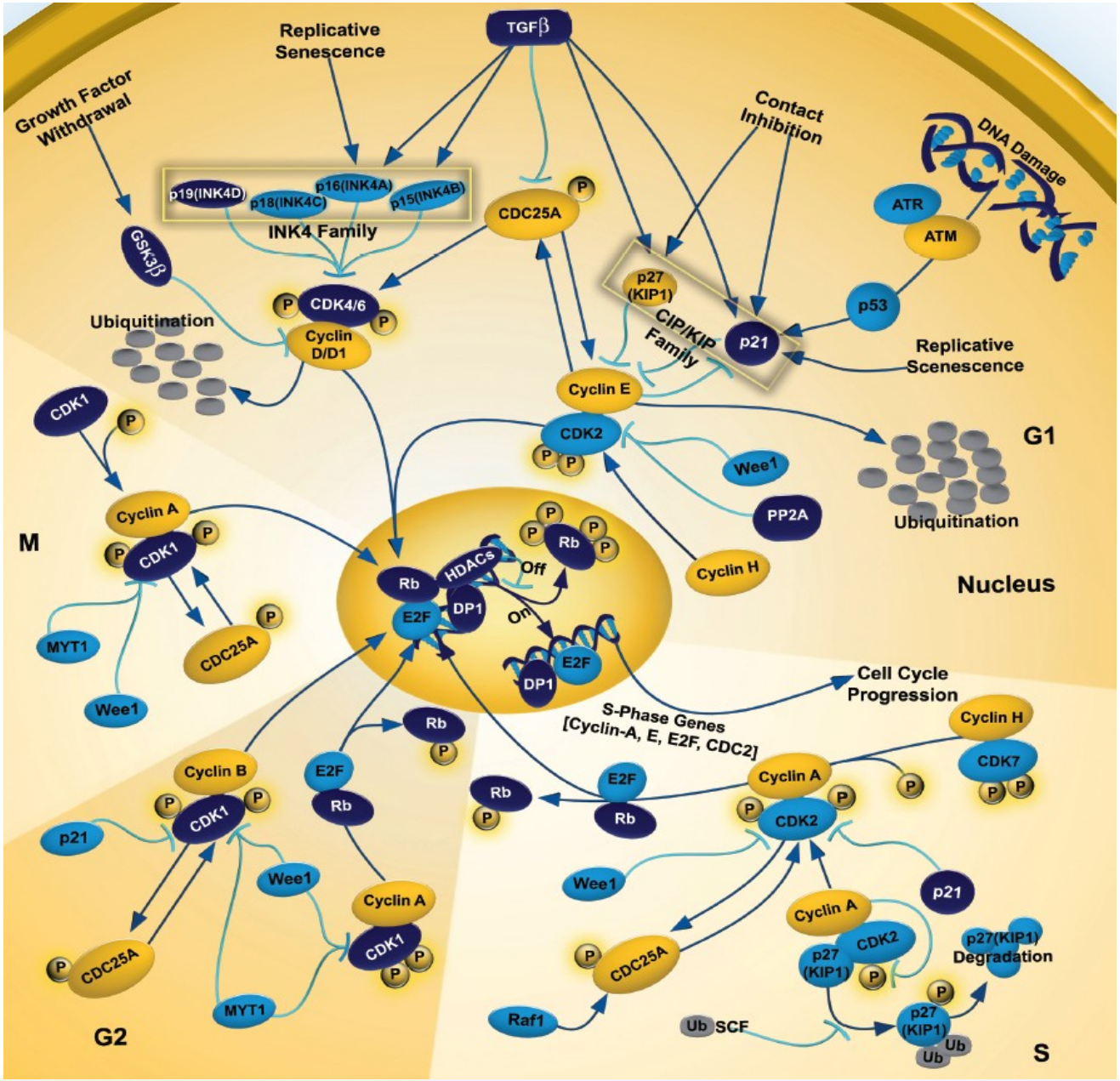
stimulates the activity of

stimulates the expression of

catalyses the formation of



Can-this be understood by biologists?



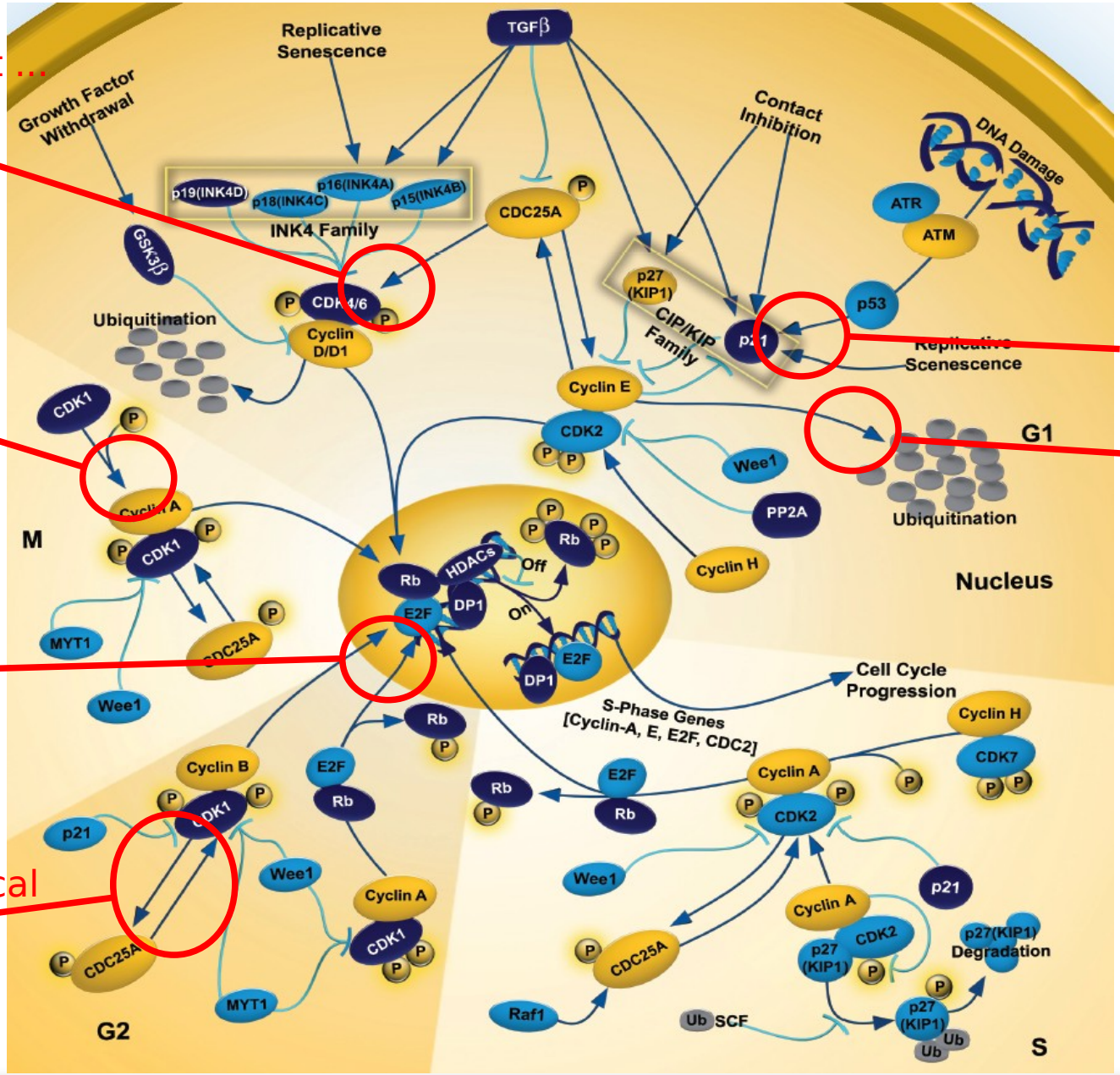
Can-this be understood by biologists?

Stimulates? but ... what exactly?

Associates into?

Translocates?

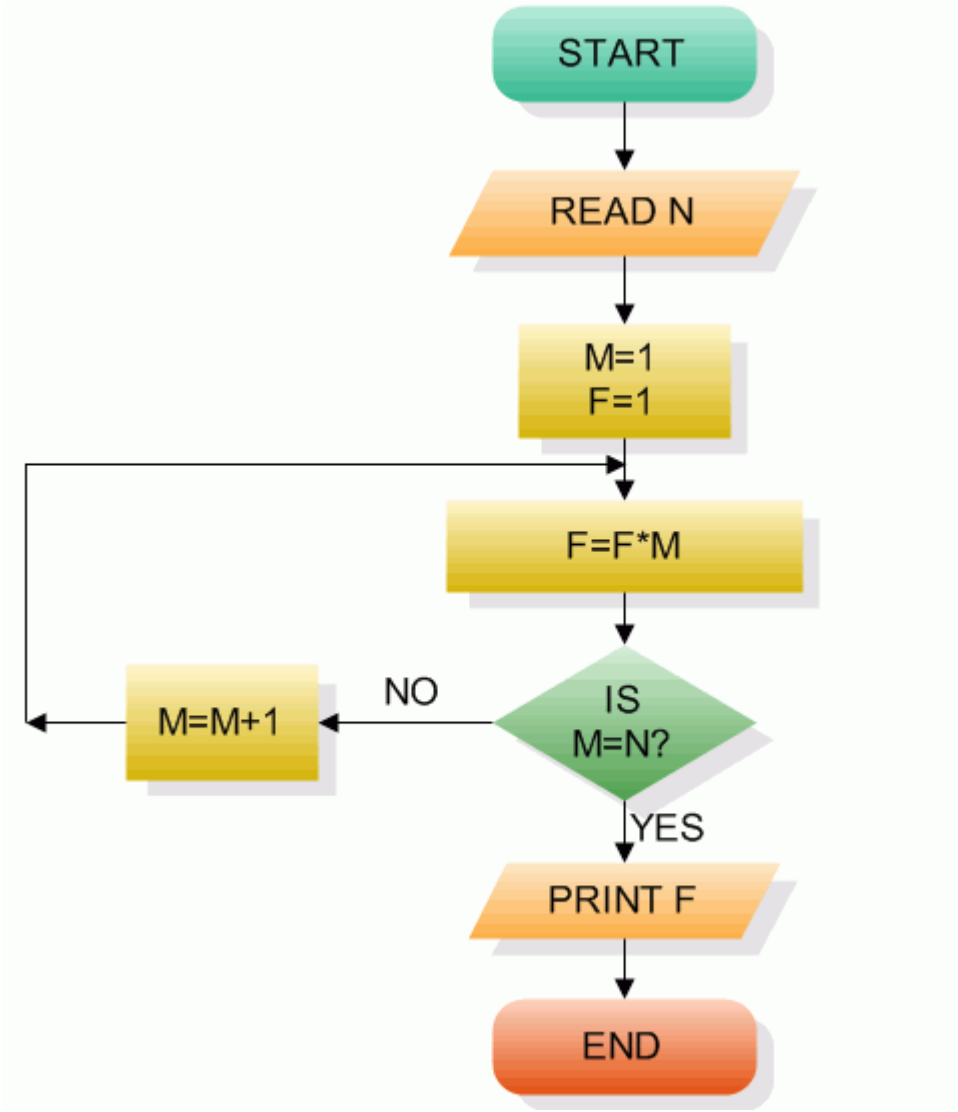
No idea. Reciprocal stimulation?

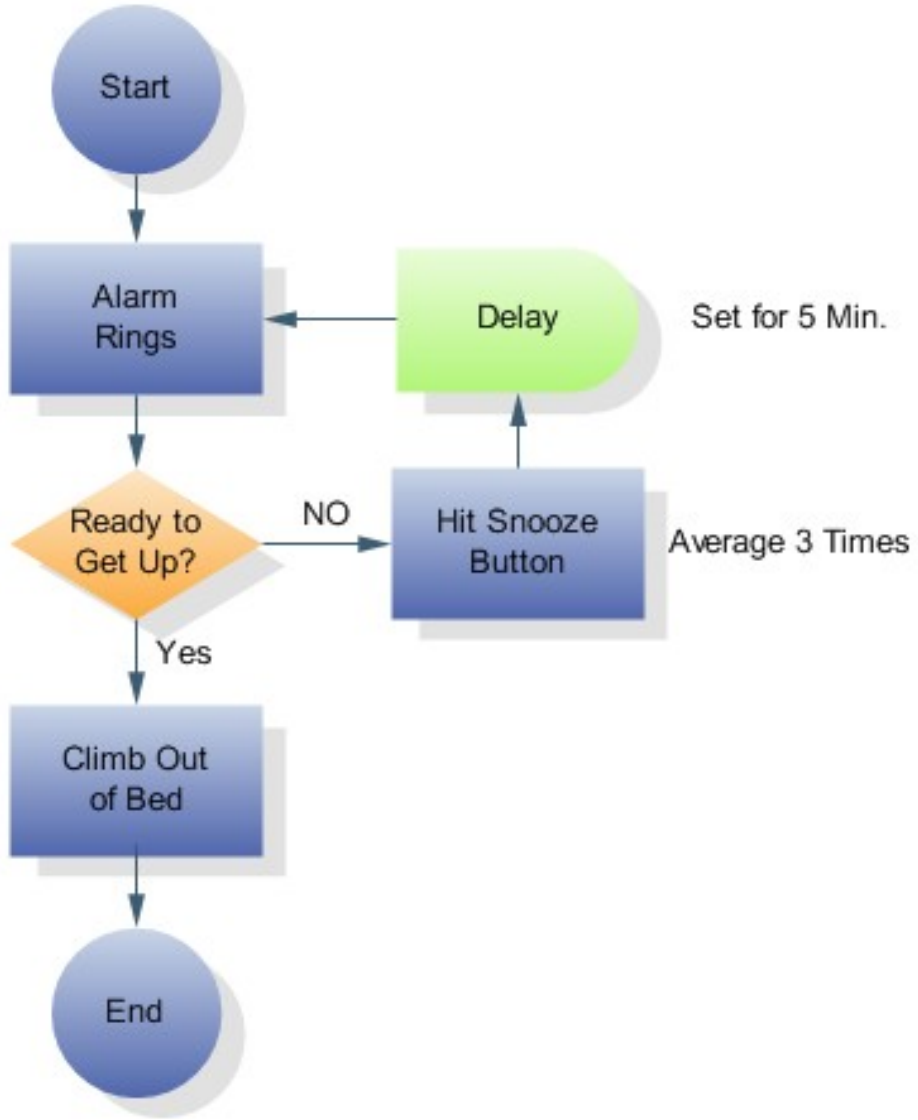


Stimulates gene transcription?

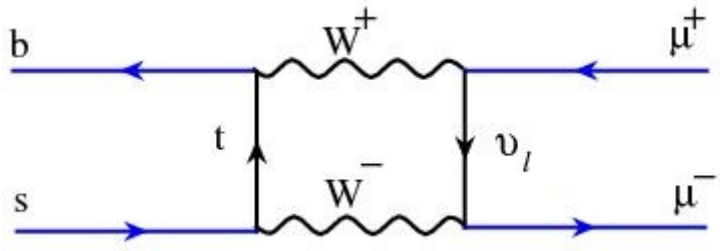
Is degraded?

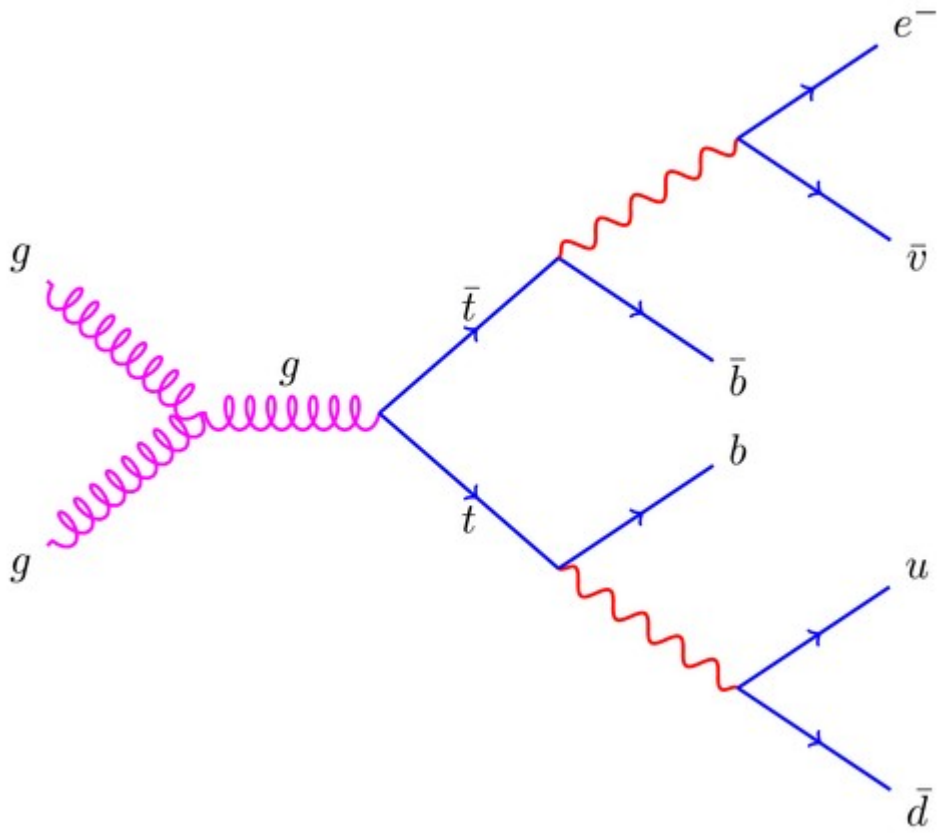
Every computer scientist understands those



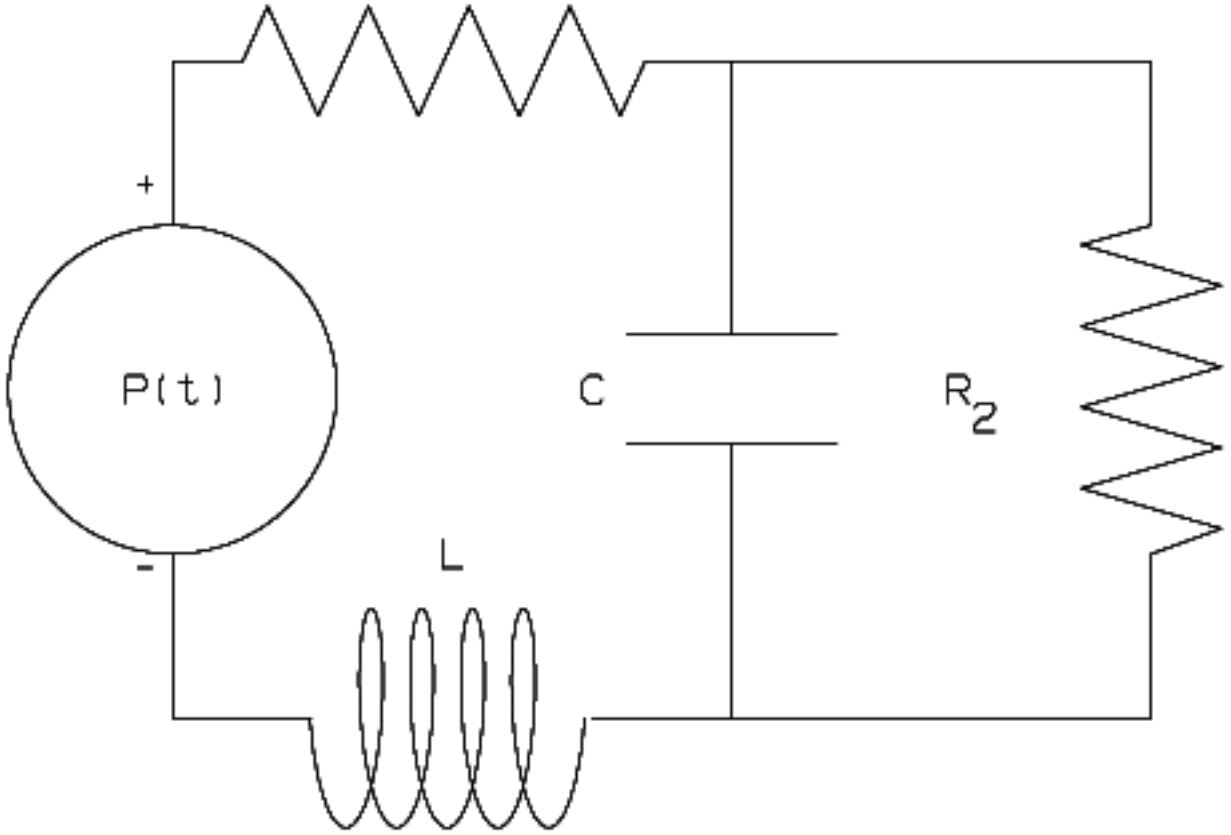


Every physicist understands those

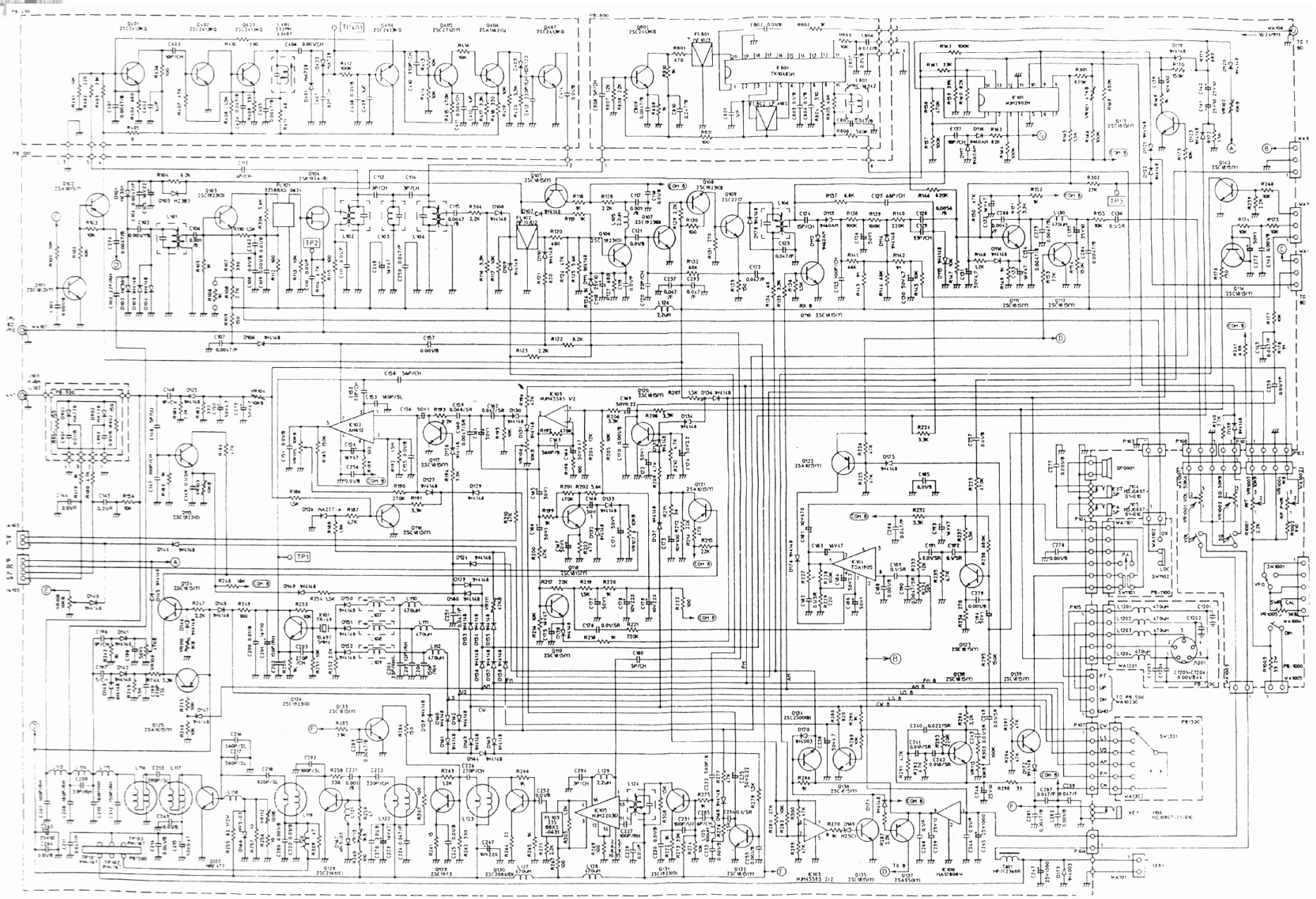




Every engineer understands that



Or that

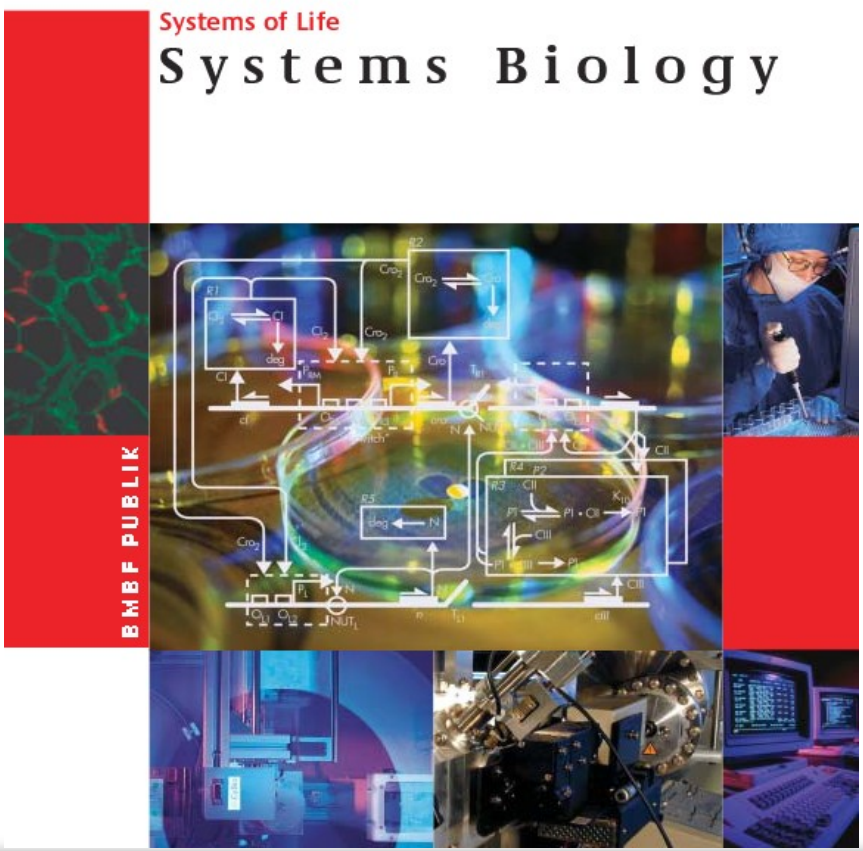


What did-those diagram bring?

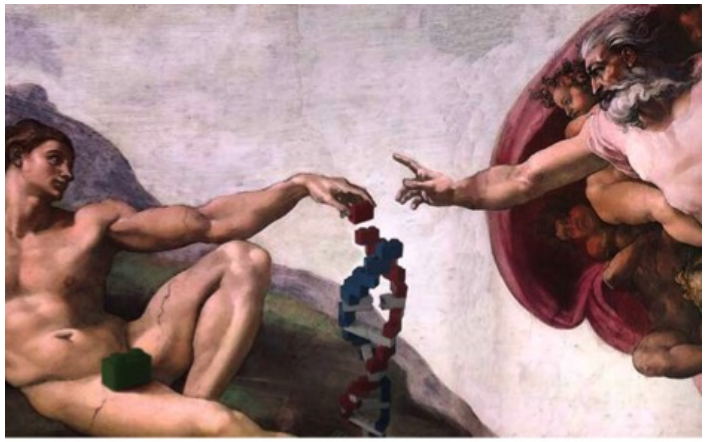
What do-we expect in modern (future) life science

Federal Ministry
of Education
and Research

Basic science



Technology



EXTREME GENETIC ENGINEERING

An Introduction to Synthetic Biology

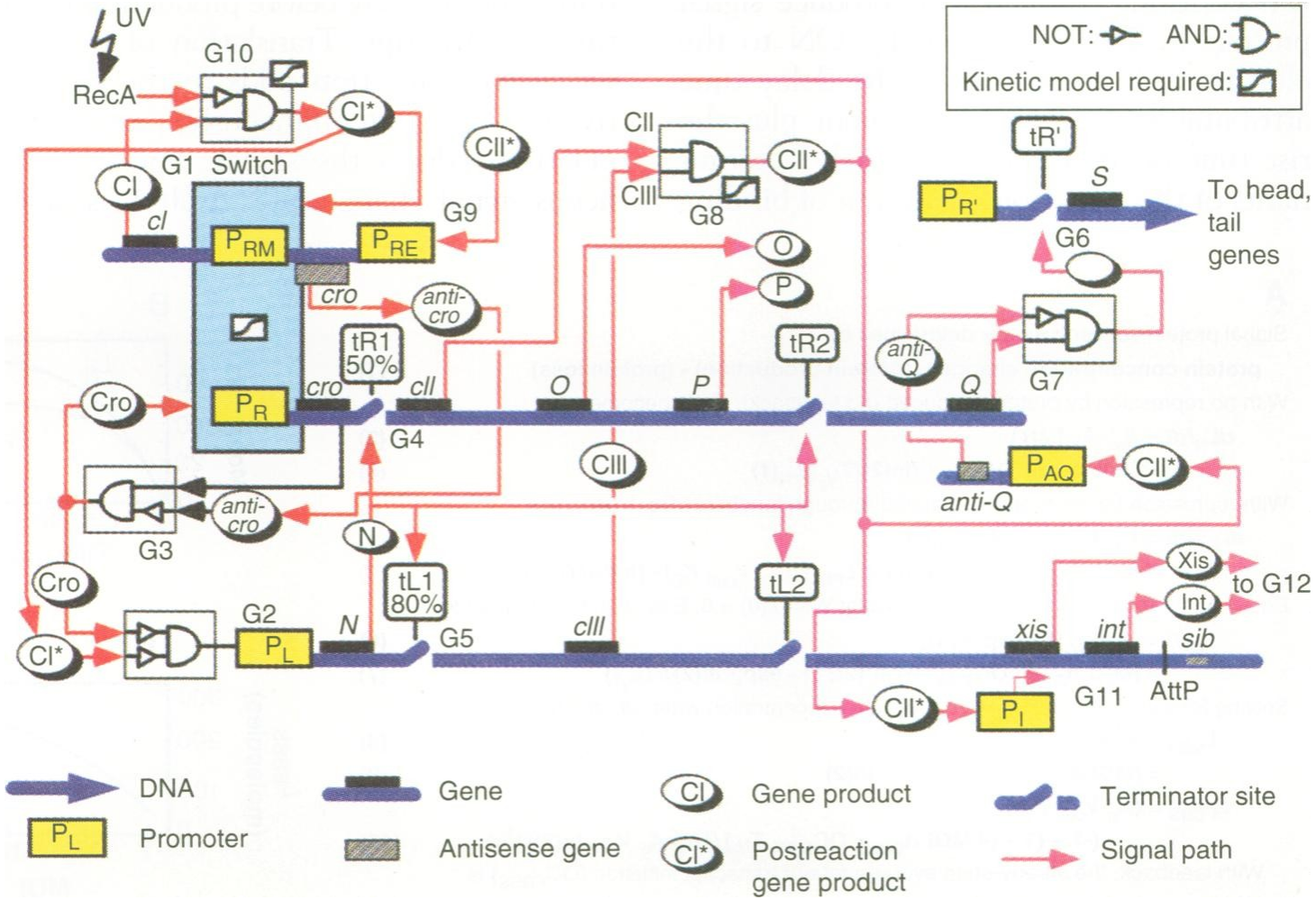
etc group

January 2007

Why can-we understand electric diagrams?

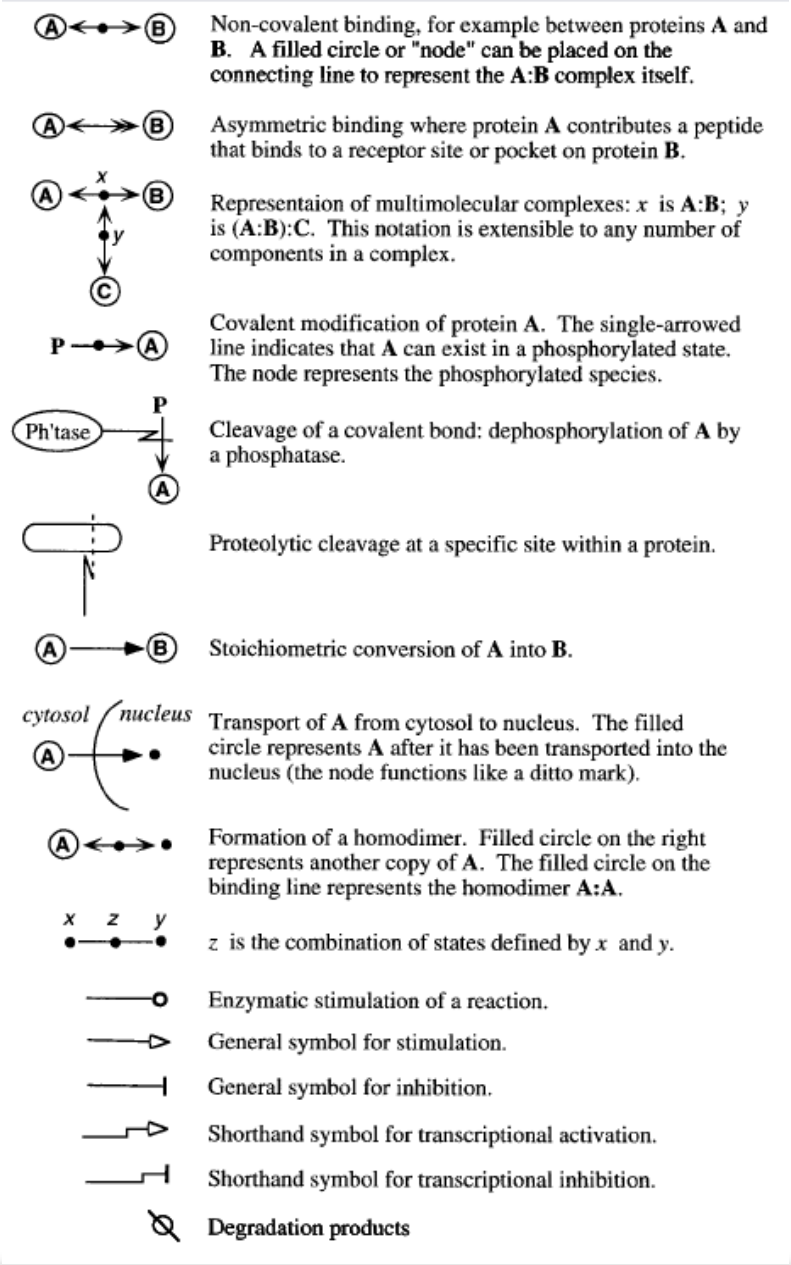
- Standard symbols
  - Simple shapes, easily recognisable
  - Limited number of basic symbols (<70)
  - Similarity of shapes reflects similarity of functions
- Unambiguous interpretation of the circuits
- Endorsed by the community for practical reasons
  - End-users: manufacturers
  - Tool developers
  - Publishing industry
  - Teaching communities

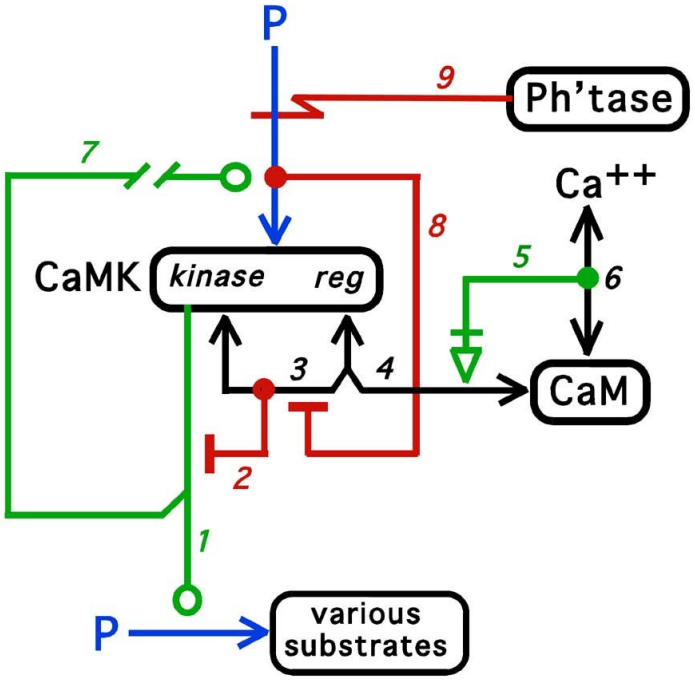
Using electrical diagrams



- McAdams and Shapiro (1995) Science, 269: 650-657

Kohn's Molecular Interaction Maps





- Kurt W Kohn (1998) *Oncogene*, 16: 1065-1075
- Kurt W. Kohn (1999) *Mol Biol Cell*, 10(8):2703-2734

MIM are Entity Relationship diagrams

- Non-sequential interactions between entities
- Each *entity* appears only once
- Complexes are defined by interactions
- Modulations of relationships possible
- Order of interactions is meaningful (only binary ones)
- Angles between edges are meaningful
- ...
- Designed manually

Cell Cycle in MIM

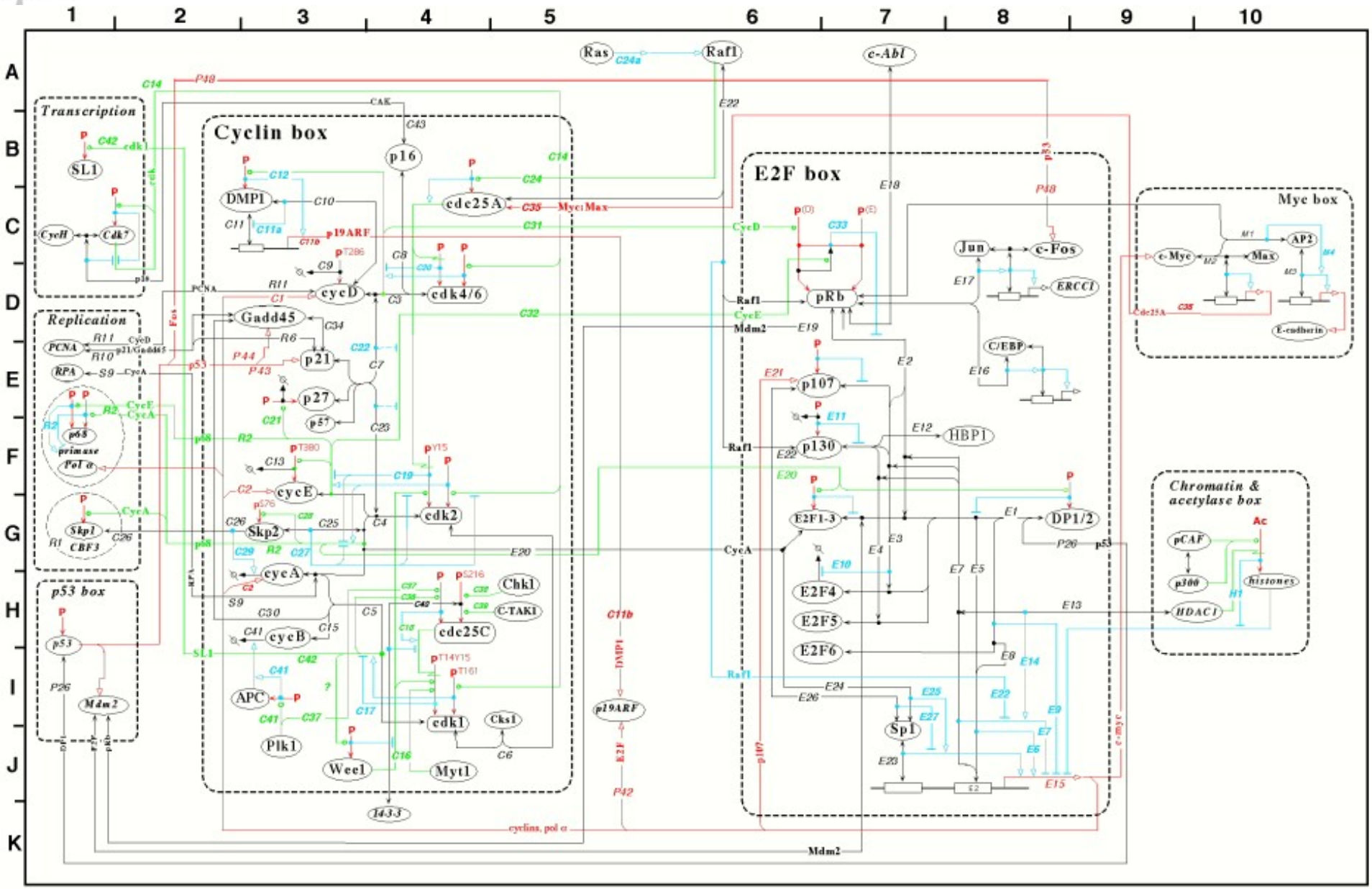
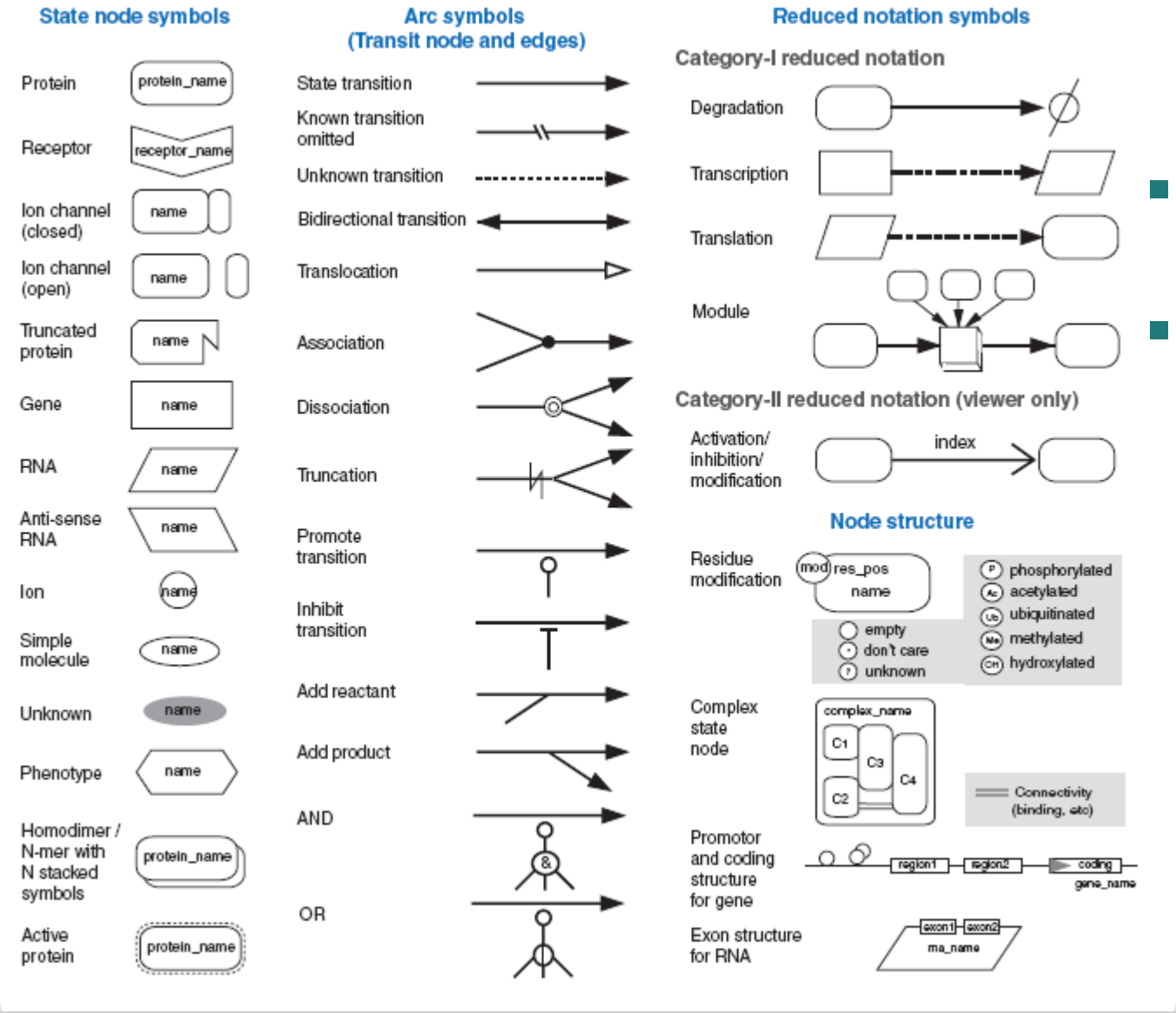


Figure 6A: **The Cyclin - E2F cell cycle control system** (version 3a - June 8, 1999)

Kitano's Notation

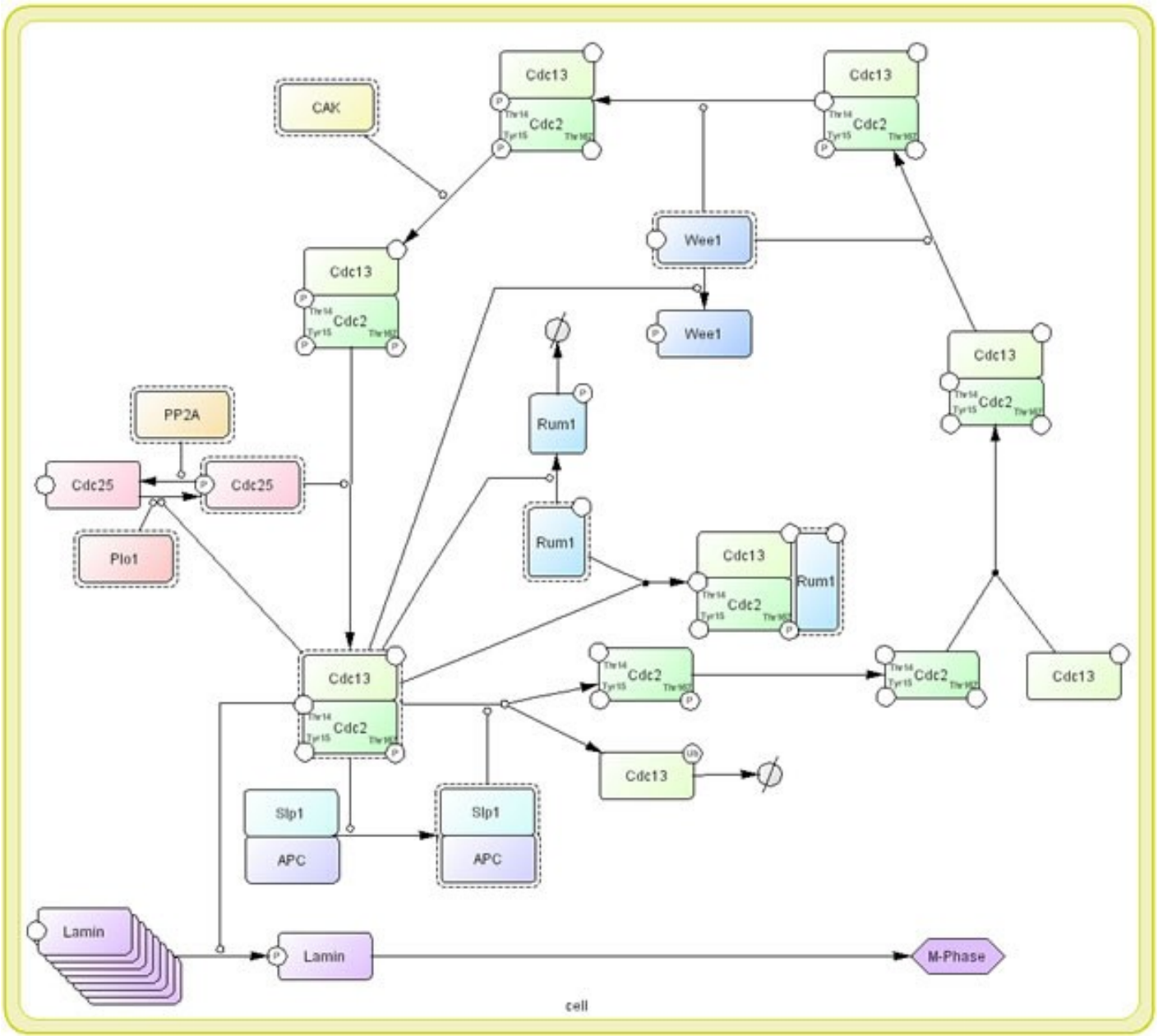


- Kitano (2003) *Biosilico*, 1: 169-170
- Kitano et al (2005) *Nat Biotech*, 23: 961-966

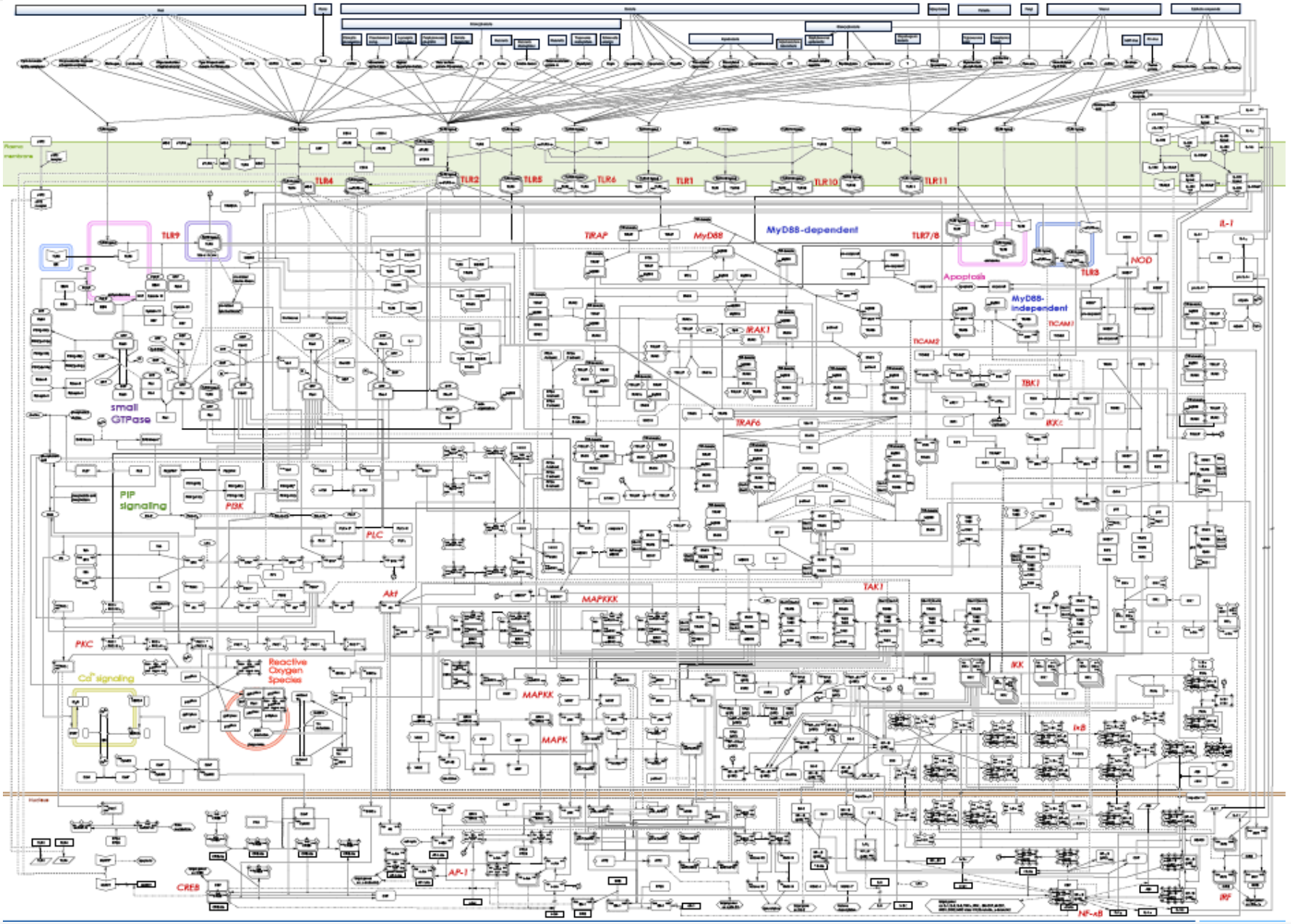
Kitano's graphs are Process Diagrams

- Causal sequence of molecular processes and their results
- Each *state* appears only once
- Complexes are defined as independent entity nodes
- Correspondence with *pathway* descriptions and reaction-based *models*
- Some software support (CellDesigner)

Cell Cycle in Kitano's Process Diagram



Toll-like signalling



Why were those pioneer efforts not suitable?

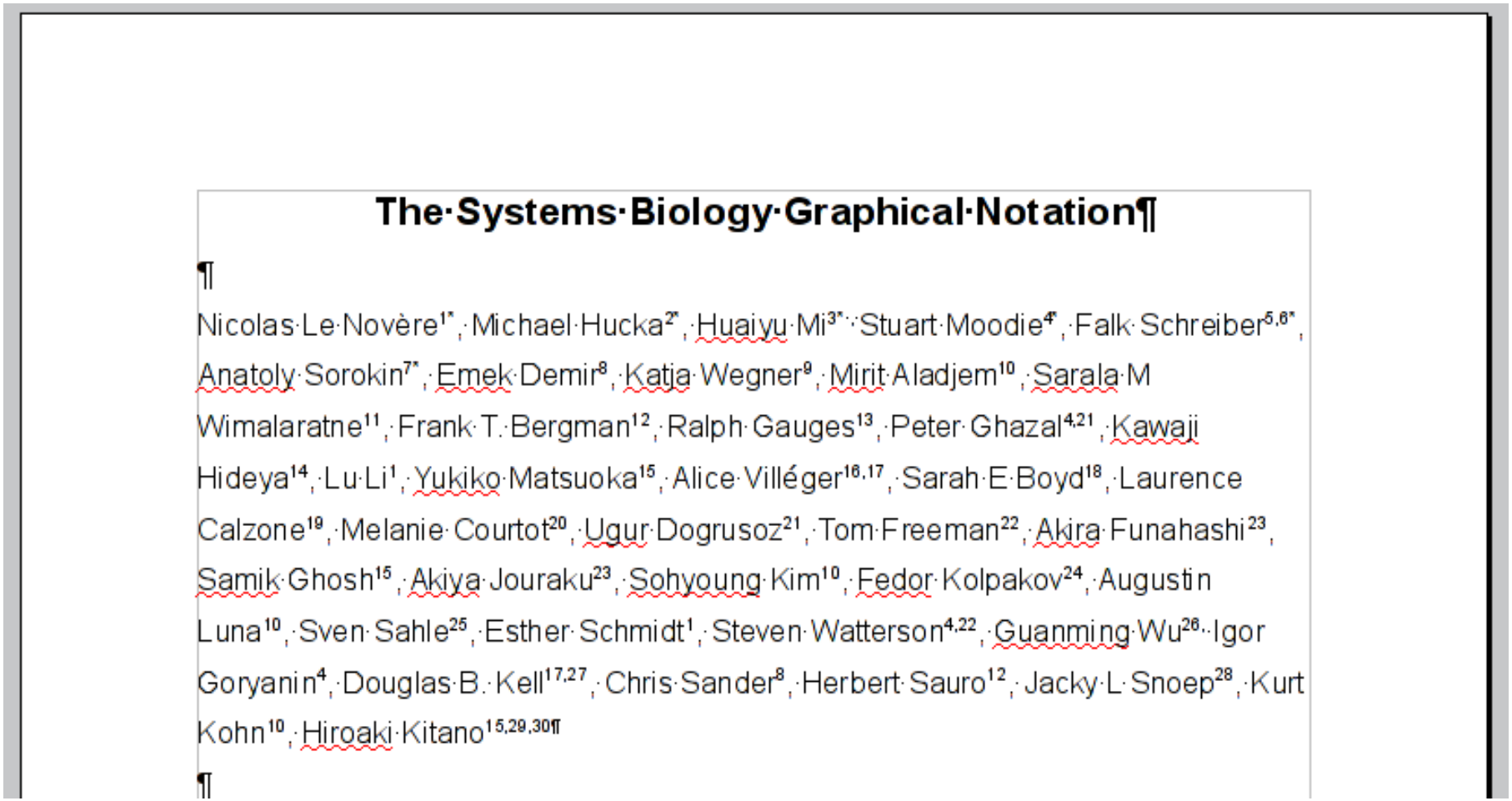
- Somehow fuzzy semantics
  - No structured data model or ontology behind the notation
  - Overlapping concepts rather than sub-classing
  - Gaps in the coverage of biochemistry or modelling
  - Ambiguous interpretation of the graph
- Little software support (except CellDesigner for Kitano's Process Diagrams)
- No community involvement
  - No systematic bug tracking and consistency checking
  - No comprehensive coverage (focussed on some use-cases)
  - No endorsement by the tool developers or by the end-users

Enters
The Systems Biology Markup Language

http://www.sbgn.org/

What is SBGN?

- A way to unambiguously describe biochemical and cellular events in graphs
- Limited amount of symbols ☞ Smooth learning curve
- Can graphically represent quantitative models, biochemical pathways, at different levels of granularity
- Developed over three years by a growing community



The·Systems·Biology·Graphical·Notation¶

¶

Nicolas·Le·Novère[1*],·Michael·Hucka[2*],·Huaiyu·Mi[3*],·Stuart·Moodie[4*],·Falk·Schreiber[5,6*],·Anatoly·Sorokin[7*],·Emek·Demir[8],·Katja·Wegner[9],·Mirit·Aladjem[10],·Sarala·M Wimalaratne[11],·Frank·T.·Bergman[12],·Ralph·Gauges[13],·Peter·Ghazal[4,21],·Kawaji Hideya[14],·Lu·Li[1],·Yukiko·Matsuoka[15],·Alice·Villéger[16,17],·Sarah·E·Boyd[18],·Laurence Calzone[19],·Melanie·Courtot[20],·Ugur·Dogrusoz[21],·Tom·Freeman[22],·Akira·Funahashi[23],·Samik·Ghosh[15],·Akiya·Jouraku[23],·Sohyoung·Kim[10],·Fedor·Kolpakov[24],·Augustin Luna[10],·Sven·Sahle[25],·Esther·Schmidt[1],·Steven·Watterson[4,22],·Guanming·Wu[26],·Igor Goryanin[4],·Douglas·B.·Kell[17,27],·Chris·Sander[8],·Herbert·Sauro[12],·Jacky·L·Snoep[28],·Kurt Kohn[10],·Hiroaki·Kitano[15,29,30]¶

¶

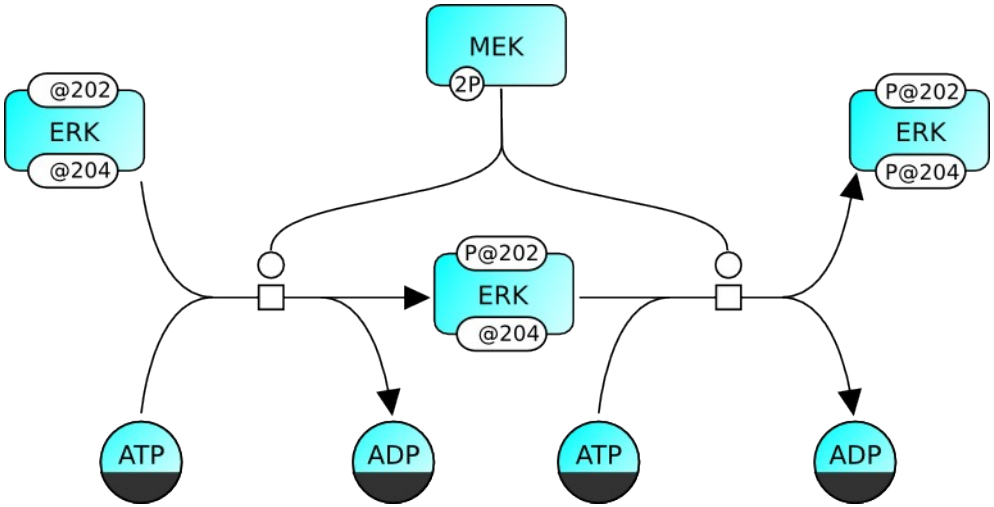
39 authors, 30 affiliations

What is SBGN?

- A way to unambiguously describe biochemical and cellular events in graphs
- Limited amount of symbols ☞ Smooth learning curve
- Can graphically represent quantitative models, biochemical pathways, at different levels of granularity
- Developed over three years by a growing community
- Three languages
  - Process Diagrams ☞ one state = one glyph, biochemical level
  - Entity Relationships ☞ one entity = one glyph, biochemical level
  - Activity Flow ☞ conceptual level

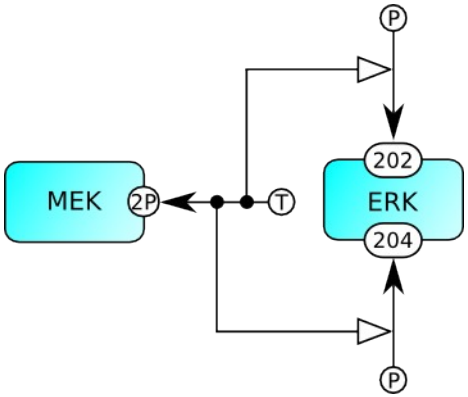
Graph trinity: three languages in one

Process diagrams



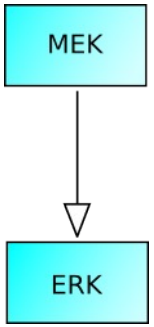
- Unambiguous
- Mechanistic
- Sequential
- Subject to combinatorial explosion

Entity Relationships diagrams



- Unambiguous
- Mechanistic
- Non-sequential

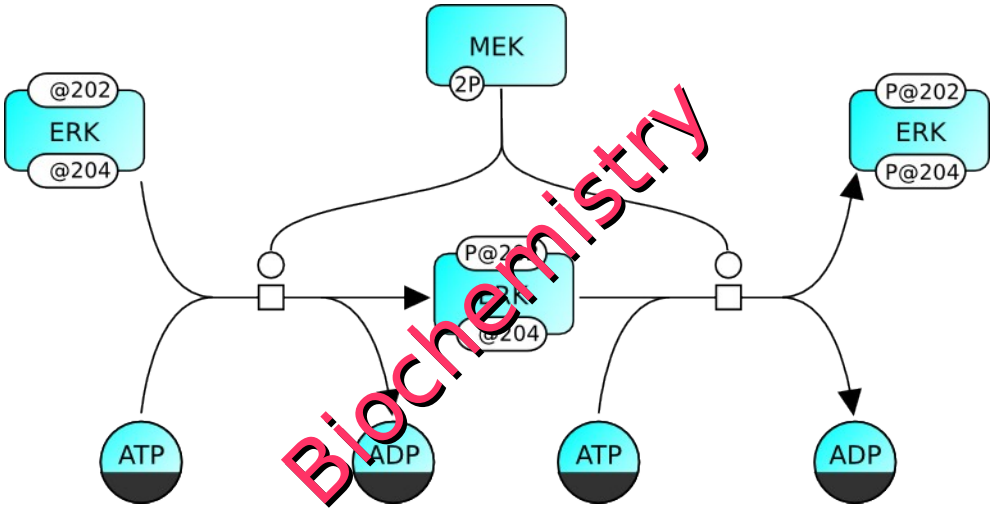
Activity Flow diagrams



- Ambiguous
- Conceptual
- Sequential

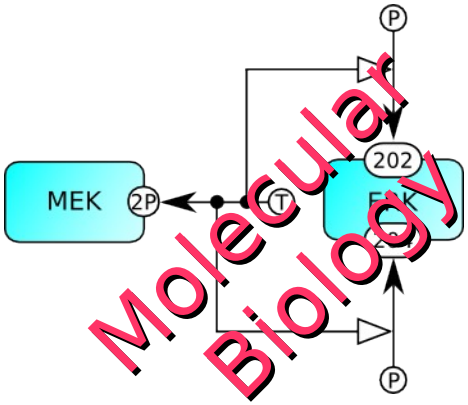
Graph trinity: three languages in one

Process diagrams



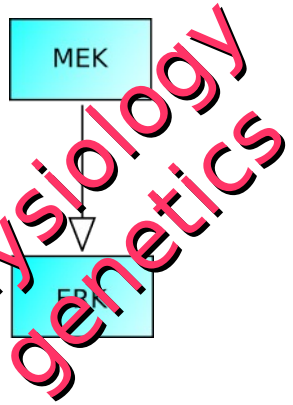
- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships diagrams



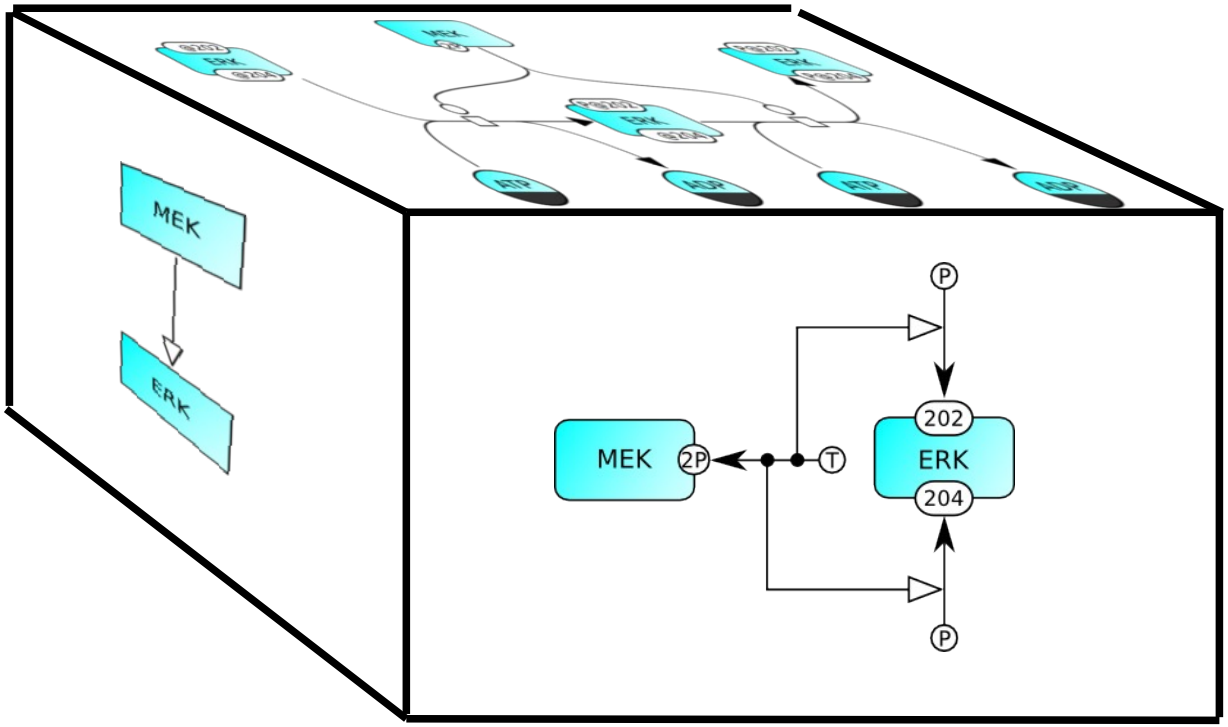
- Unambiguous
- Mechanistic
- Non-sequential
- Independence of relationships

Activity Flow diagrams

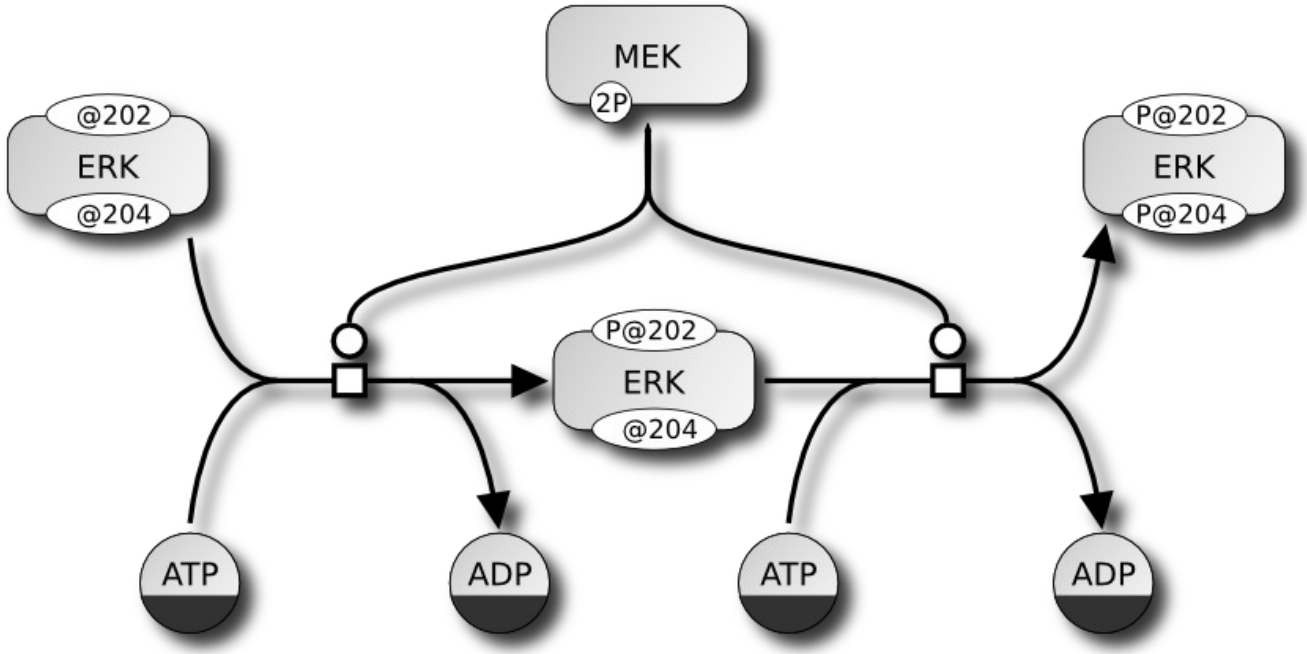


- Ambiguous
- Conceptual
- Sequential

Three orthogonal projections of biology

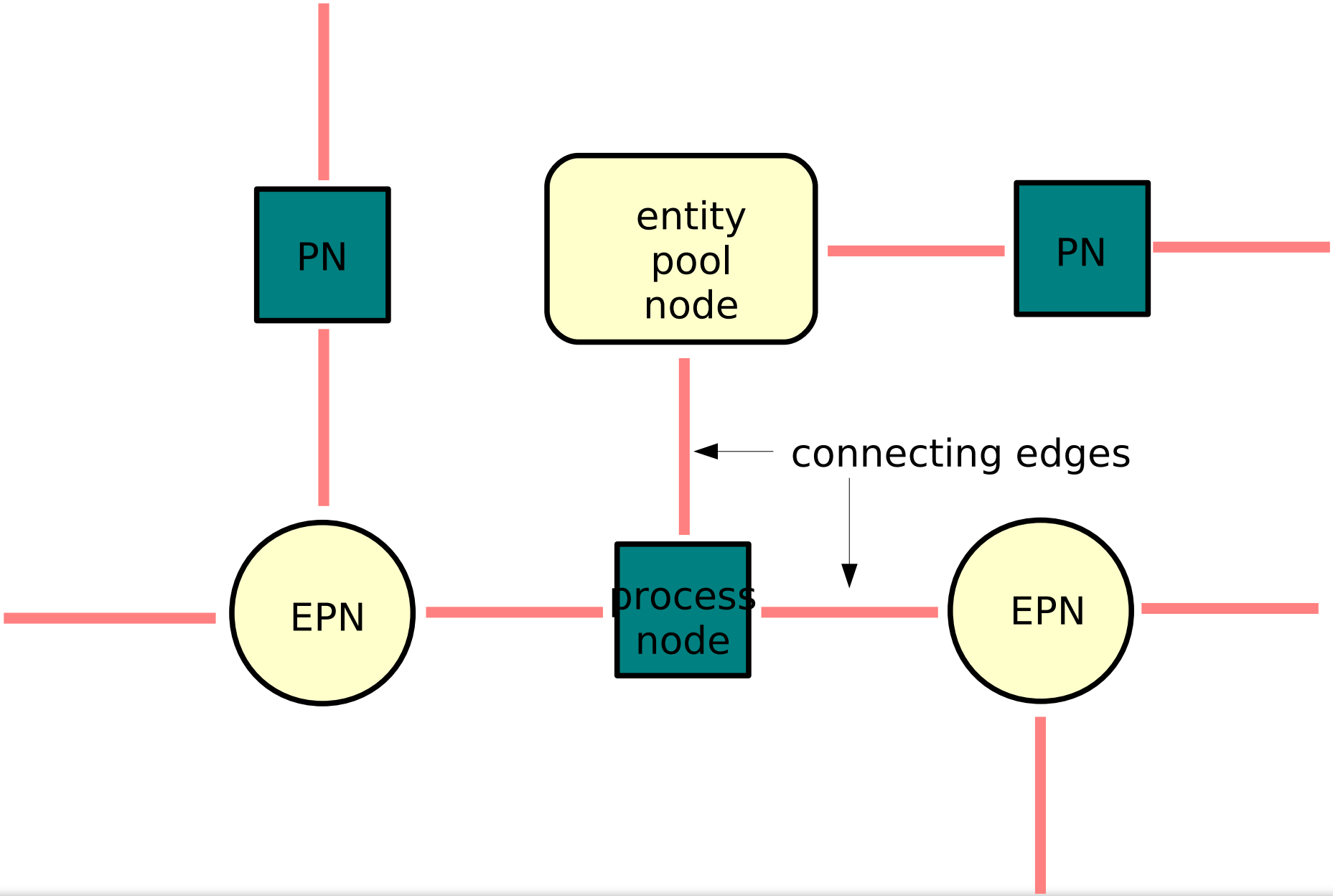


SBGN Process Diagram Level 1

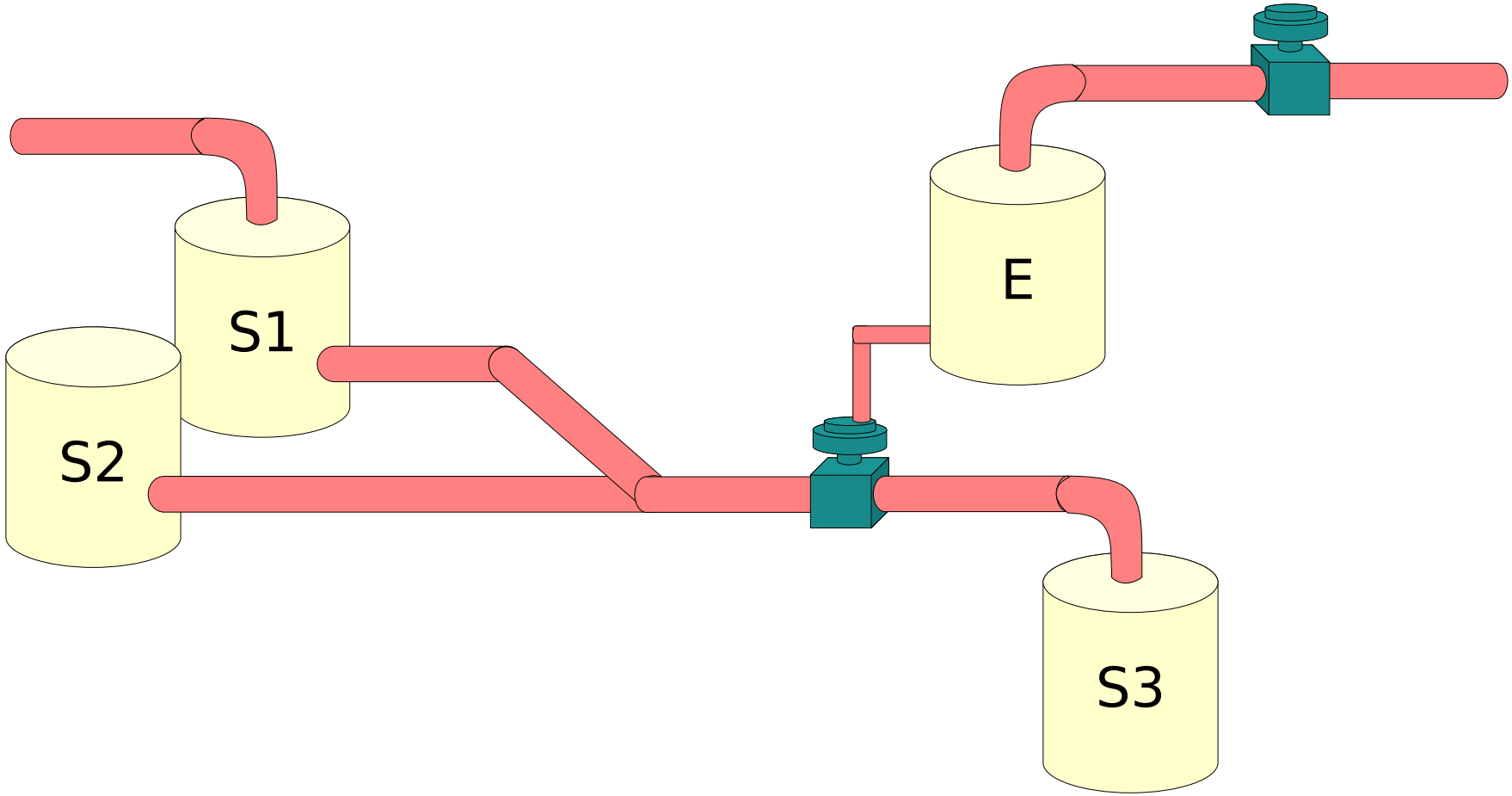


Le Novère, Moodie, Sorokin, Hucka, Schreiber, Demir, Mi, Matsuoka, Wegner, Kitano
Systems Biology Graphical Notation: Process Diagram Level 1 (2008)
Available from *Nature Precedings* <http://hdl.handle.net/10101/npre.2008.2320.1>

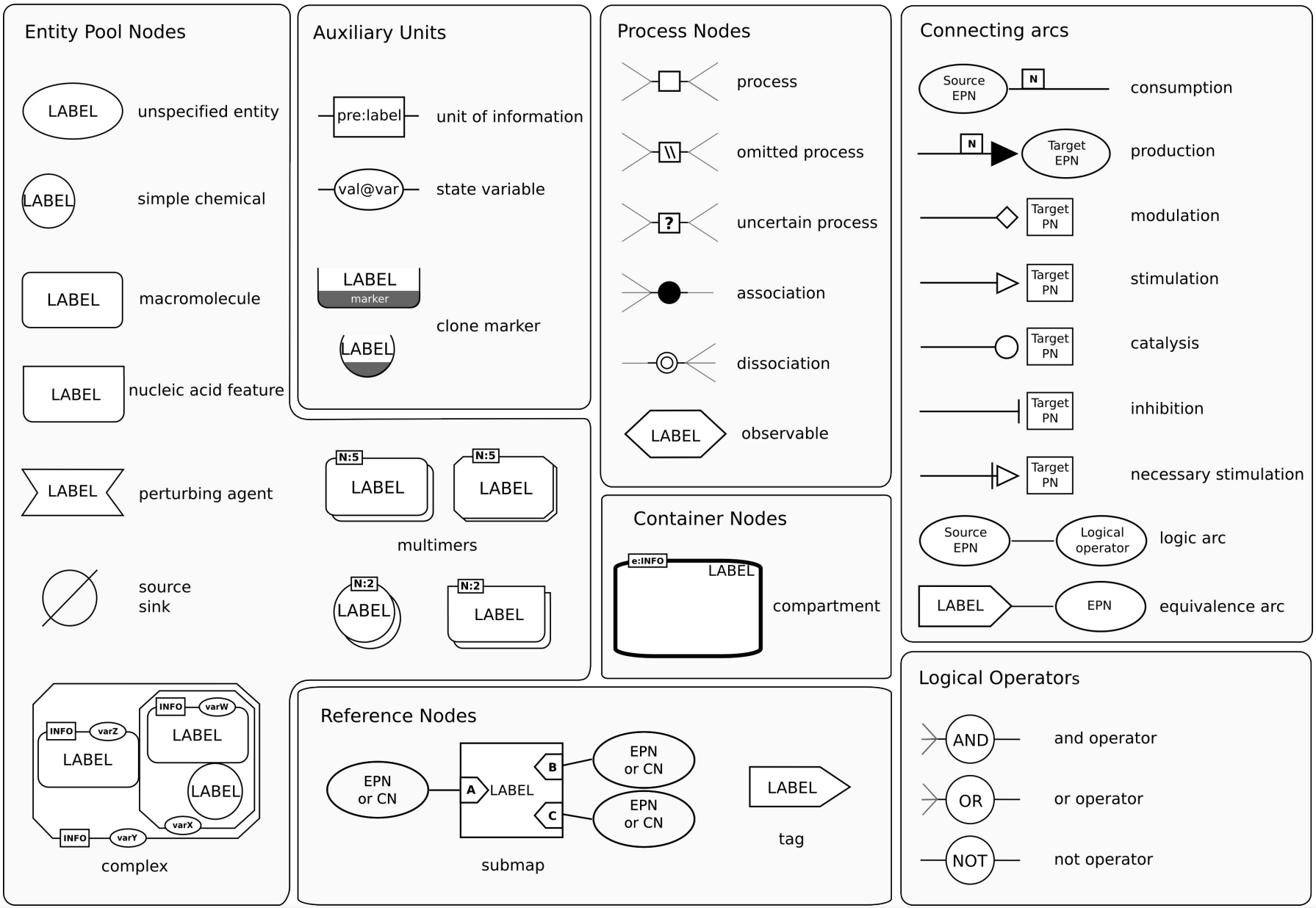
Process Diagrams are bipartite graphs



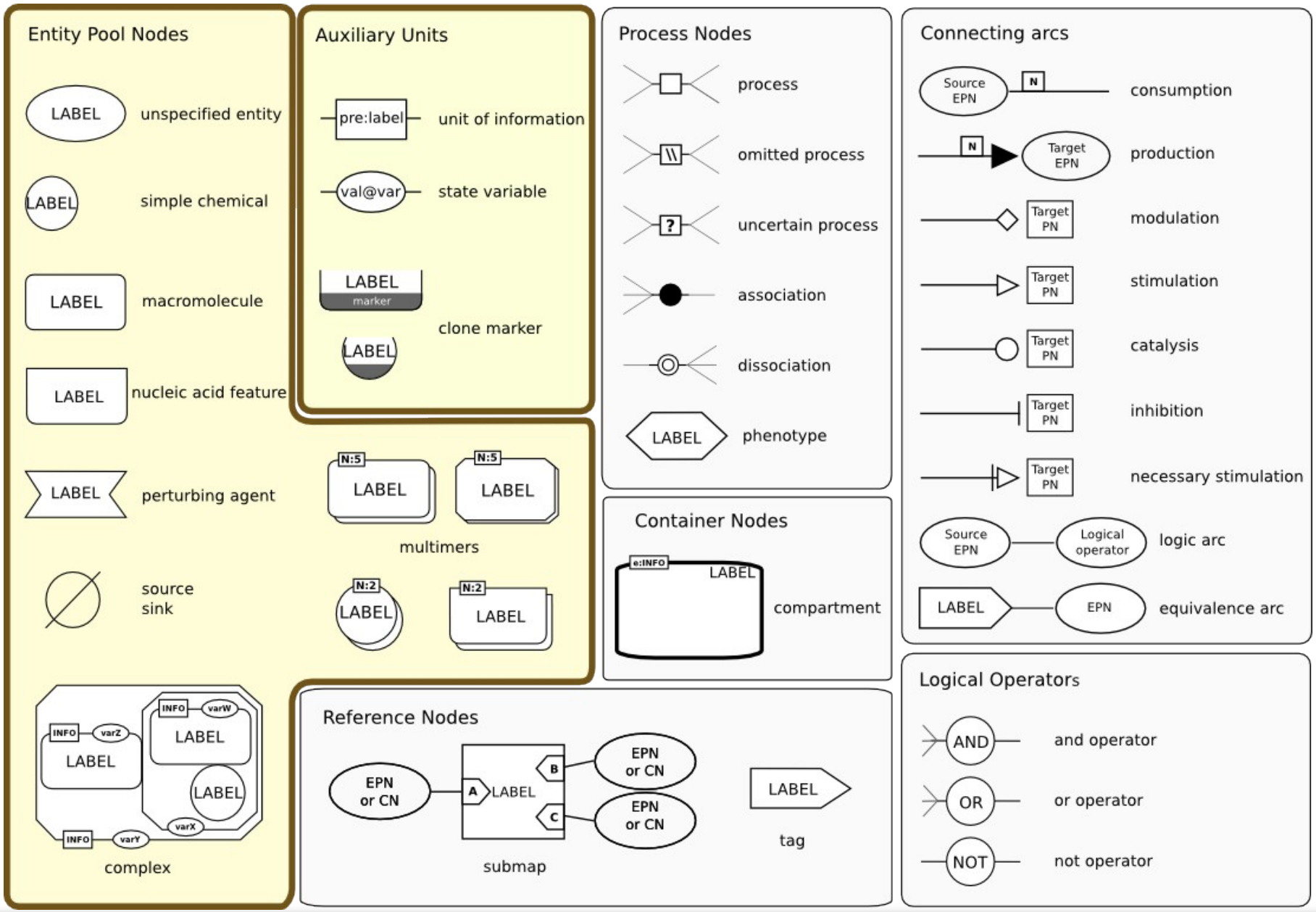
Process Diagrams can be viewed as pipelines



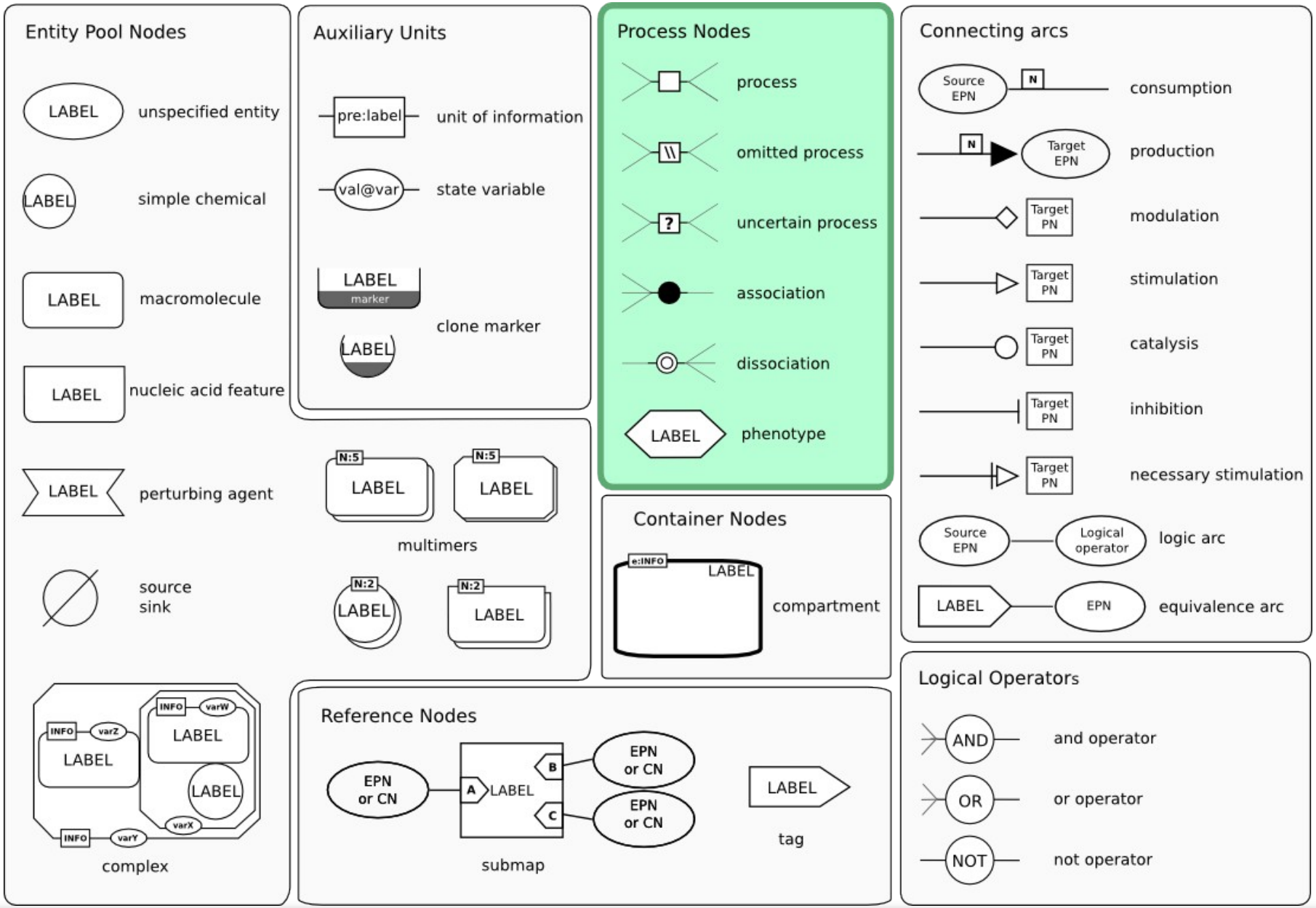
SBGN Process Diagram L1 reference card



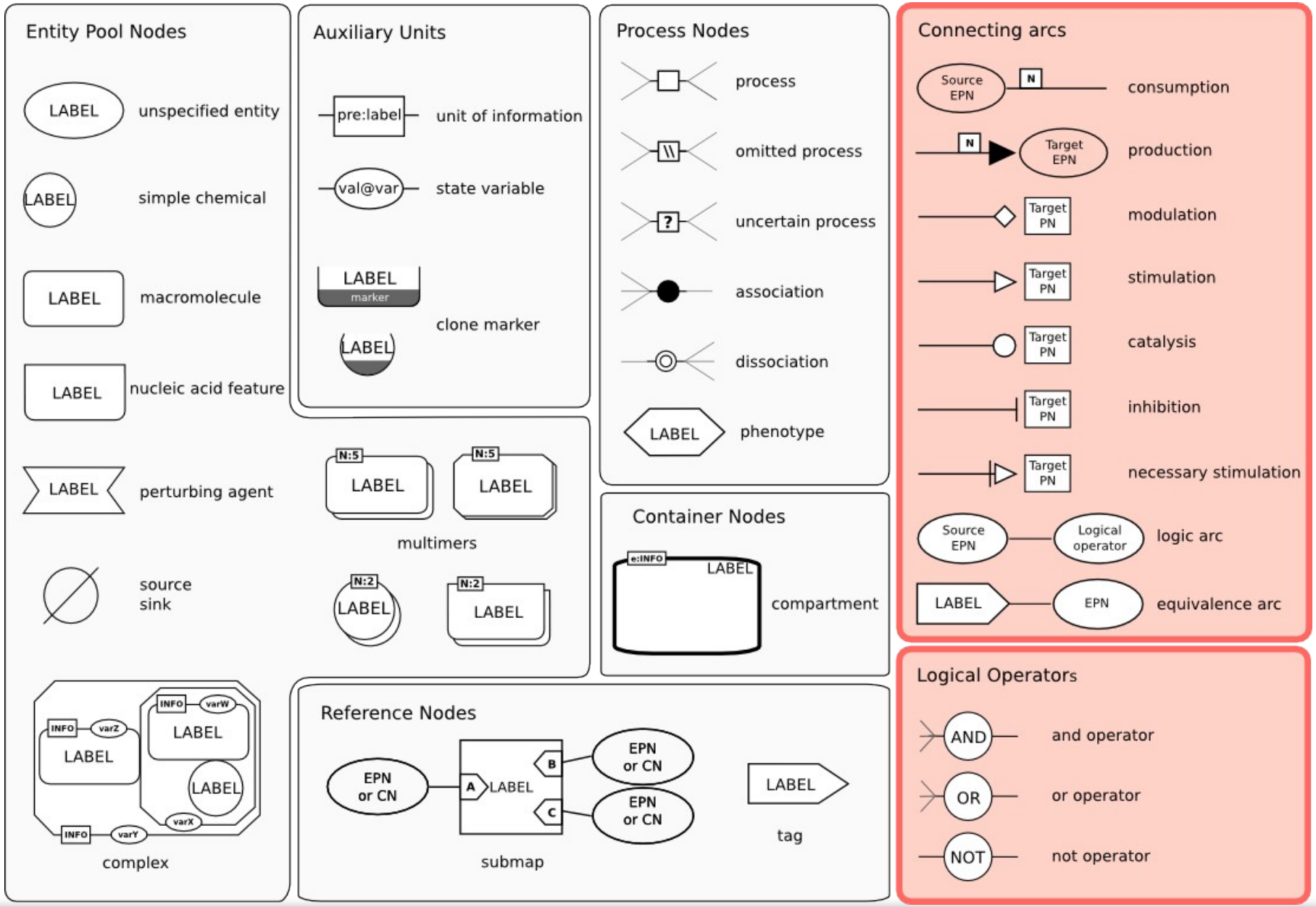
Entity Pool Nodes



Process Nodes



Edges

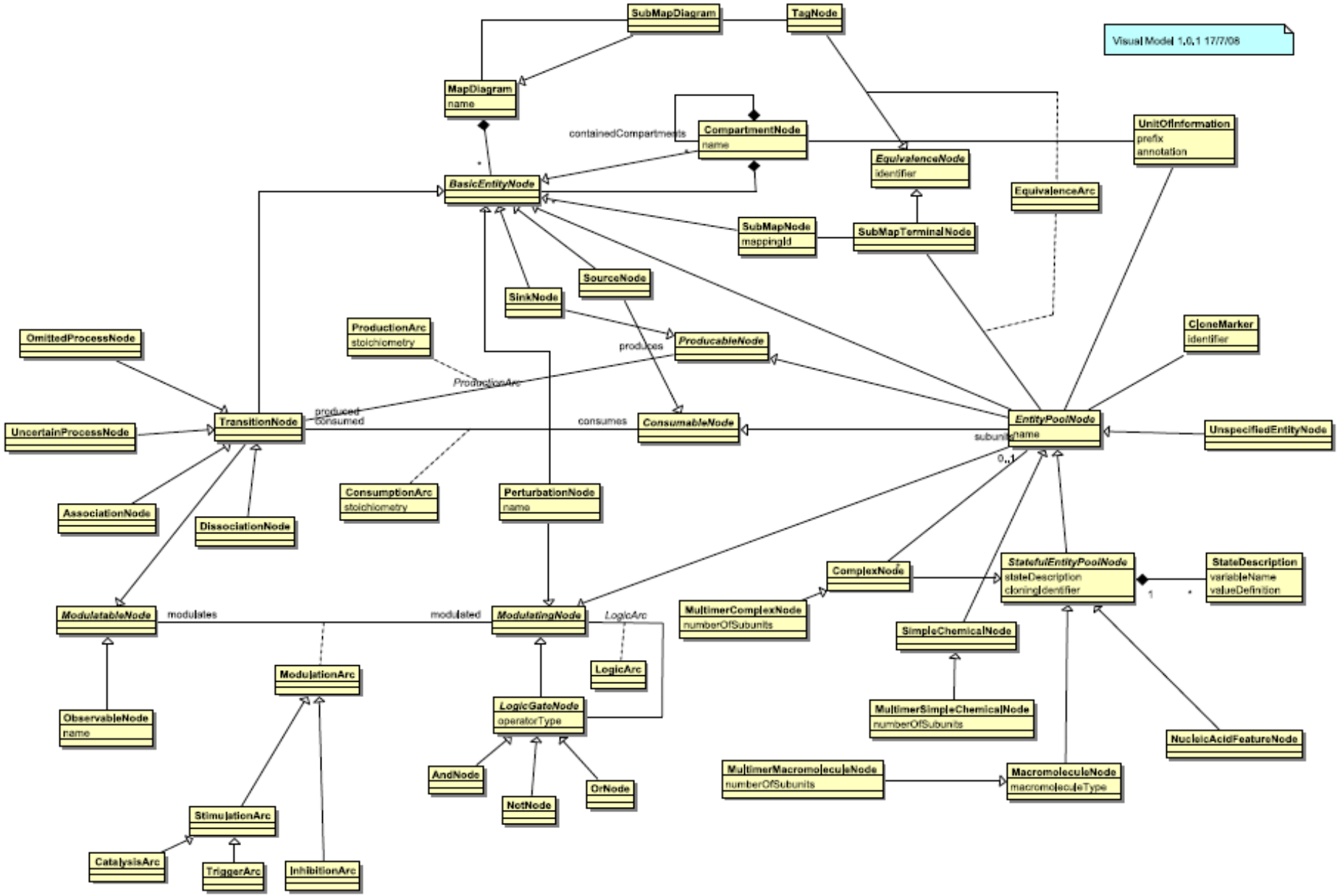

Entity Pool Nodes
LABEL
unspecified entity
LABEL
simple chemical
LABEL
macromolecule
LABEL
nucleic acid feature
LABEL
perturbing agent
source
sink
complex
Auxiliary Units
pre:label
unit of information
val@var
state variable
LABEL
marker
clone marker
LABEL
N:5
LABEL
multimers
N:2
LABEL
Process Nodes
process
omitted process
?
uncertain process
association
dissociation
LABEL
phenotype
Container Nodes
e:INFO
LABEL
compartment
Reference Nodes
EPN
or CN
A
LABEL
B
C
submap
LABEL
tag
Connecting arcs
Source
EPN
N
consumption
Target
EPN
production
Target
PN
modulation
stimulation
catalysis
inhibition
necessary stimulation
Logical
operator
logic arc
EPN
equivalence arc
Logical Operators
AND
and operator
OR
or operator
NOT
not operator

Syntax definition

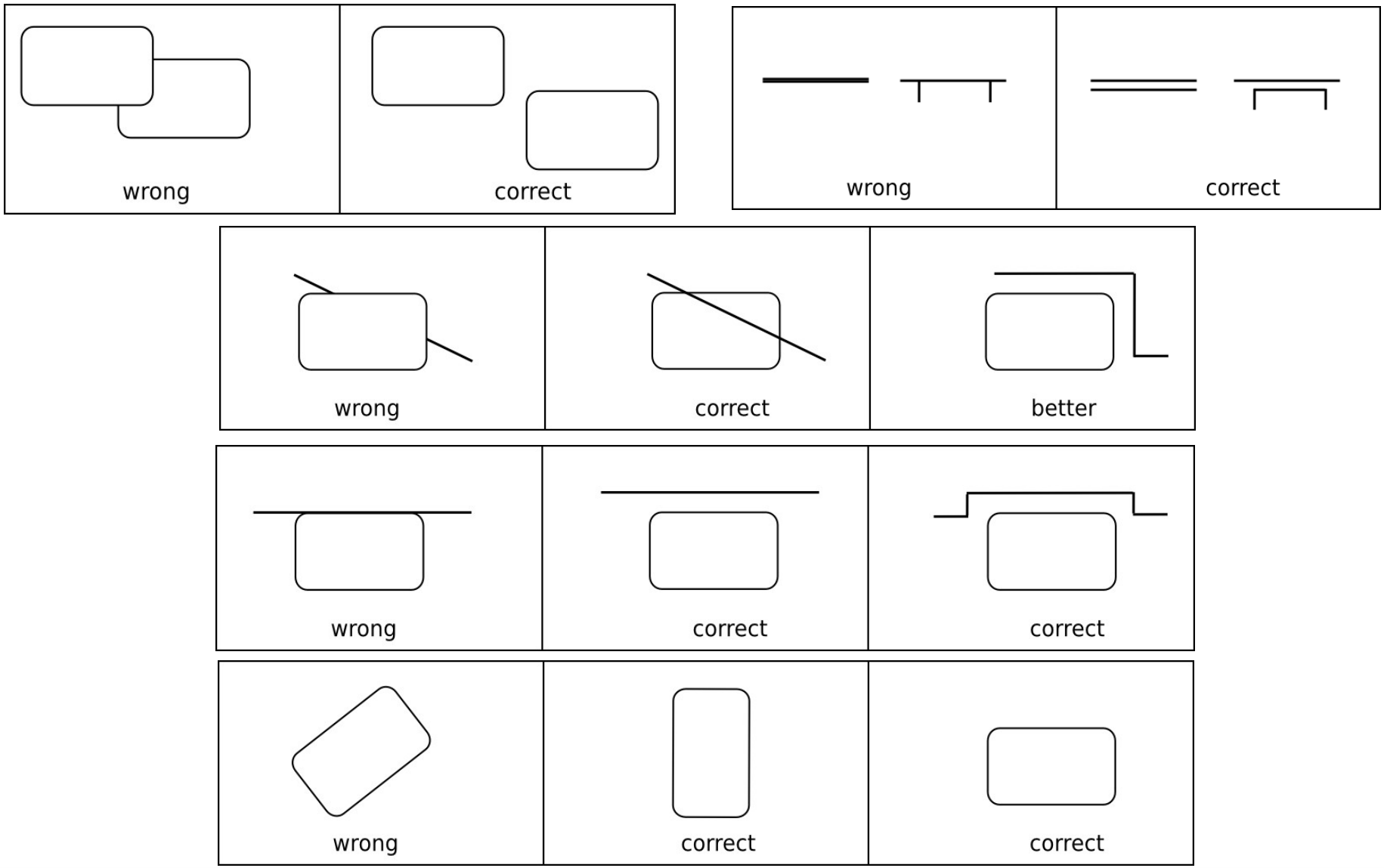
Arc\EPN	*macromolecule*	*simple chemical*	*unspecified entity*	*multimer*	*complex*	*nucleic acid feature*	*tag*	*source/sink*	*perturbation*	*observable*	*submap*
consumption	I	I	I	I	I	I		I			
production	O	O	O	O	O	O		O			
modulation	I	I	I	I	I	I			I	O	
stimulation	I	I	I	I	I	I			I	O	
catalysis	I	I	I	I	I				I	O	
inhibition	I	I	I	I	I	I			I	O	
trigger	I	I	I	I	I	I			I	O	
logic arc	I	I	I	I	I	I					
equivalence arc	I	I	I	I	I	I	O				O

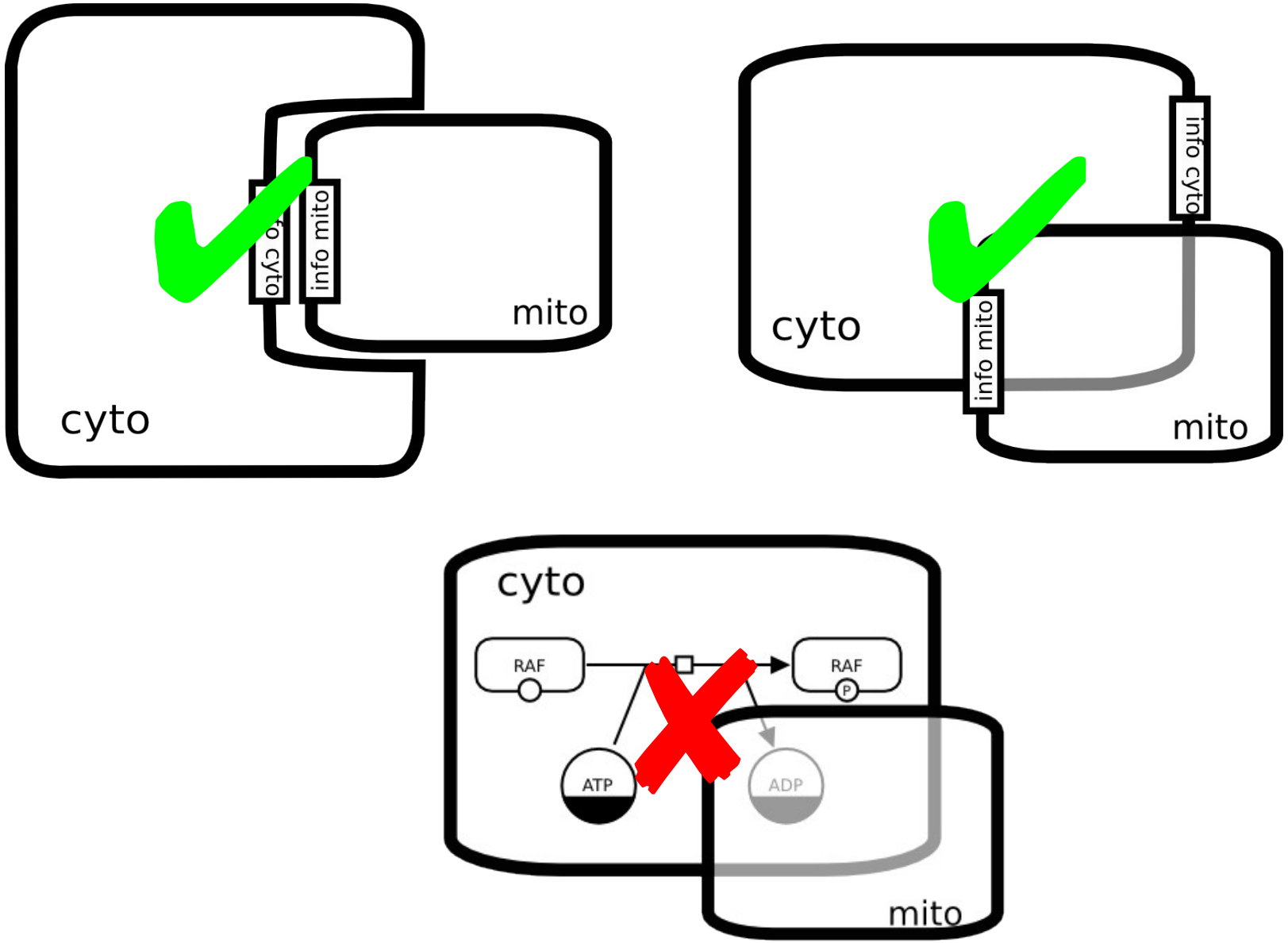
Arc\PN	*transition*	*omitted process*	*uncertain process*	*association*	*dissociation*	*and*	*or*	*not*
consumption	O	O	O	O	O(1)			
production	I	I	I	I(1)	I			
modulation	O	O	O			I(1)	I(1)	I(1)
stimulation	O	O	O			I(1)	I(1)	I(1)
catalysis	O	O	O			I(1)	I(1)	I(1)
inhibition	O	O	O			I(1)	I(1)	I(1)
trigger	O	O	O			I(1)	I(1)	I(1)
logic arc						O	O	O(1)
equivalence arc								

SBGN PD data model

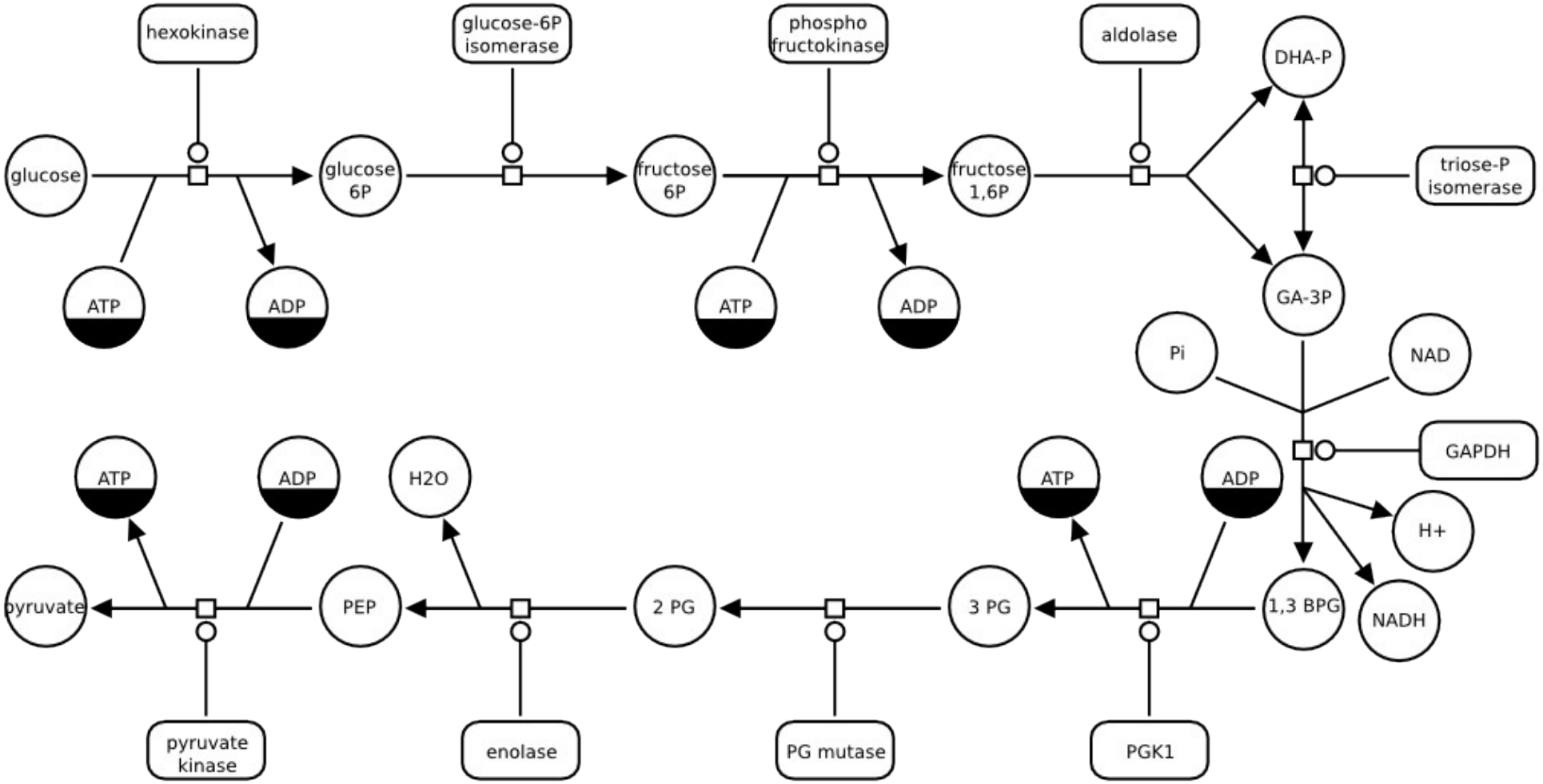


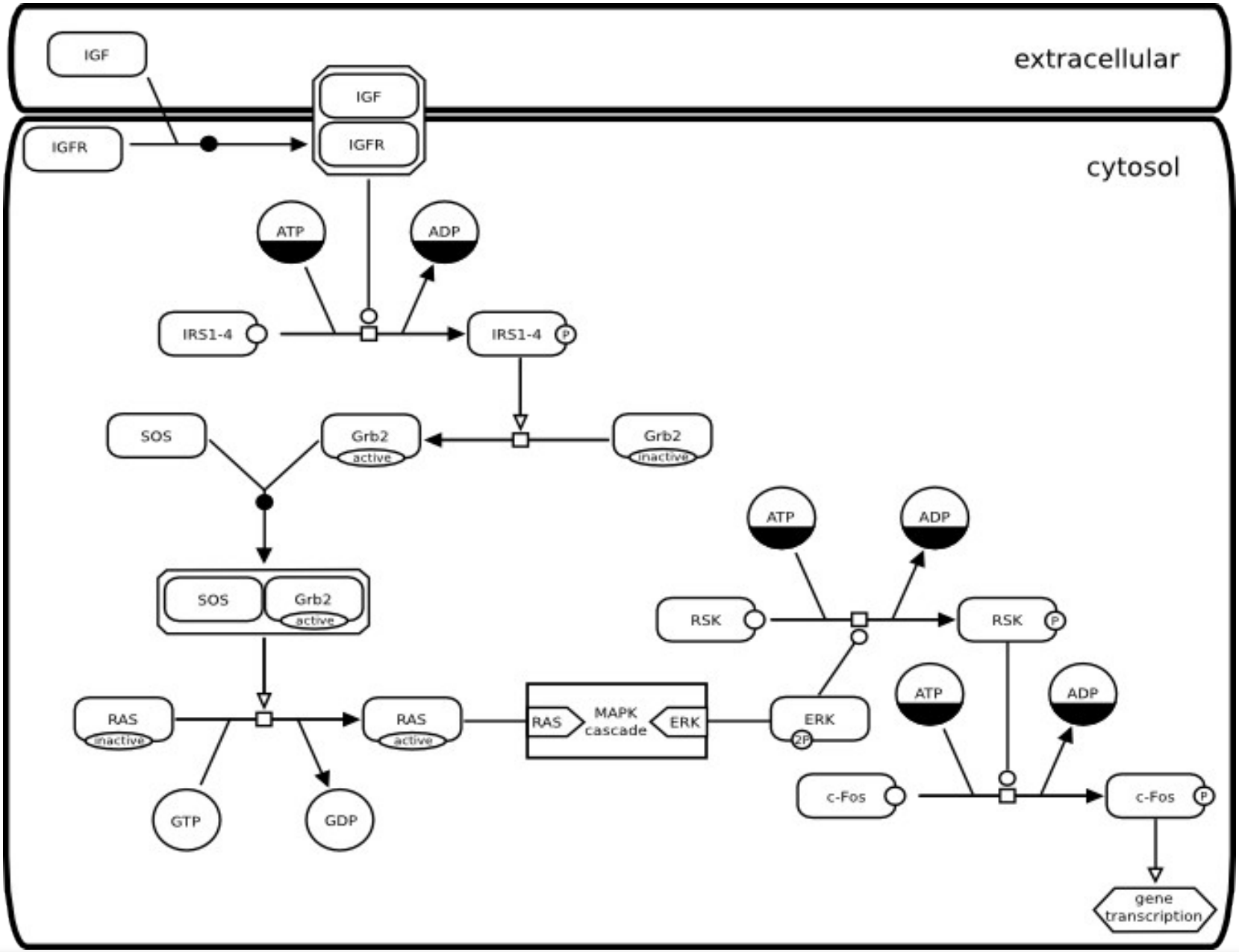
Layout constraints



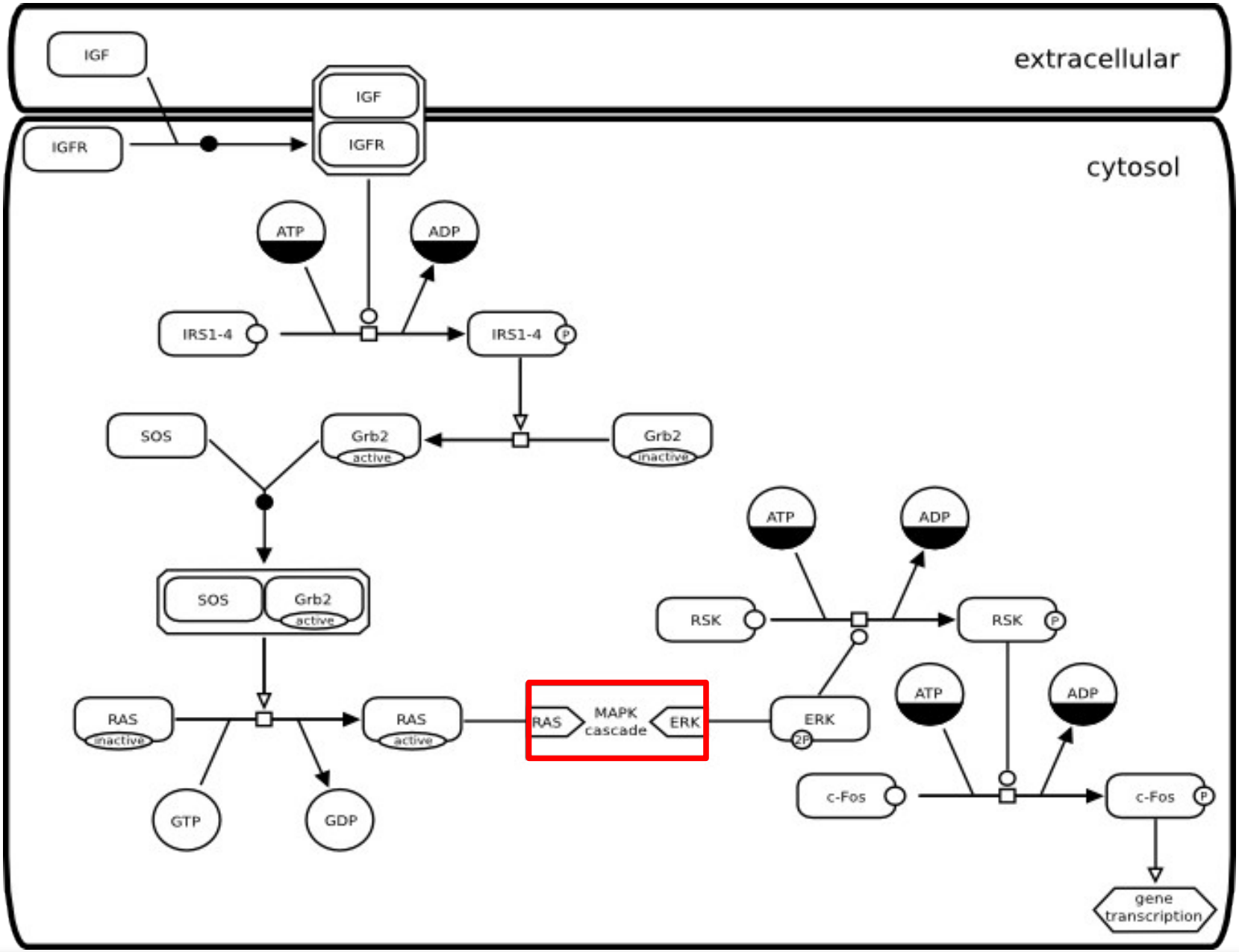

info cyto
info mito
cyto
mito
info cyto
cyto
info mito
mito
cyto
RAF
RAF
P
ATP
ADP
mito

Metabolic network

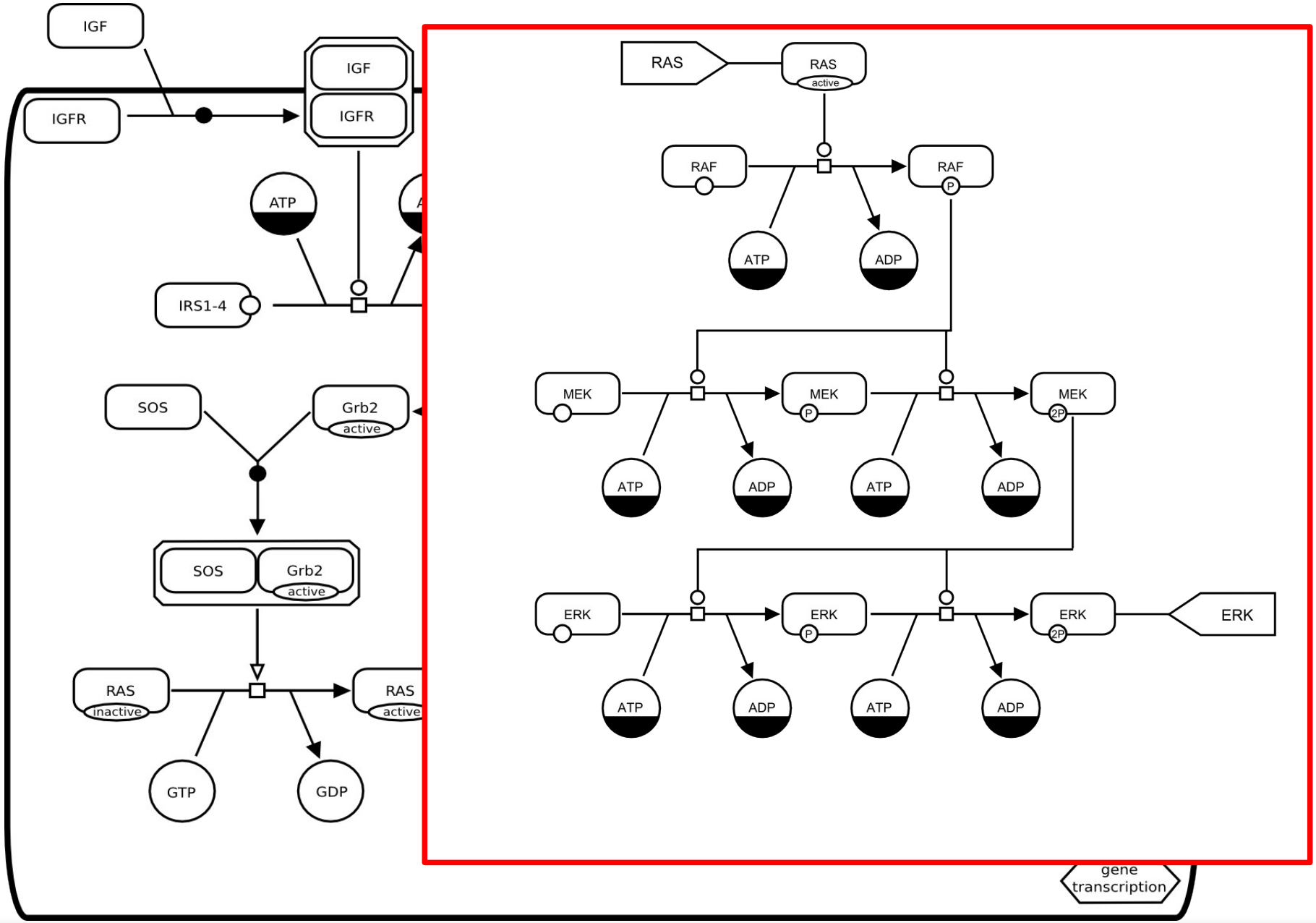




Signalling pathways

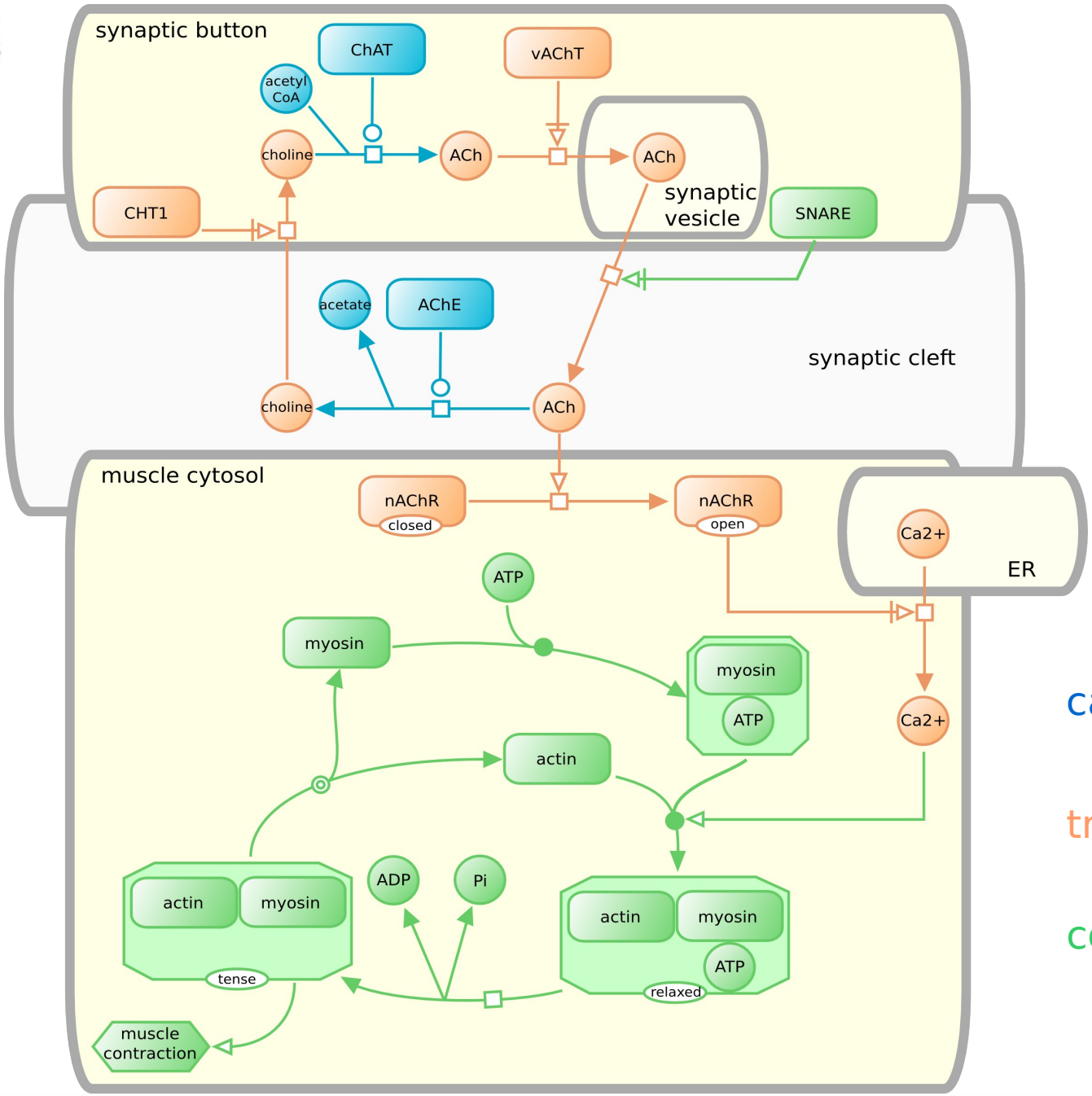

IGF
extracellular
IGF
IGFR
IGFR
cytosol
ATP
ADP
IRS1-4
IRS1-4
P
SOS
Grb2
active
Grb2
inactive
SOS
Grb2
active
ATP
ADP
RSK
RSK
P
RAS
inactive
RAS
active
RAS
MAPK
cascade
ERK
ERK
2P
ATP
ADP
GTP
GDP
c-Fos
c-Fos
P
gene
transcription

Submaps


IGF
IGFR
IGF
IGFR
ATP
IRS1-4
SOS
Grb2
active
SOS
Grb2
active
RAS
inactive
RAS
active
GTP
GDP
RAS
RAS
active
RAF
RAF
P
ATP
ADP
MEK
MEK
P
MEK
2P
ATP
ADP
ATP
ADP
ERK
ERK
P
ERK
2P
ERK
ATP
ADP
ATP
ADP
gene
transcription

Submaps

multi-cellular processes

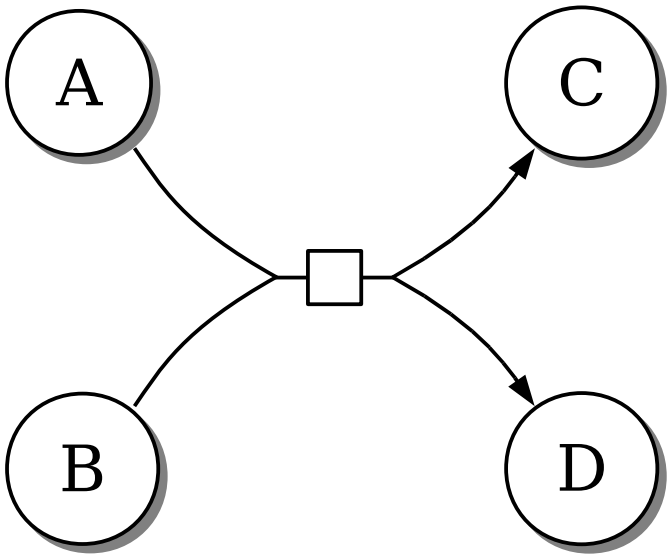


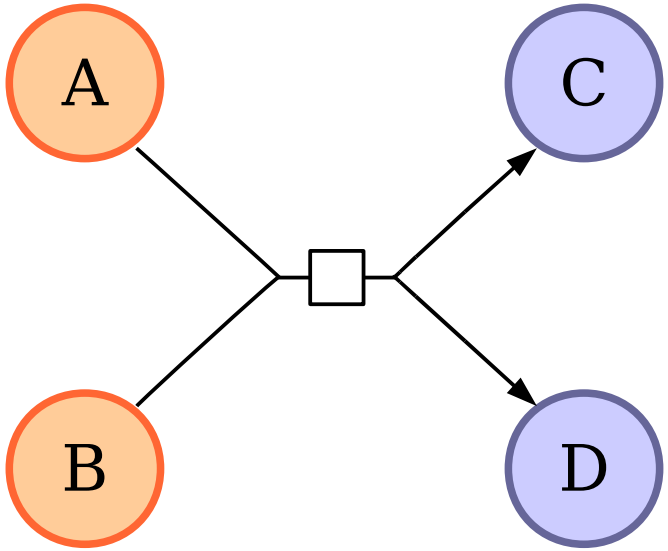
catalytic processes

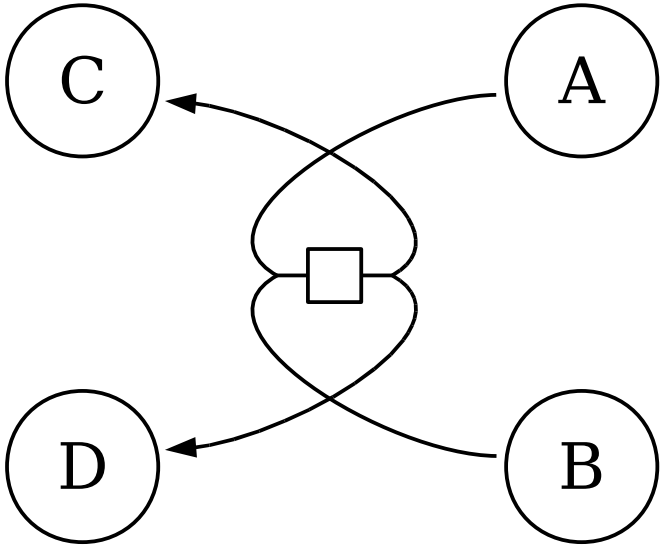
transport processes

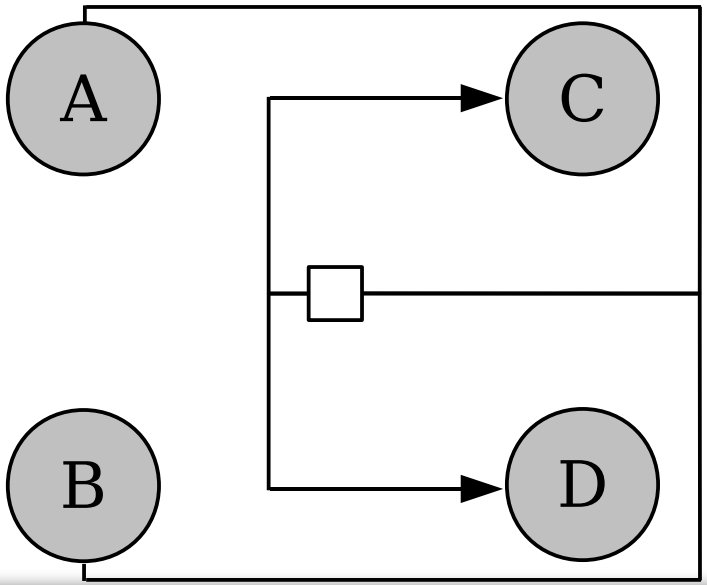
contractile proteins

All those diagrams are identical

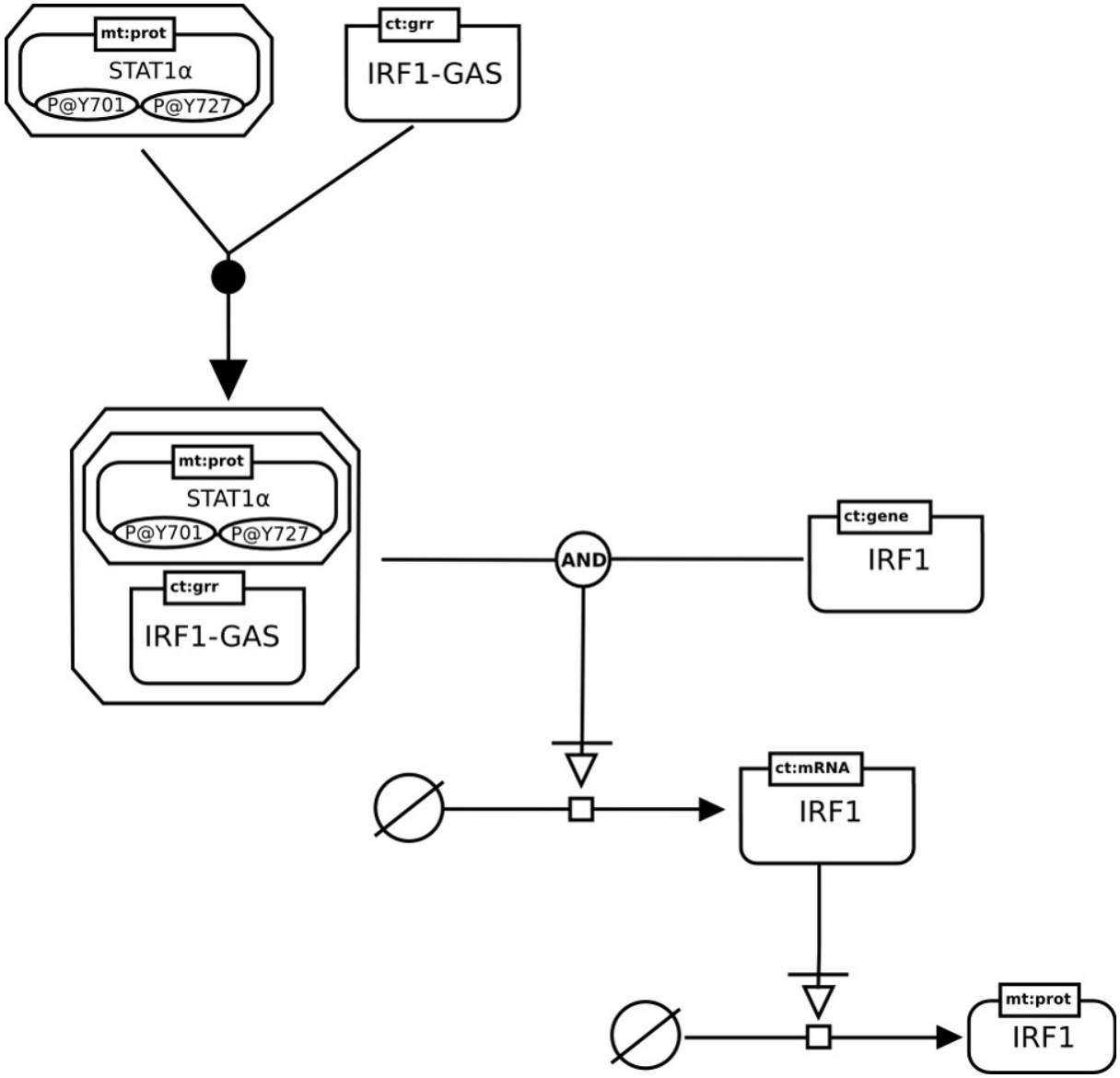








Genetic regulation



Pending issues

- Generics (entity pools representing several possible biochemical types)
- Trans-compartment (e.g. transmembrane) structures
- Logical combination of state-variable values (and close-world/open-world position)
- Moving and transforming compartments
- Non-chemical entity pool nodes (“voltage”, “pH” …)

SBGN Entity Relationships L1 reference card

Entity Nodes

Interactors

LABEL — entity

outcome

LABEL — perturbing agent

Logical Operators

AND — and operator

OR — or operator

NOT — not operator

τ — delay operator

Auxiliary units

pre:label — unit of information

variable — state variable

existence

location

reference nodes

LABEL — annotation

Relationship Nodes

Statements

assignment

N — interaction

N — Non-interaction

LABEL — observable

Influence

modulation

stimulation

necessary stimulation

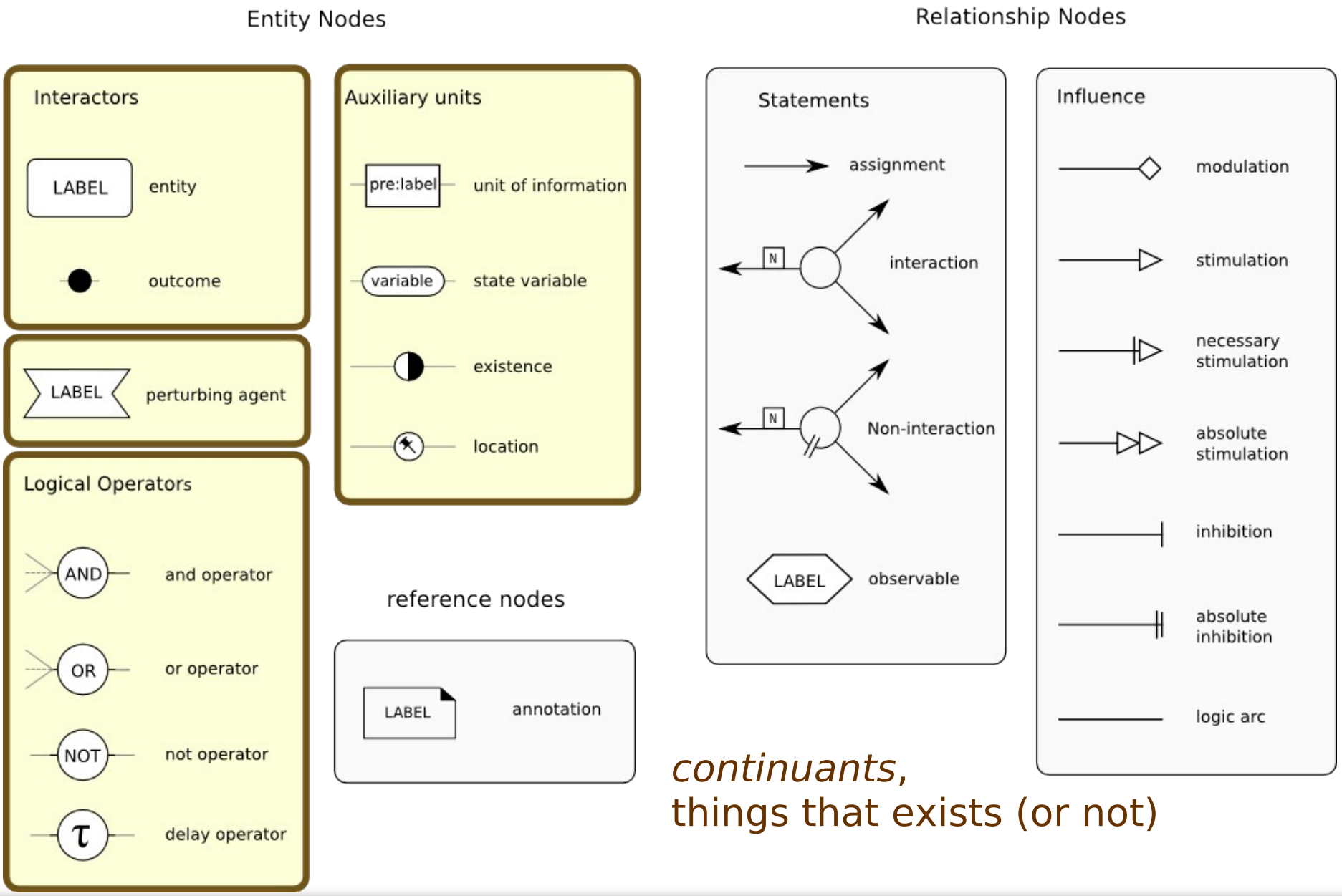
absolute stimulation

inhibition

absolute inhibition

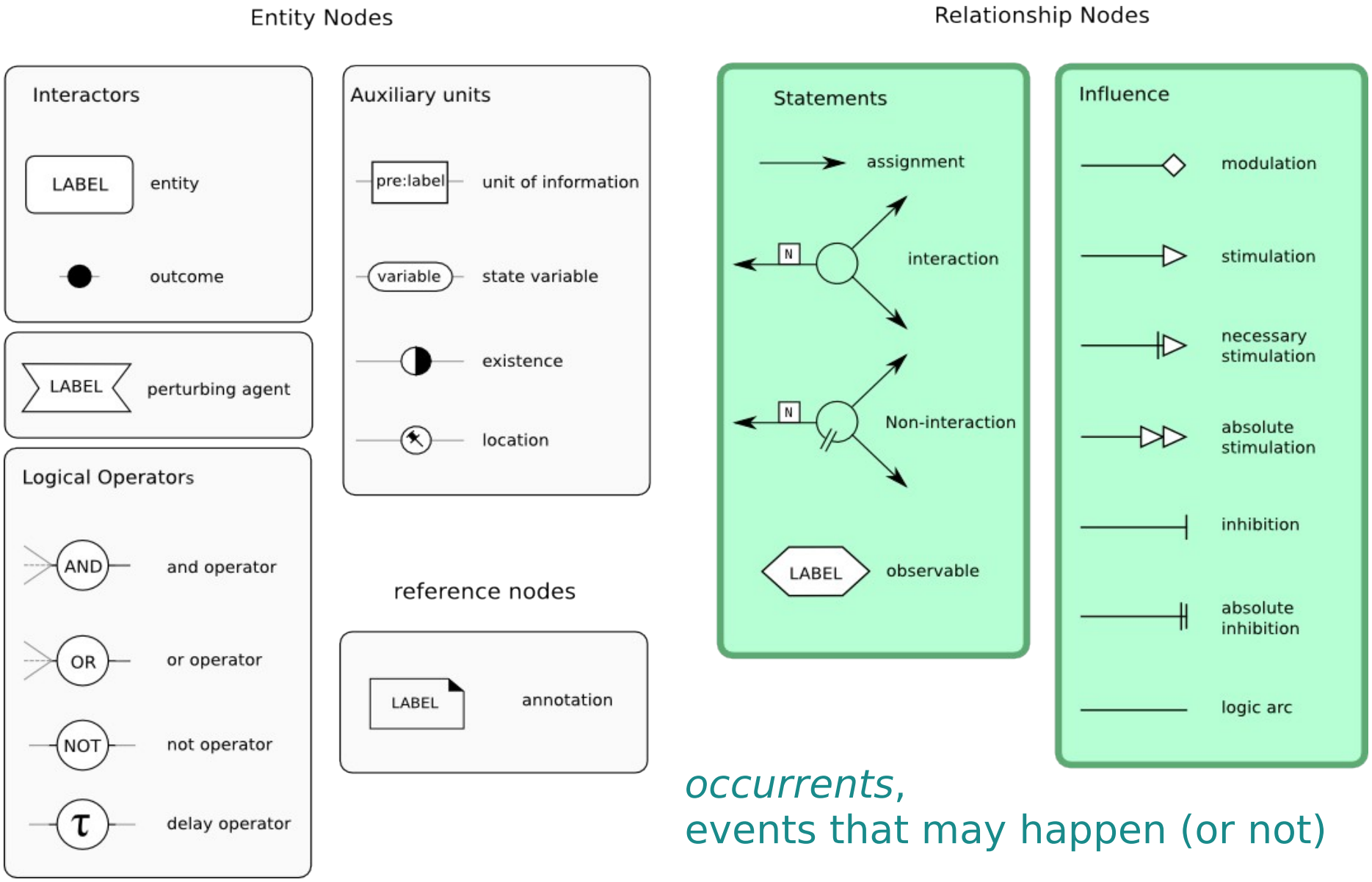
logic arc

SBGN Entity Relationships L1 reference card



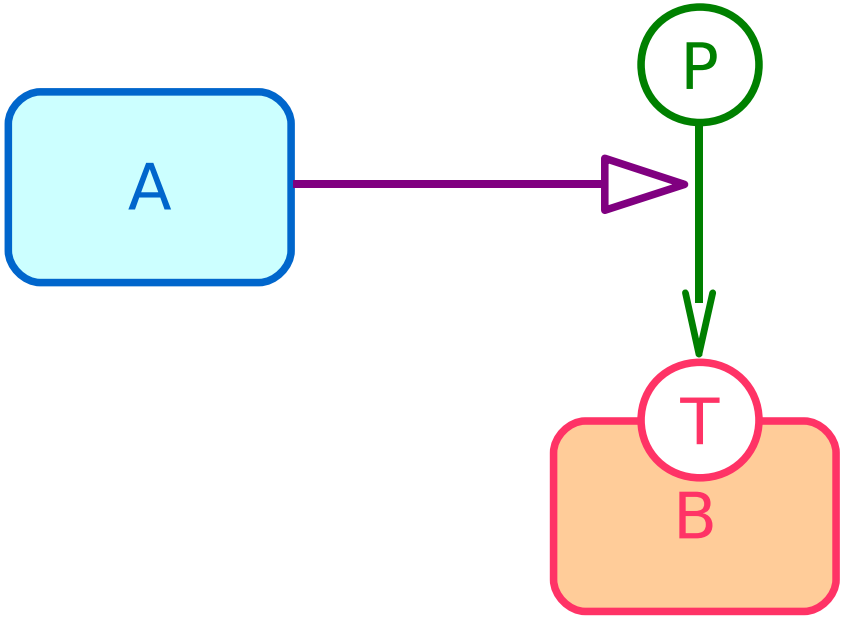
continuants,
things that exists (or not)

SBGN Entity Relationships L1 reference card

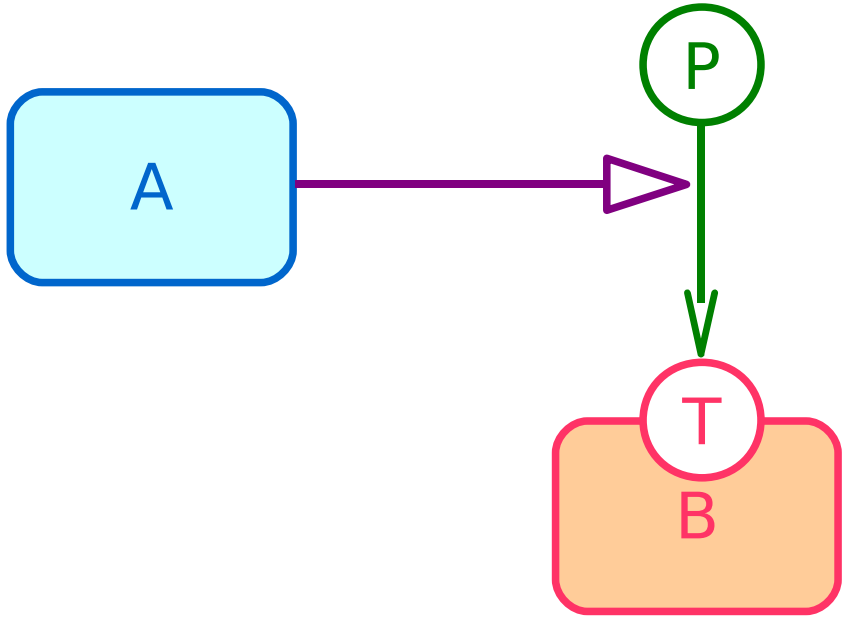


occurrents,
events that may happen (or not)

Entity Relationships can be viewed as rules

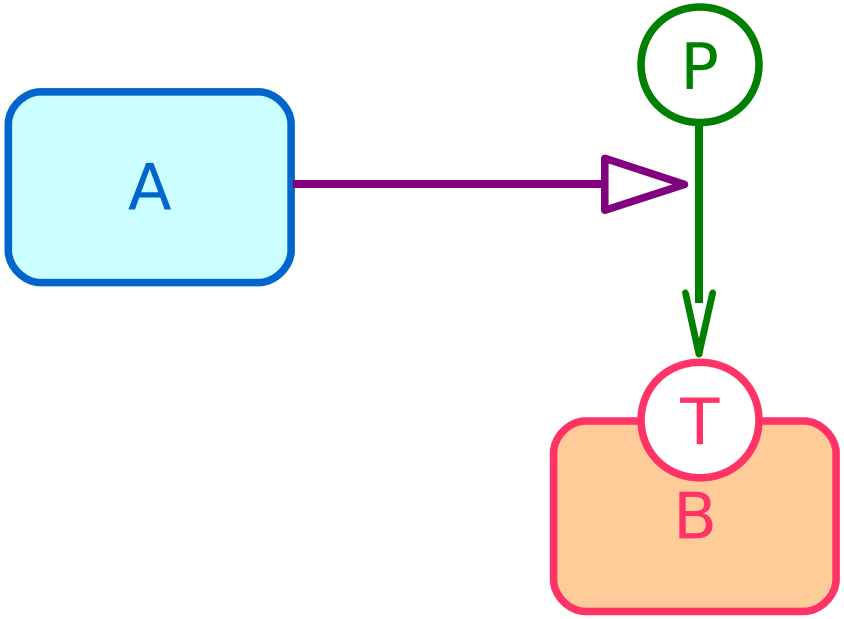


Entity Relationships can be viewed as rules



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

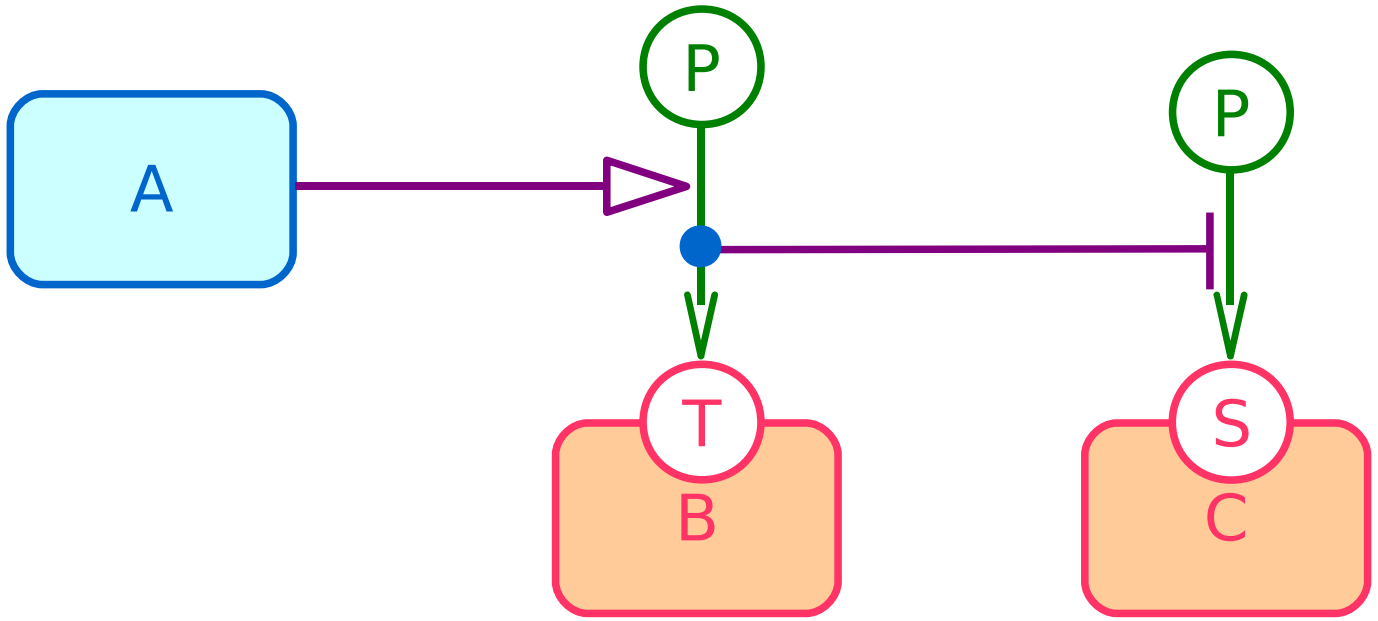
Entity Relationships can be viewed as rules



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)

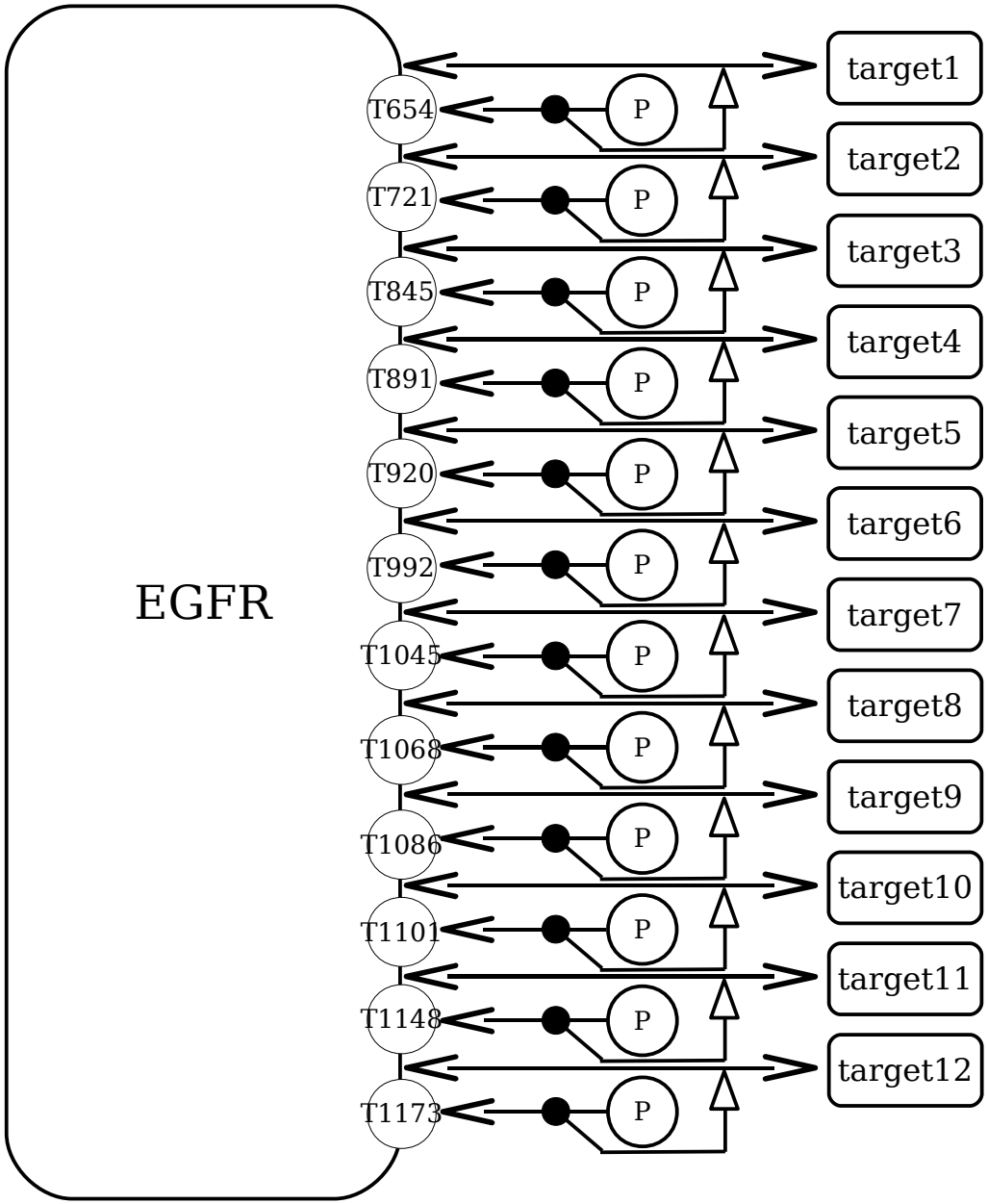
Entity Relationships can be viewed as rules



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

If P is assigned to the state variable T of B, the assignment of the value P to the state variable S of B is decreased

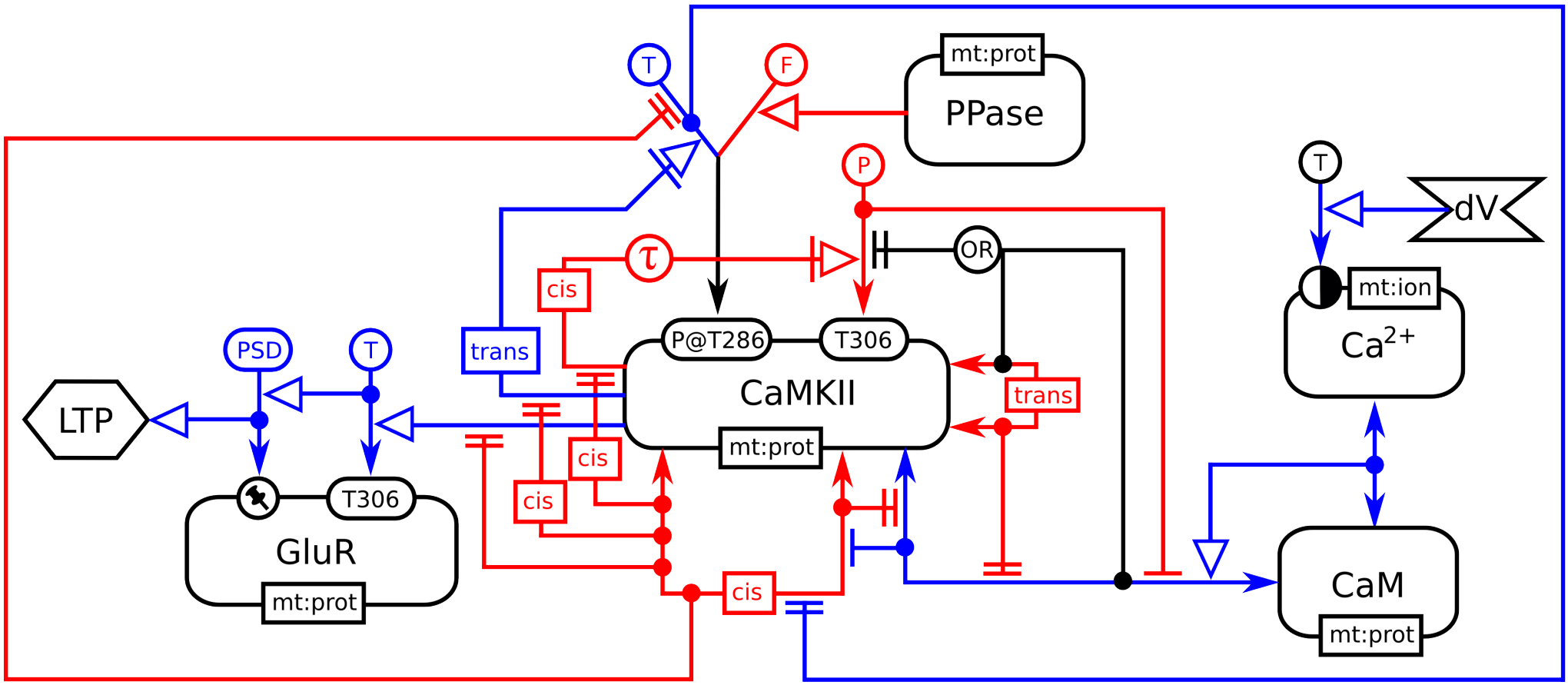
Multistate and combinatorial explosion



Process Diagram:
“once a state variable value,
always a state variable value”

2^{12} = 4096 states
(i.e. EPN glyphs) for EGFR
and 4096 complexes between
EGFR and targets

Example of Entity Relationships map



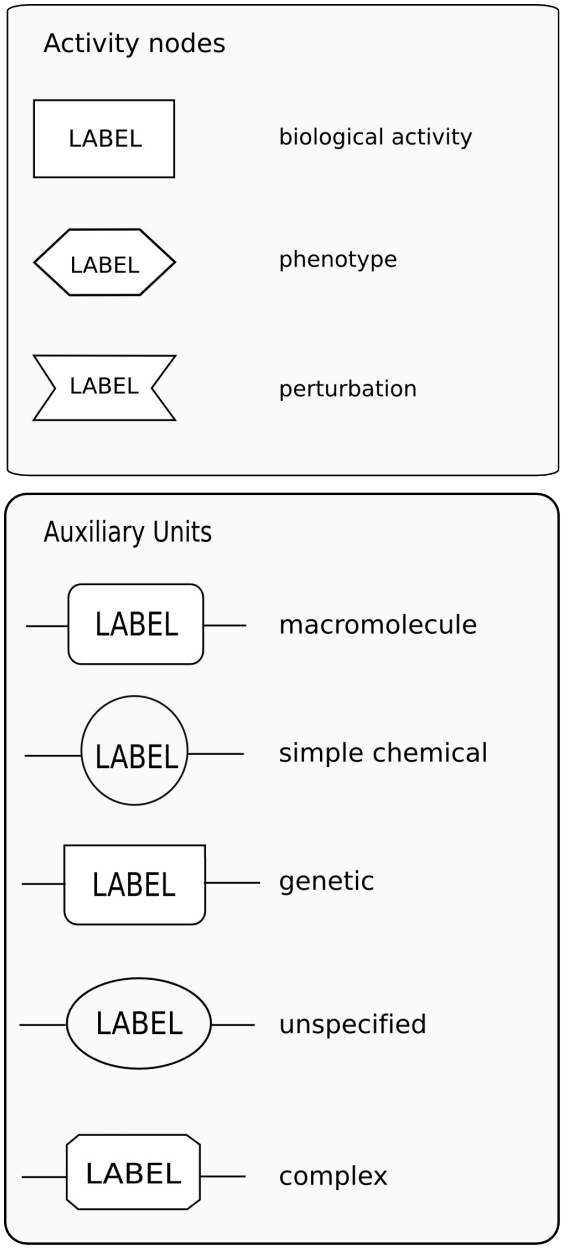
increases synaptic weight

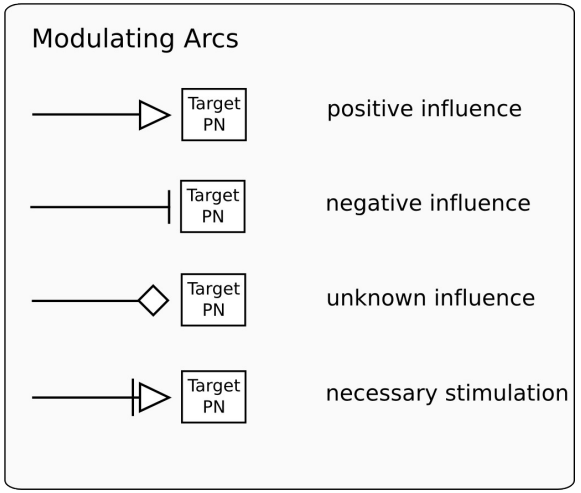
decreases synaptic weight

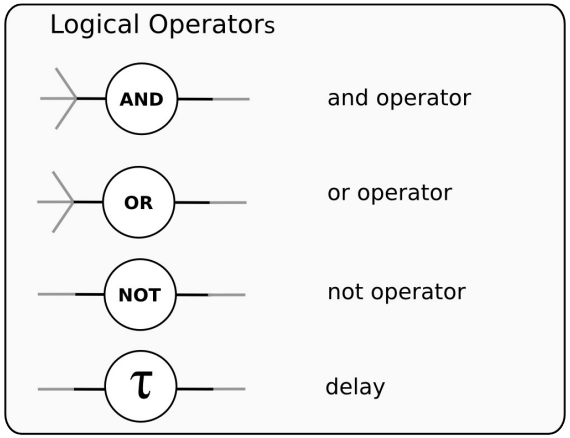
Pending issues

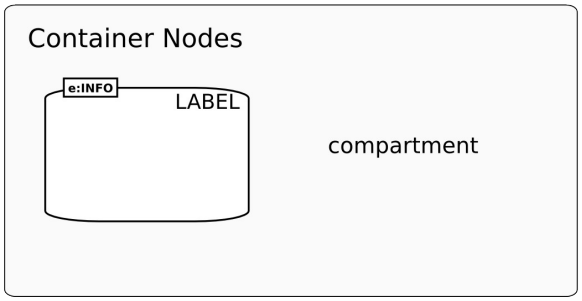
- Internal structure of entities (domains, sites, complexes)
- Identification of instances: How to differentiate between several instances of the same entity, differentially involved in a relationships?

SBGN Activity Flow L1 reference card

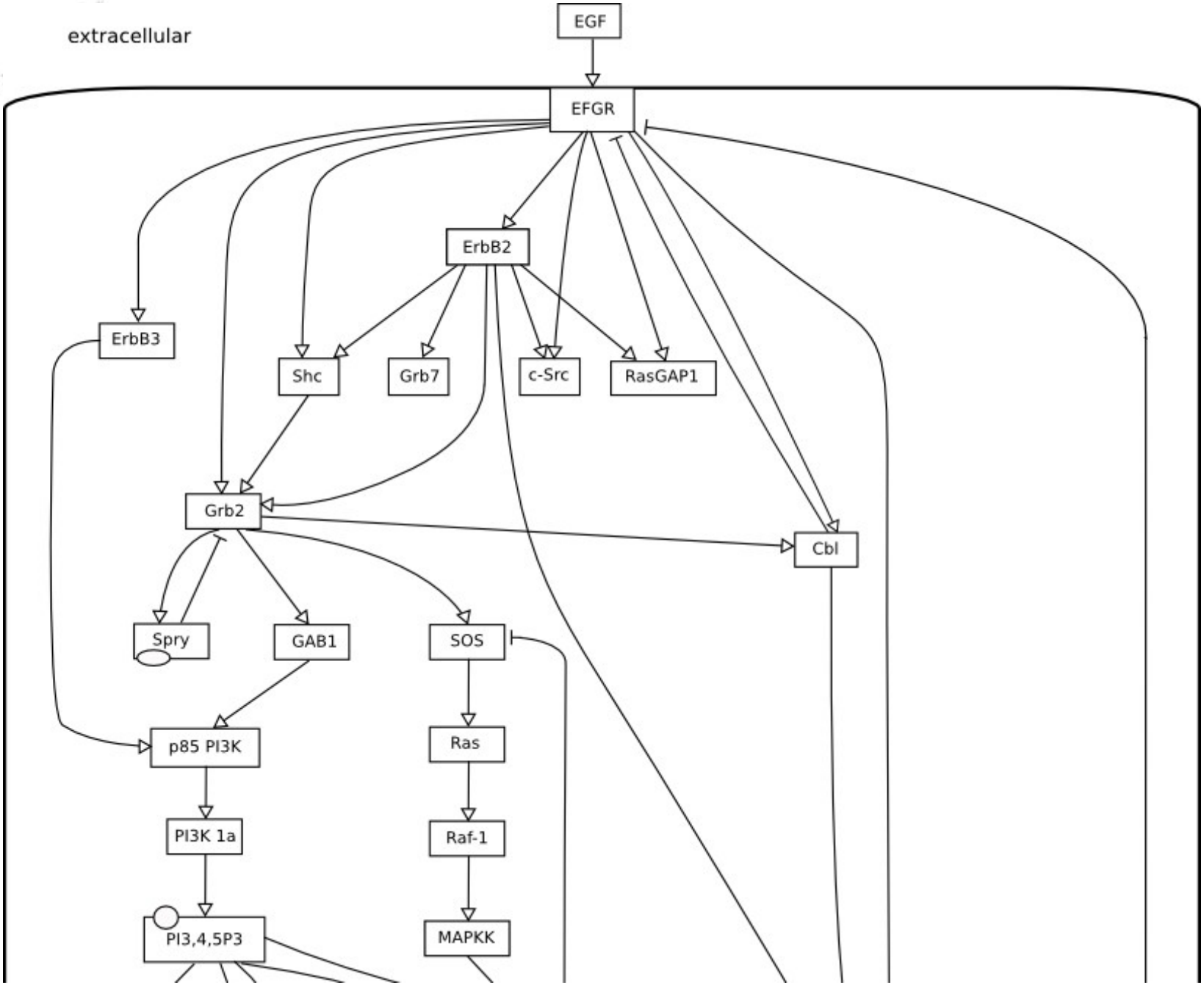


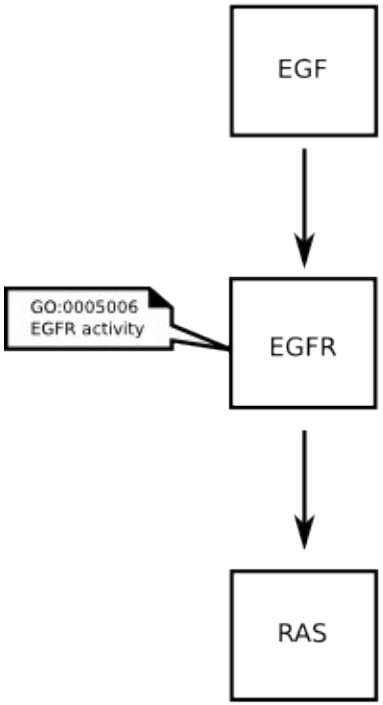


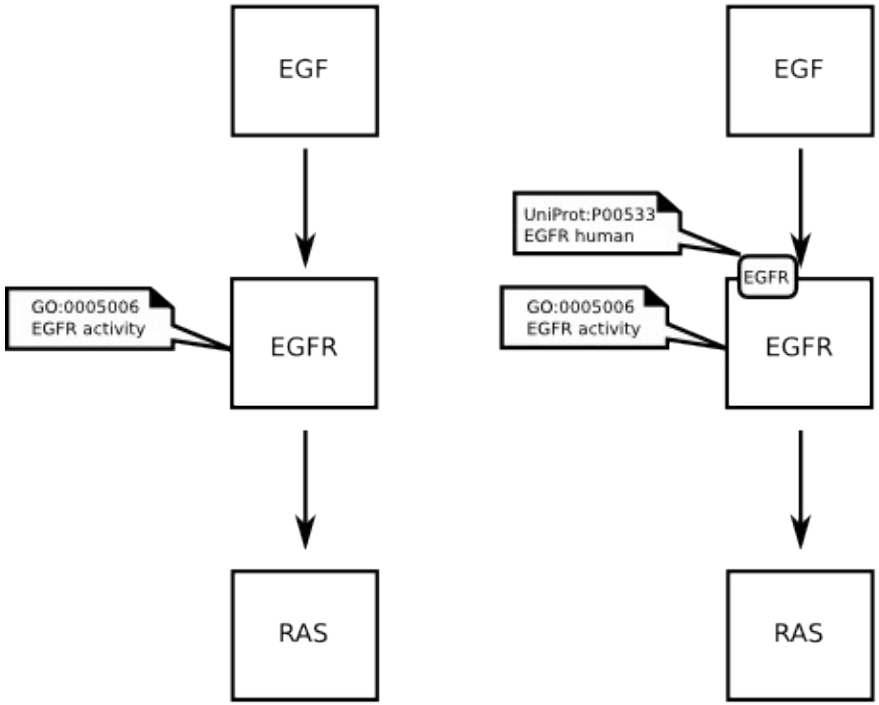


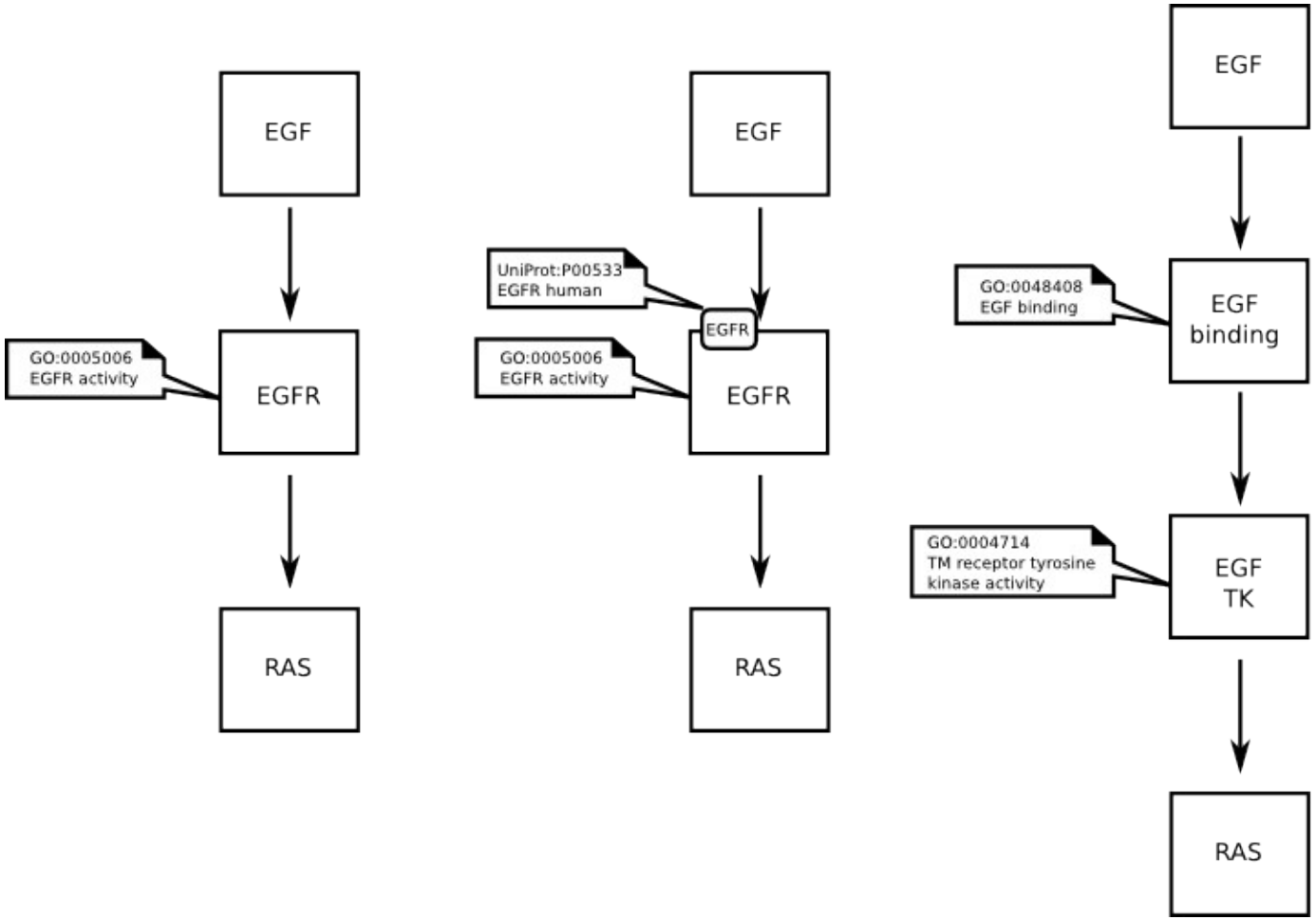


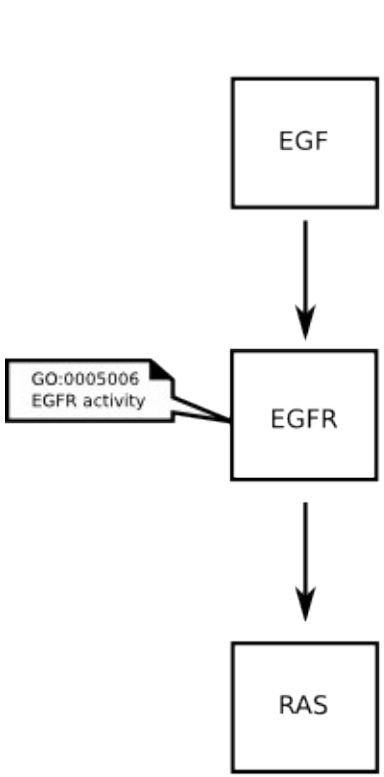
Example of Activity Flow map

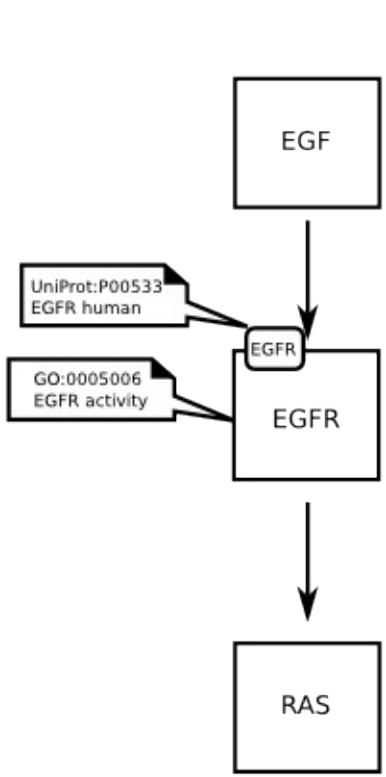


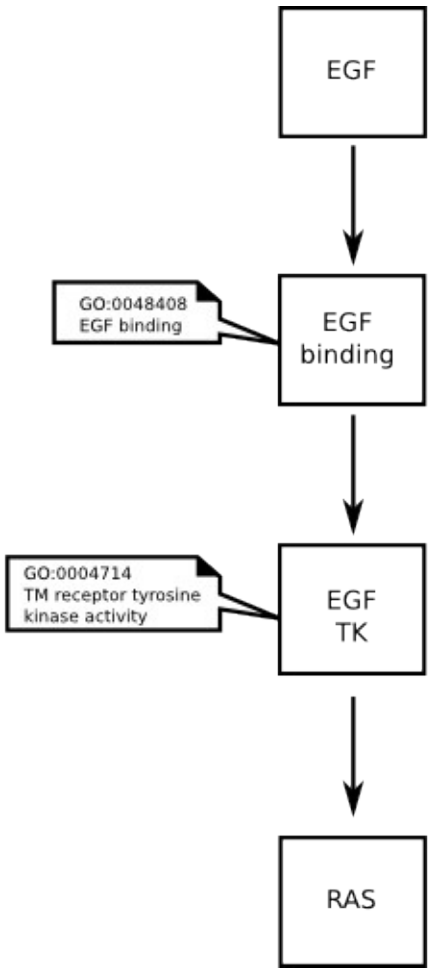

EGF
GO:0005006
EGFR activity
EGFR
RAS

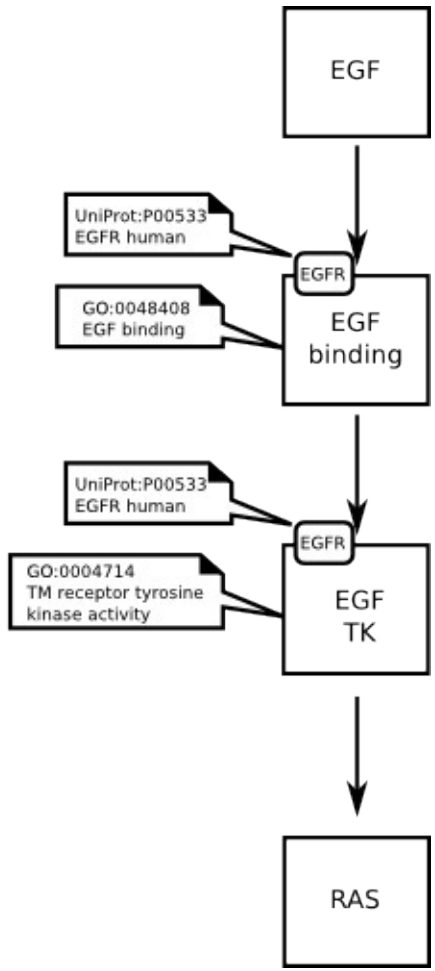

EGF
GO:0005006
EGFR activity
EGFR
RAS
EGF
UniProt:P00533
EGFR human
EGFR
GO:0005006
EGFR activity
EGFR
RAS


EGF
GO:0005006
EGFR activity
EGFR
RAS
EGF
UniProt:P00533
EGFR human
EGFR
GO:0005006
EGFR activity
EGFR
RAS
EGF
GO:0048408
EGF binding
EGF
binding
GO:0004714
TM receptor tyrosine
kinase activity
EGF
TK
RAS


EGF
GO:0005006
EGFR activity
EGFR
RAS

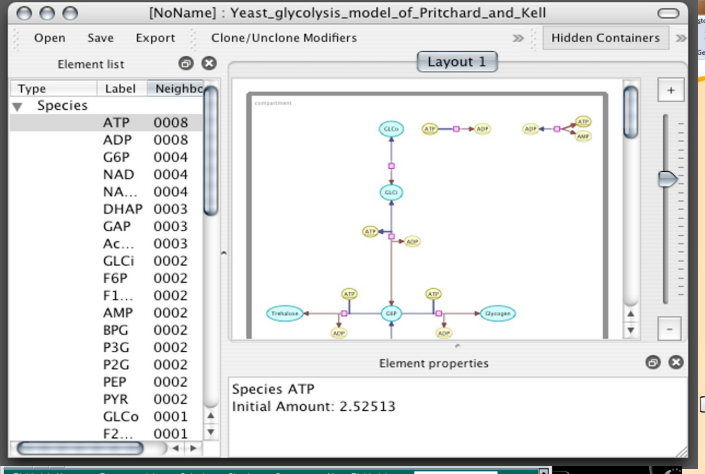

EGF
UniProt:P00533
EGFR human
EGFR
GO:0005006
EGFR activity
EGFR
RAS

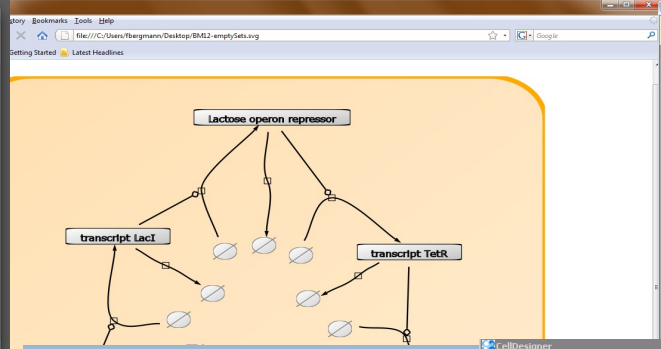

EGF
GO:0048408
EGF binding
EGF
binding
GO:0004714
TM receptor tyrosine
kinase activity
EGF
TK
RAS

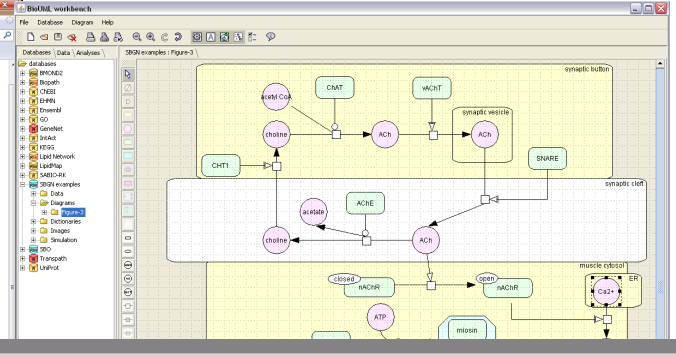

EGF
UniProt:P00533
EGFR human
EGFR
GO:0048408
EGF binding
EGF
binding
UniProt:P00533
EGFR human
EGFR
GO:0004714
TM receptor tyrosine
kinase activity
EGF
TK
RAS

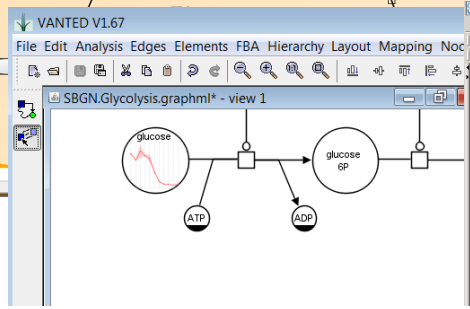
Software support for SBGN PD L1

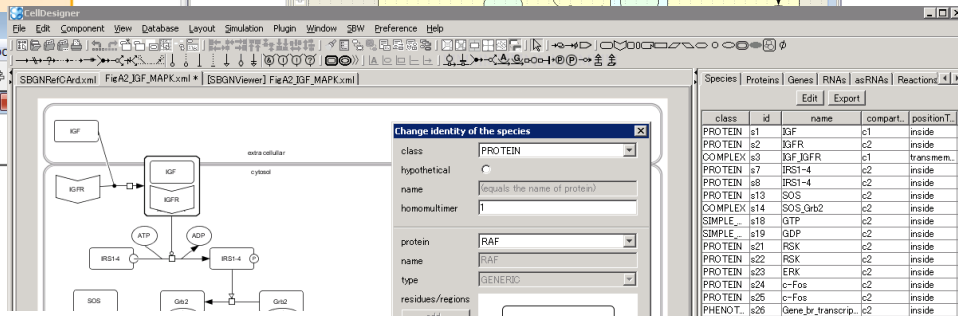
- **Arcadia** (MCISB, Manchester, UK) http://arcadiapathways.sourceforge.net/
- **Athena** (Univ Washington, Seattle, US) http://www.codeplex.com/athena/
- **BioModels DB** (EMBL-EBI, UK) http://www.ebi.ac.uk/biomodels/
- **BioUML** (Inst Systems Biology, Novosibirsk, RU) http://www.biouml.org/
- **CellDesigner** (SBI, Tokyo, JP) http://www.celldesigner.org/
- **EPE** (CISBE, Edinburgh, UK) http://www.bioinformatics.ed.ac.uk/epe/
- **JWS Online** (Stellenbosh University, ZA) http://jjj.biochem.sun.ac.za/
- **NetBuilder** (Univ Hertfords, UK) http://strc.herts.ac.uk/bio/maria/Apostrophe/
- **PANTHER** (SRI international, USA) http://www.pantherdb.org/pathway/
- **Reactome** (EMBL-EBI, UK) http://www.reactome.org/
- **Vanted** (IPK Gatersleben, DE) http://vanted.ipk-gatersleben.de/
- **VISIOweb** (Bilkent Univ, Turkey) http://www.bilkent.edu.tr/~bcbi/pvs.html

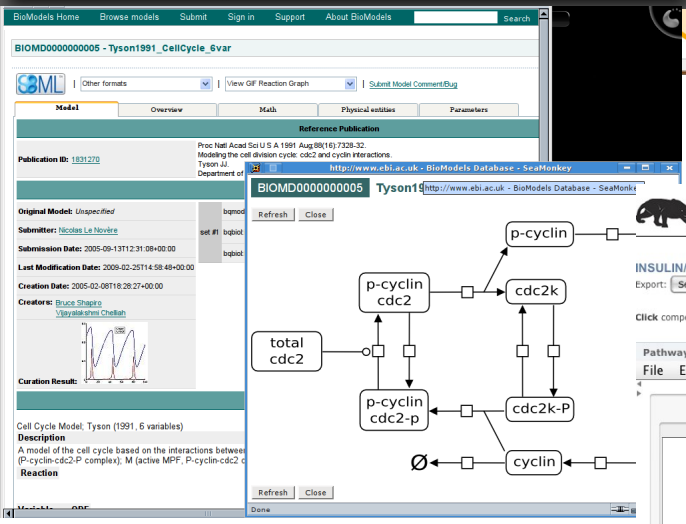

[NoName] : Yeast_glycolysis_model_of_Pritchard_and_Kell
Open
Save
Export
Clone/Unclone Modifiers
Hidden Containers
Element list
Layout 1
Type
Label
Species
ATP 0008
ADP 0008
G6P 0004
NAD 0004
DHAP 0003
GAP 0003
GLCi 0002
F6P 0002
AMP 0002
BPG 0002
P3G 0002
P2G 0002
PEP 0002
PYR 0002
GLCo 0001
Element properties
Species ATP
Initial Amount: 2.52513


Lactose operon repressor
transcript LacI
transcript TetR

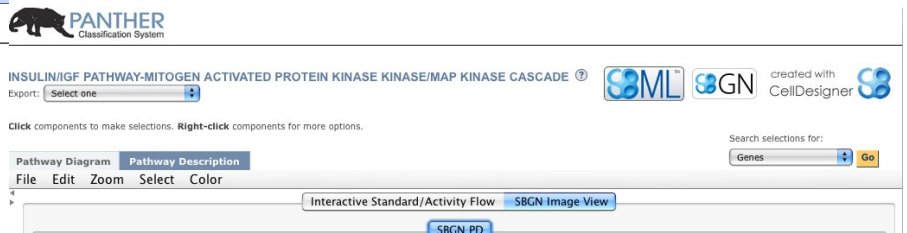

BioUML workbench
synaptic button
synaptic vesicle
synaptic cleft
muscle cytosol
ChAT
vAChT
ACh
SNARE
AChE
choline
acetate
nAChR
Ca2+
ATP
myosin

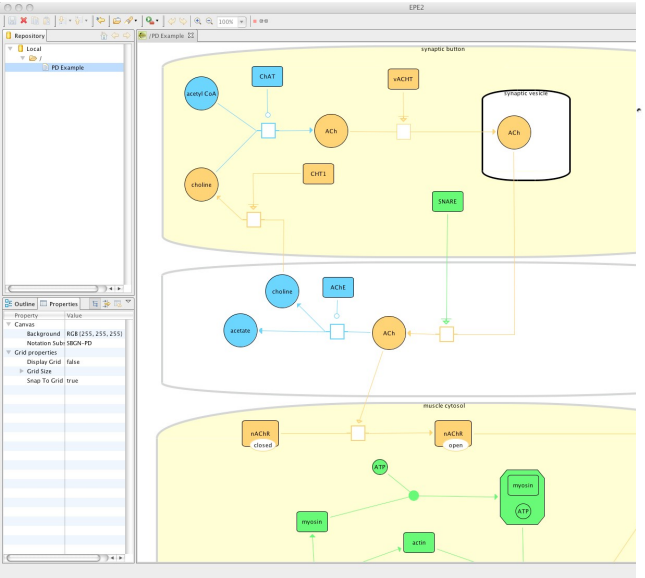

VANTED V1.67
File Edit Analysis Edges Elements FBA Hierarchy Layout Mapping
SBGN.Glycolysis.graphml* - view 1
glucose
glucose 6P
ATP
ADP

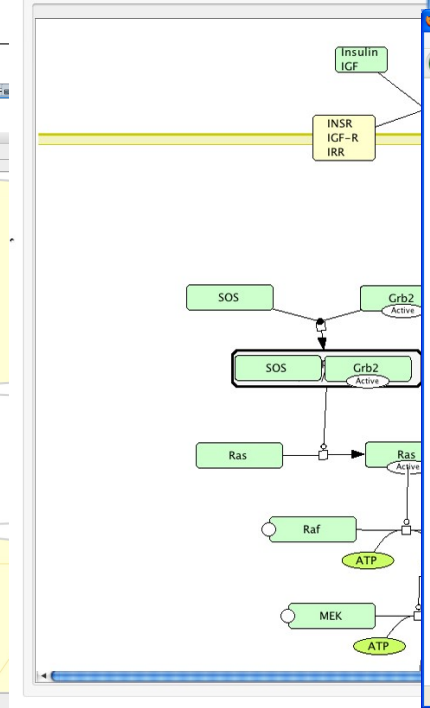

CellDesigner
File Edit Component View Database Layout Simulation Plugin Window SBW Preference Help
Change identity of the species
class
PROTEIN
hypothetical
name
homomultimer
protein
RAF
type
GENERIC
residues/regions
modification
state
empty
text input
Apply
Reset
Cancel
IGF
IGFR
IRS1-4
SOS
Grb2

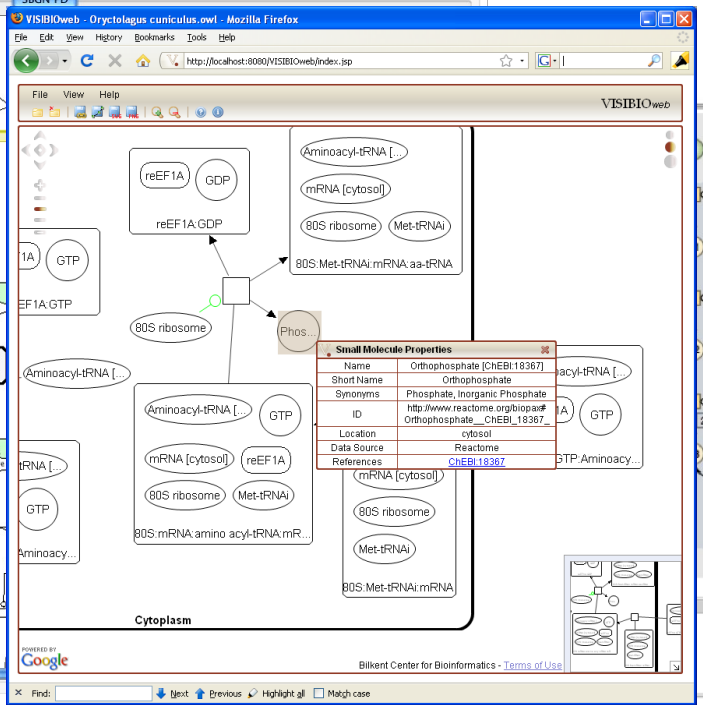

BioModels Home
Browse models
Submit
Sign in
Support
About BioModels
BIOMD0000000005 - Tyson1991_CellCycle_6var
Publication ID: 1831270
Original Model: Unspecified
Submitter: Nicolas Le Novère
Submission Date: 2005-09-13T12:31:08+00:00
Last Modification Date: 2009-02-25T14:58:48+00:00
Creation Date: 2005-02-08T18:28:27+00:00
Creators: Bruce Shapiro
Curation Result:
Cell Cycle Model; Tyson (1991, 6 variables)
Description
Reaction

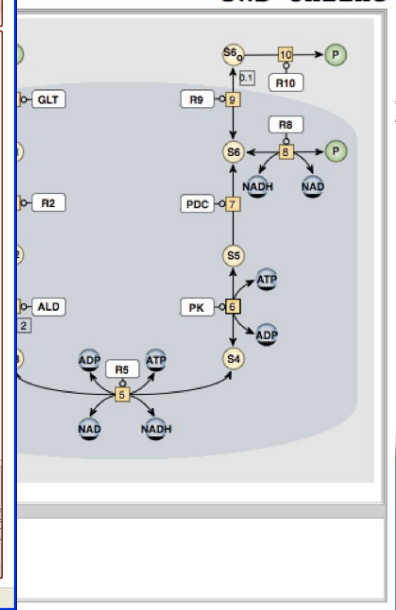
BIOMD0000000005
Refresh
Close
p-cyclin
p-cyclin cdc2
cdc2k
total cdc2
p-cyclin cdc2-p
cdc2k-P
cyclin
Ø


PANTHER
Classification System
INSULIN/IGF PATHWAY-MITOGEN ACTIVATED PROTEIN KINASE KINASE/MAP KINASE CASCADE
Export:
Select one
created with CellDesigner
Click components to make selections. Right-click components for more options.
Search selections for:
Genes
Go
Pathway Diagram
Pathway Description
File Edit Zoom Select Color
Interactive Standard/Activity Flow
SBGN Image View
SBGN PD


EPE2
synaptic button
ChAT
vAChT
synaptic vesicle
ACh
choline
CHT1
SNARE
AChE
acetate
muscle cytosol
nAChR
closed
open
myosin
actin
ATP


Insulin
IGF
INSR
IGF-R
IRR
SOS
Grb2
Active
Ras
Raf
ATP
MEK


VISIBIOweb - Oryctolagus cuniculus.owl - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:8080/VISIBIOweb/index.jsp
VISIBIOweb
reEF1A
GDP
reEF1A:GDP
Aminoacyl-tRNA [...
mRNA [cytosol]
80S ribosome
Met-tRNAi
80S:Met-tRNAi:mRNA:aa-tRNA
GTP
Phos...
Small Molecule Properties
Name
Orthophosphate [ChEBI:18367]
Short Name
Orthophosphate
Synonyms
Phosphate, Inorganic Phosphate
ID
http://www.reactome.org/biopax#Orthophosphate__ChEBI_18367_
Location
cytosol
Data Source
Reactome
References
ChEBI:18367
80S:mRNA:amino acyl-tRNA:mR...
80S:Met-tRNAi:mRNA
Cytoplasm
Bilkent Center for Bioinformatics - Terms of Use
Find:
Next
Previous
Highlight all
Match case


JWS Online
GLT
R9
R8
S6
NADH
NAD
R2
PDC
S5
ALD
PK
ATP
ADP
S4
R5

Status of Specifications

- SBGN Process Diagrams
  - Level 1 Version 1.0 release on August 23rd 2008
  - Level 1 Version 1.1 to be released over summer
  - Level 1 Version 2 to be finalised over summer
- SBGN Entity Relationships
  - Level 1 Version 1.0 to be released over summer
- SBGN Activity Flow
  - Level 1 Version 1.0 to be released over summer

Future SBGN meetings

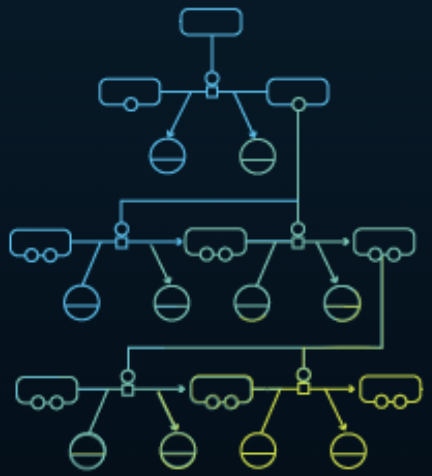
- 5th SBGN forum
  - 02-03 September 2009,
  - San Francisco
  - “Satellite” of ICSB 2009
- 4rd SBGN hackathon (SBGN 5.5)
  - 21-23 April 2010
  - Wittenberg
- 6th SBGN forum (provisional)
  - October 2010
  - Edinburgh
  - Satellite of ICSB 2010
- 5rd SBGN hackathon (SBGN 6.5)
  - Spring 2011, Bethesda, USA

SBGN Systems Biology Graphical Notation

http://www.sbgn.org/

Home | Documents | Forums | Community | Events | About

Google Site Search..



A Visual Notation for Network Diagrams in Biology

SBGN.org is the global portal for documentation, news, and other information about the Systems Biology Graphical Notation (SBGN) project, an effort to standardize the graphical notation used in diagrams of biochemical and cellular processes studied in systems biology.

Standardizing the diagrammatic notation is crucial for more efficient and accurate transmission of biological knowledge between different research communities in the expanding field of systems biology. Notations traditionally used by researchers and software have been informal, idiosyncratic and highly variable. Until SBGN, there has been no standard agreed-upon convention defining precisely how to draw biochemical interaction diagrams in a regular and systematic way that helps readers interpret them consistently and unambiguously.

SBGN defines a comprehensive set of symbols with precise semantics, together with detailed syntactic rules defining their use and how diagrams are to be interpreted. By standardizing the visual notation, SBGN can serve as a bridge between different communities in research, education, publishing, and more. The real payoff will come when researchers are as familiar with the notation as electronics engineers are familiar with the notation of circuit schematics. If researchers are saved the time and effort required to familiarize themselves with different notations, they can spend more time thinking about the biology being depicted.

On this site, you can browse some example diagrams to get a feeling for SBGN, read the SBGN specification documents, join online discussions, see current working documents in the SBGN wiki, and much more.

SBGN is the work of many people. It would not have been possible without the generous support of multiple organizations over the years, for which we are very thankful.

SBGN News

(23 Aug.'08) The first **SBGN Process Diagrams Level 1** specification is out! Download the specification and tell us what you think!

(Locked) | Watch page | Page History | Upload file | File list | Page list | My watchlist | Preferences | Help | Log out

Acknowledgements

SBGN editors

- Nicolas Le Novère (UK)
- Stuart Moodie (UK)
- Anatoly Sorokin (UK)
- Michael Hucka (US)
- Falk Schreiber (DE)
- Huaiyu Mi (US)

Special contributors

Emek Demir, Katja Wegner, Mirit Aladjem, Sarala Wimalaratne, Frank Bergman, Ralph Gauges, Peter Ghazal, Kawaji Hideya, Lu Li, Yukiko Matsuoka, Alice Villéger, Sarah Boyd, Laurence Calzone, Melanie Courtot, Ugur Dogrusoz, Akira Funahashi, Sohyoung Kim, Fedor Kolpakov, Augustin Luna, Sven Sahle, Douglas Kell, Kurt Kohn, Hiroaki Kitano

The whole community participating to SBGN meetings, and mailing-lists

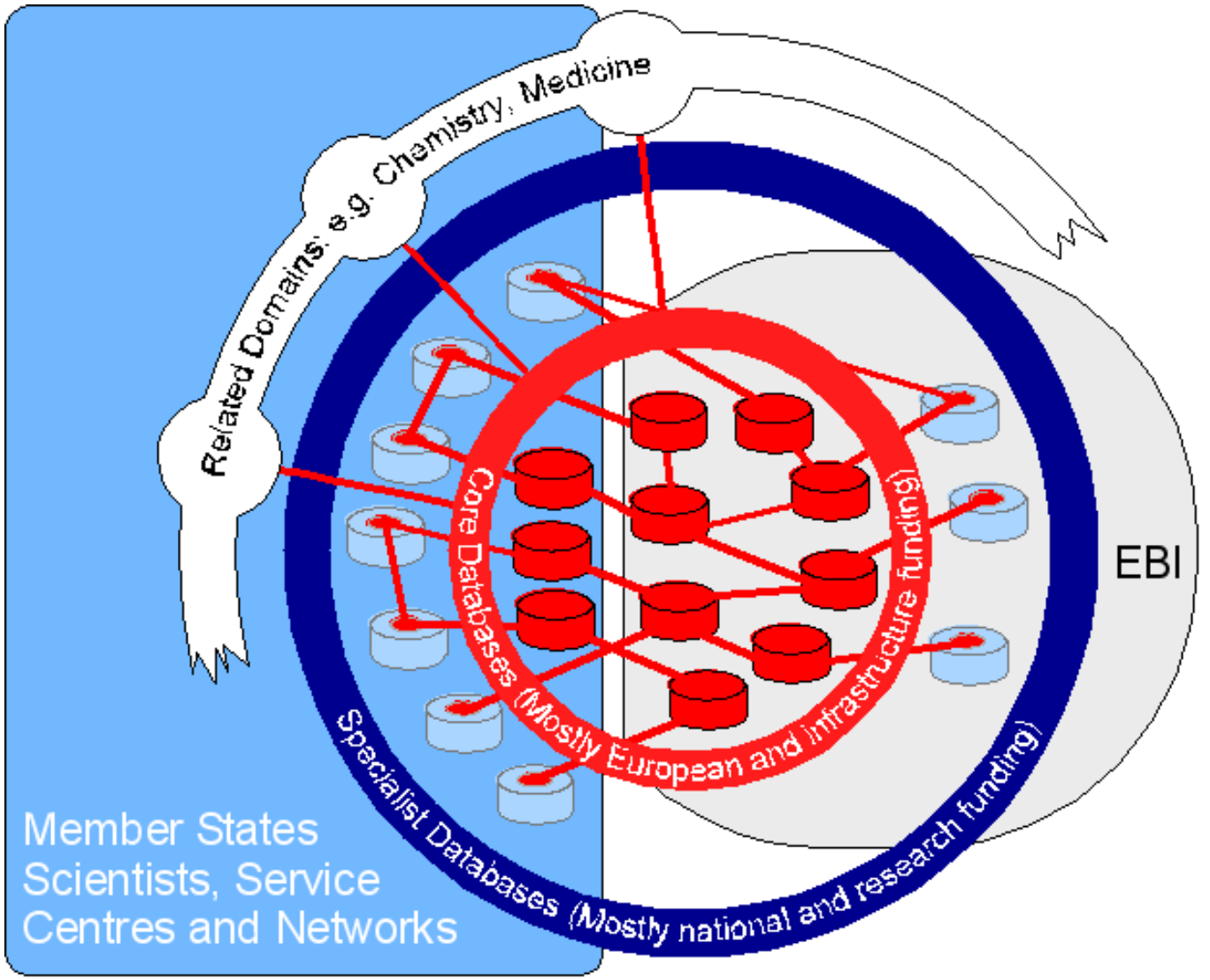
NEDO

bbsrc biotechnology and biological sciences research council

NIGMS

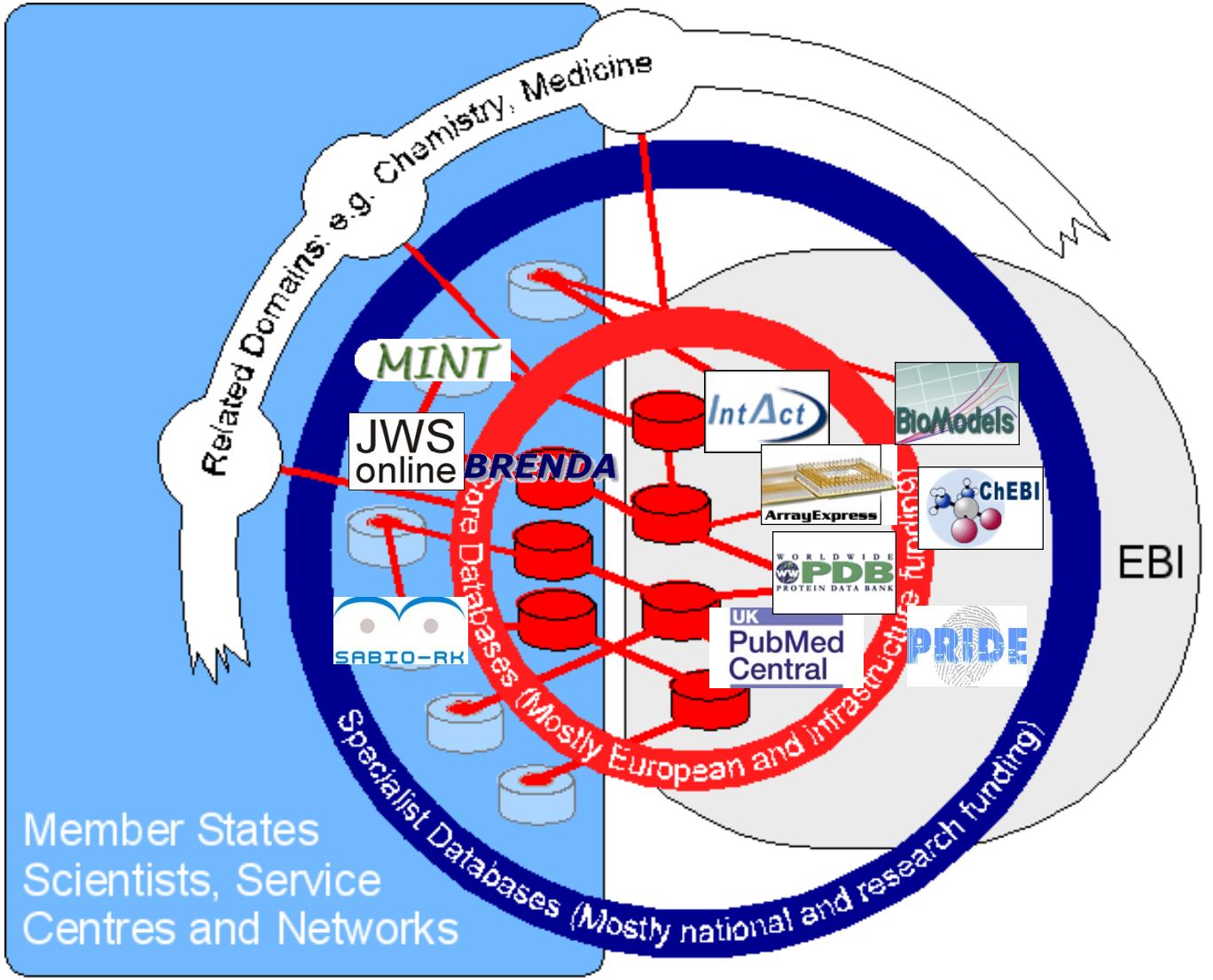
EMBL

ELIXIR content



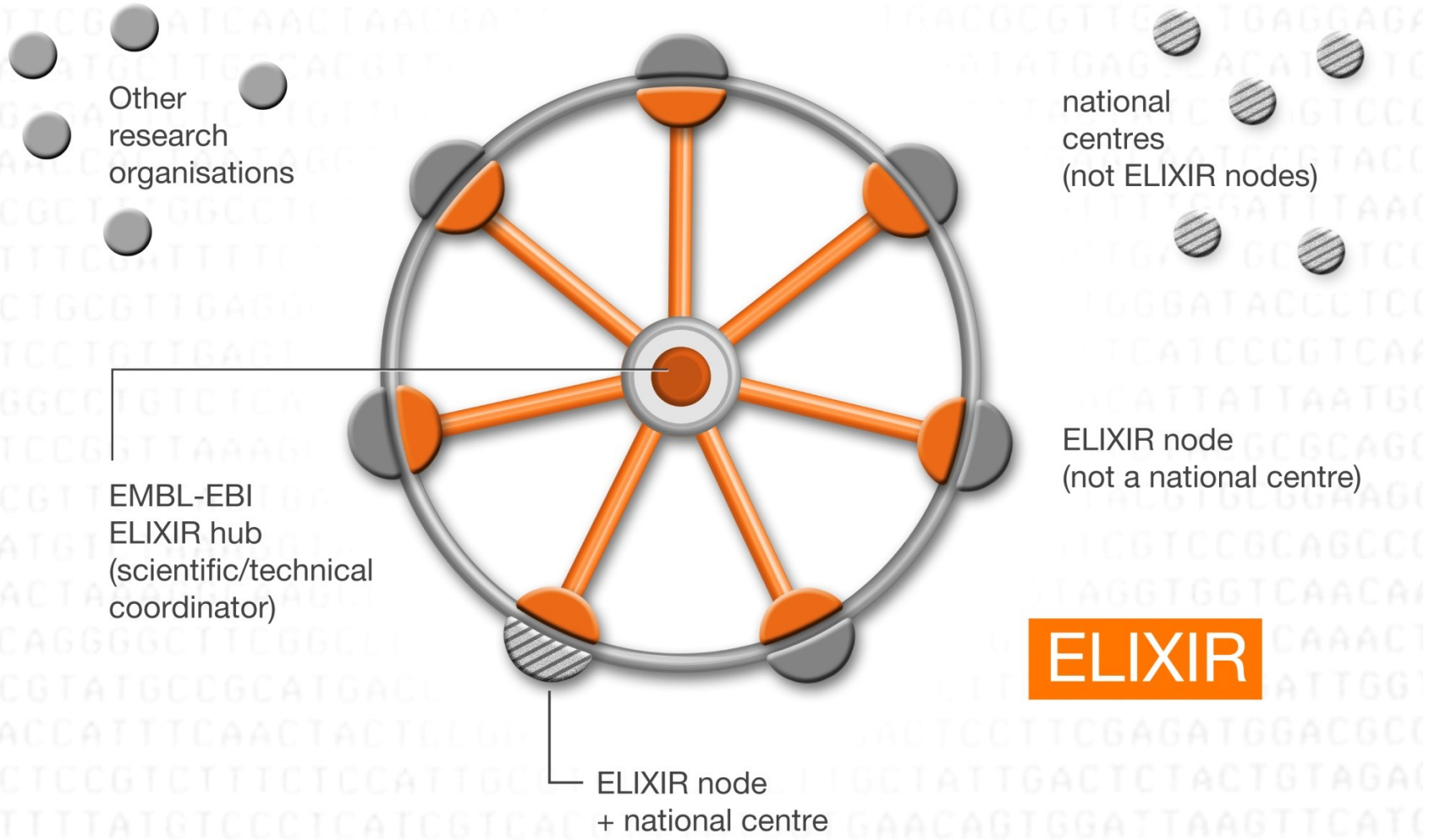
ELIXIR: *a **sustainable** infrastructure for biological information in Europe.*

ELIXIR content



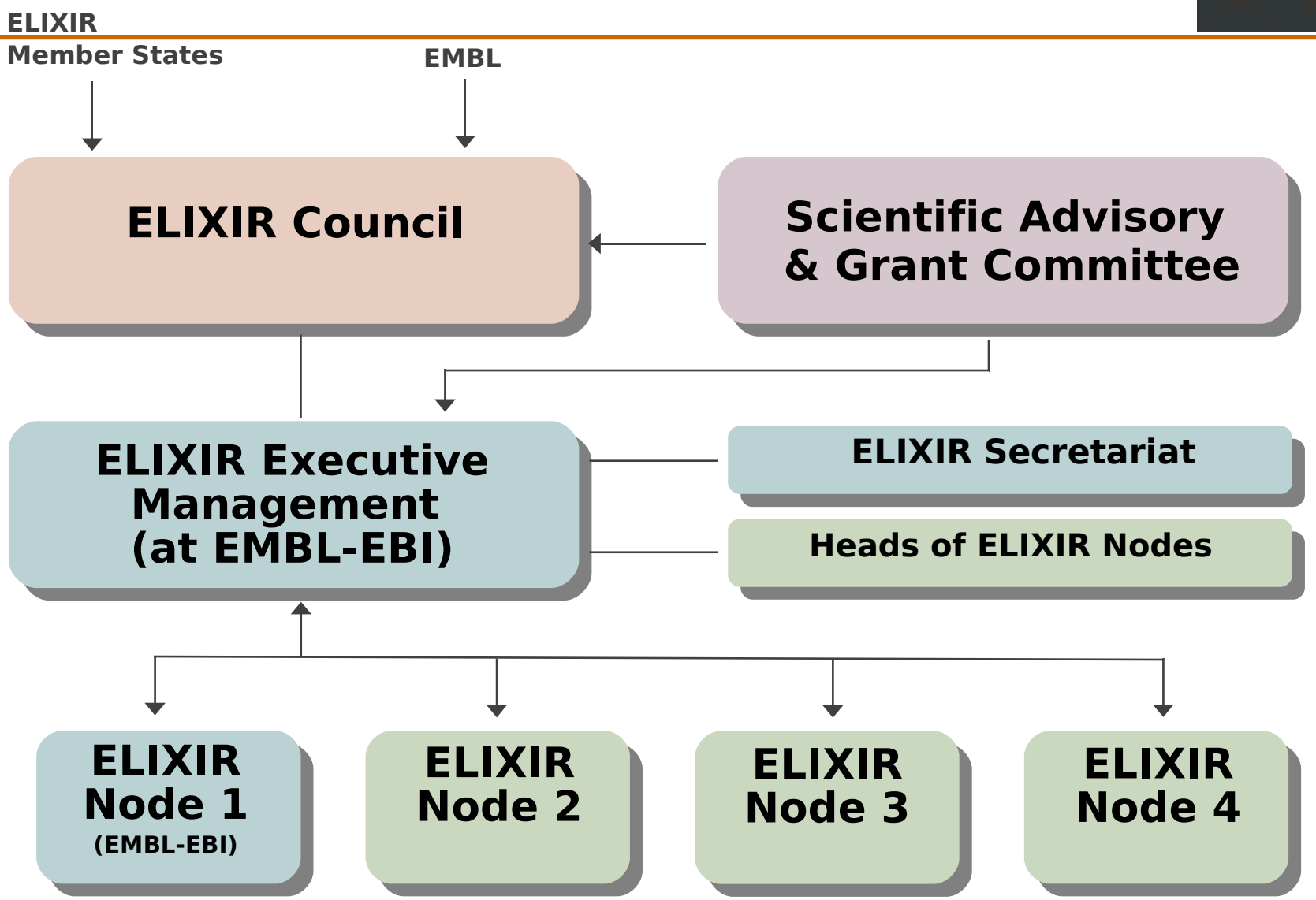
ELIXIR: *a **sustainable** infrastructure for biological information in Europe.*

ELIXIR Scientific and technical Structure

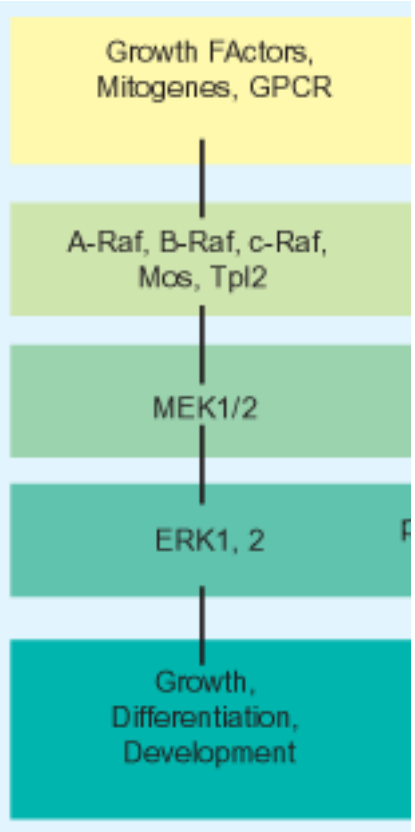


ELIXIR: *a **sustainable** infrastructure for biological information in Europe.*

ELIXIR Legal and Governance

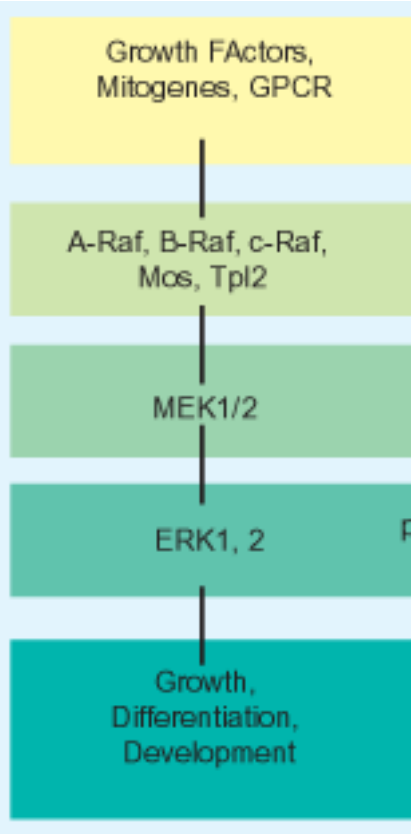


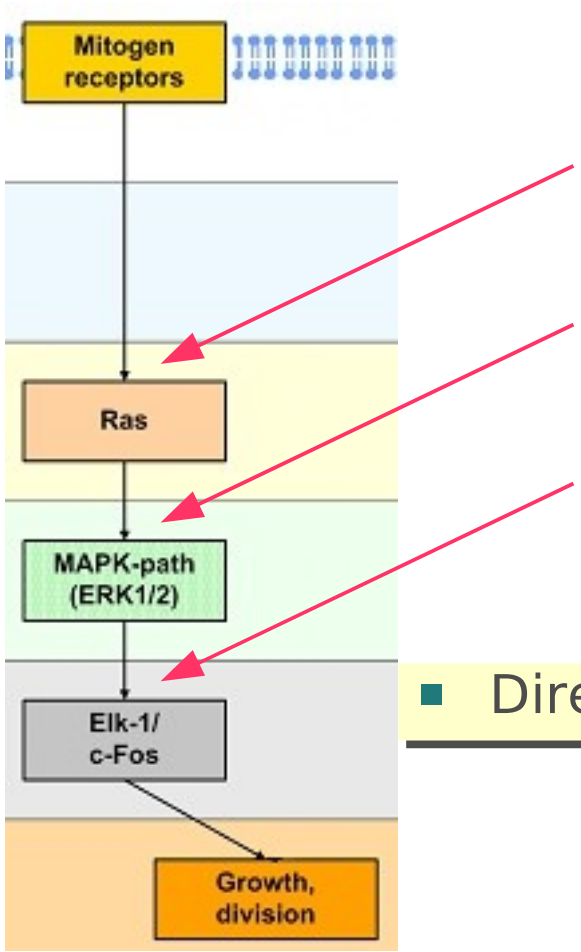
Different representations of a pathway



- No temporal sequence
- No directionality
- No biochemical effects
- No mechanistic descriptions

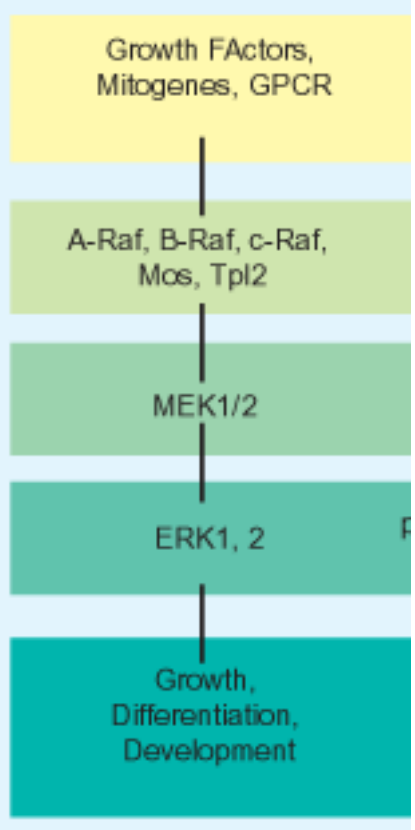
Different representations of a pathway

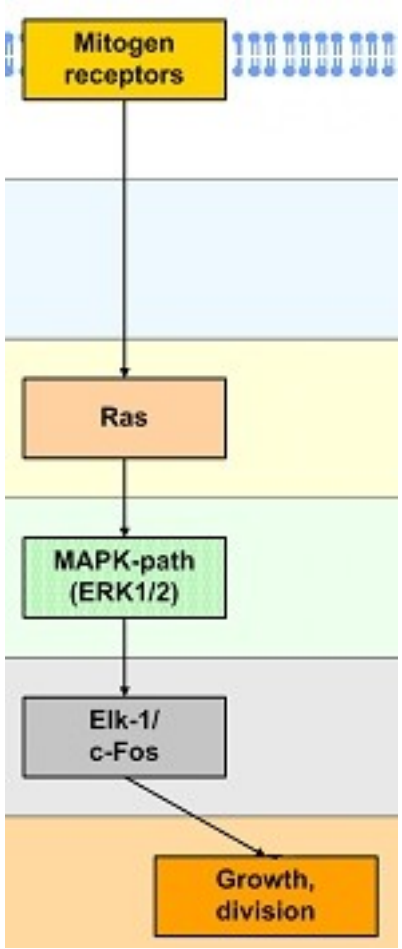


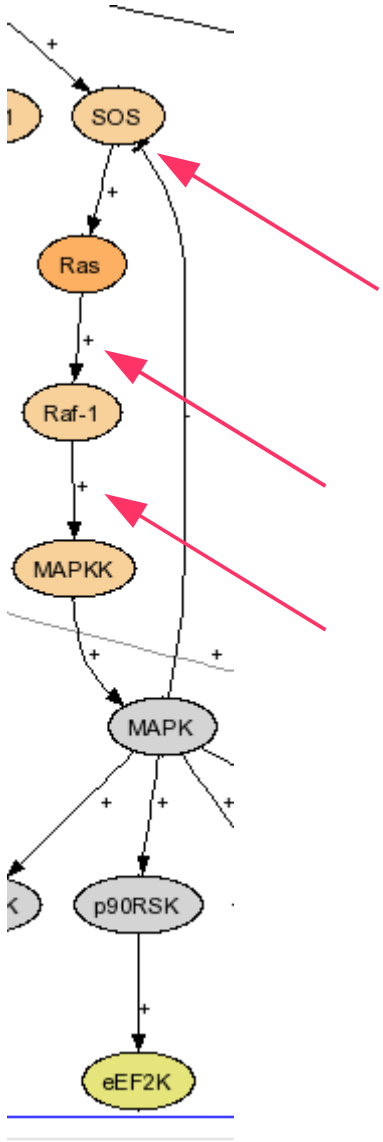


- Directionality of influence

Different representations of a pathway

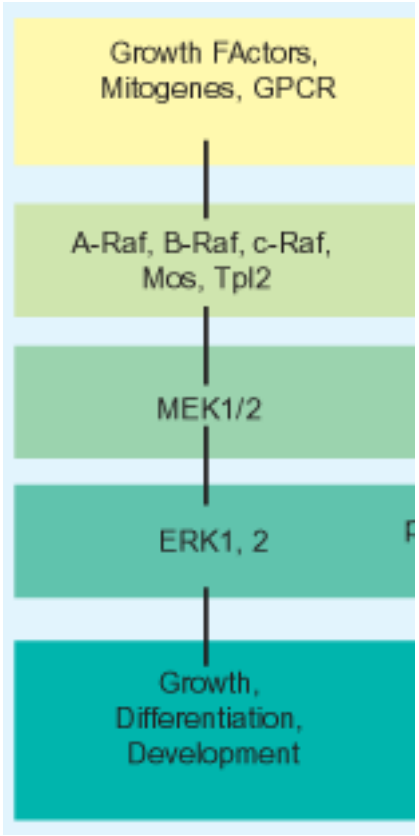


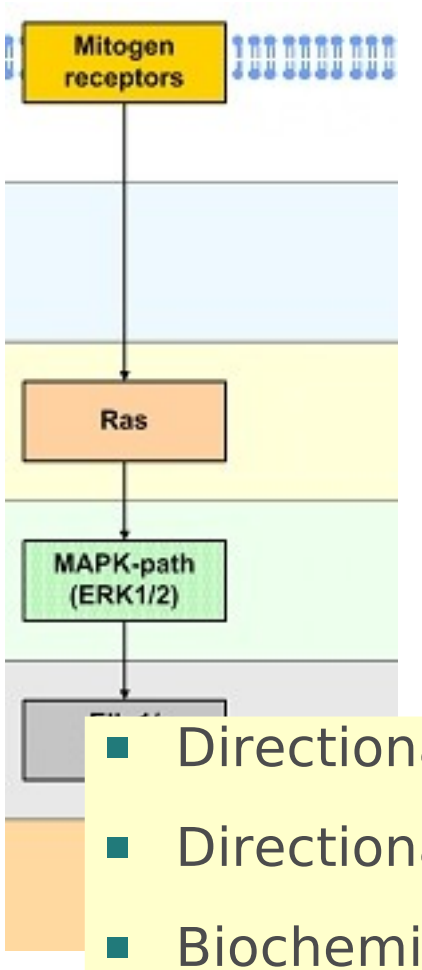


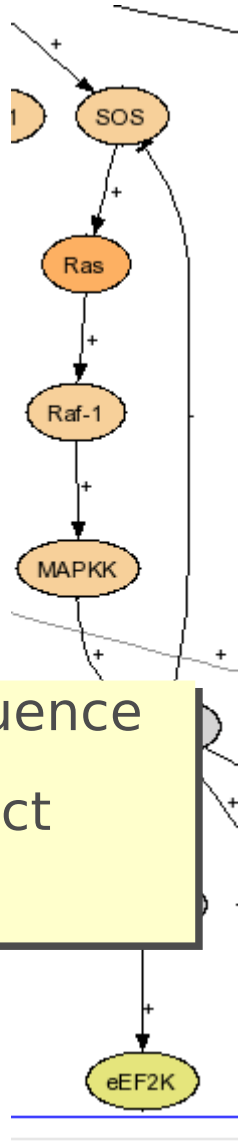


- Directionality of influence
- Directionality of effect

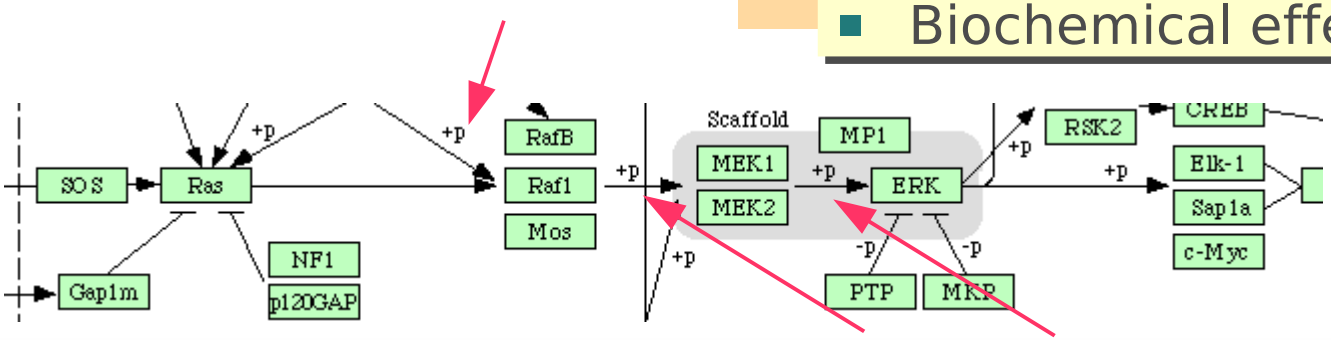
Different representations of a pathway



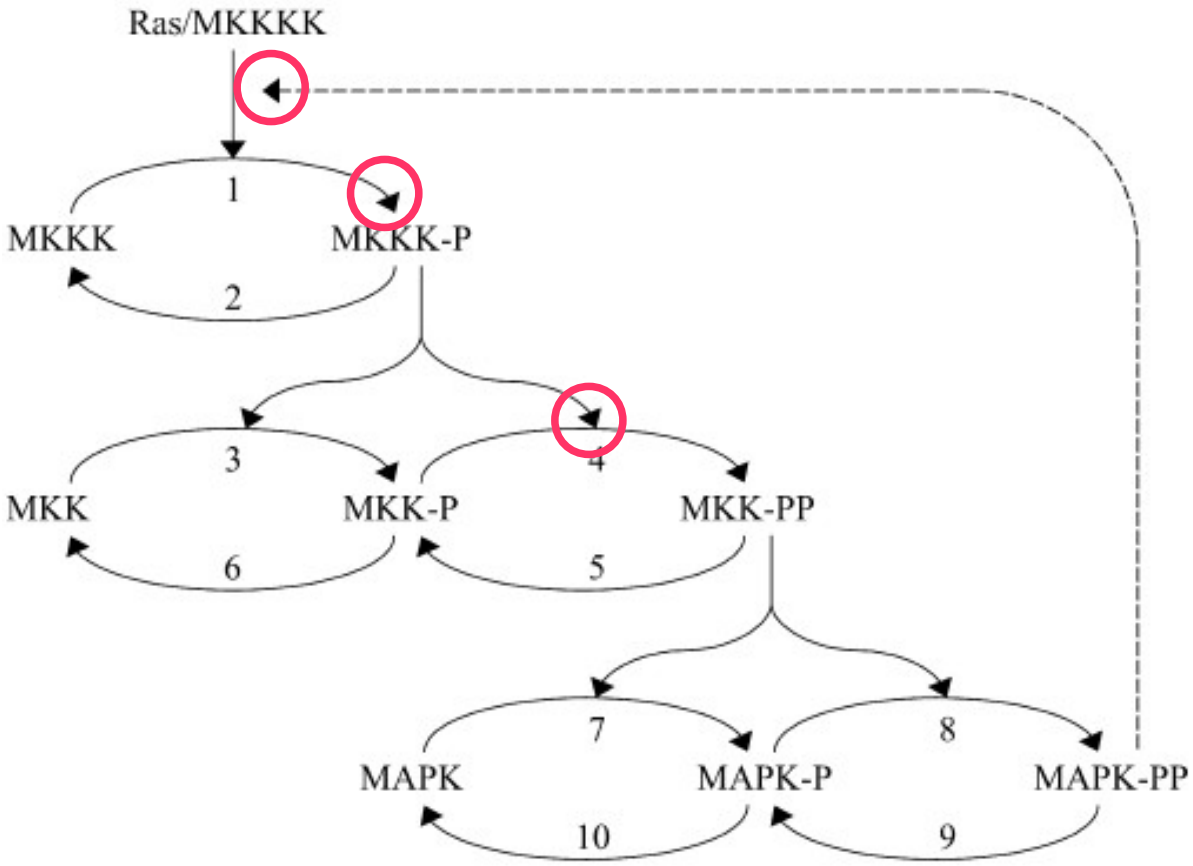




- Directionality of influence
- Directionality of effect
- Biochemical effect



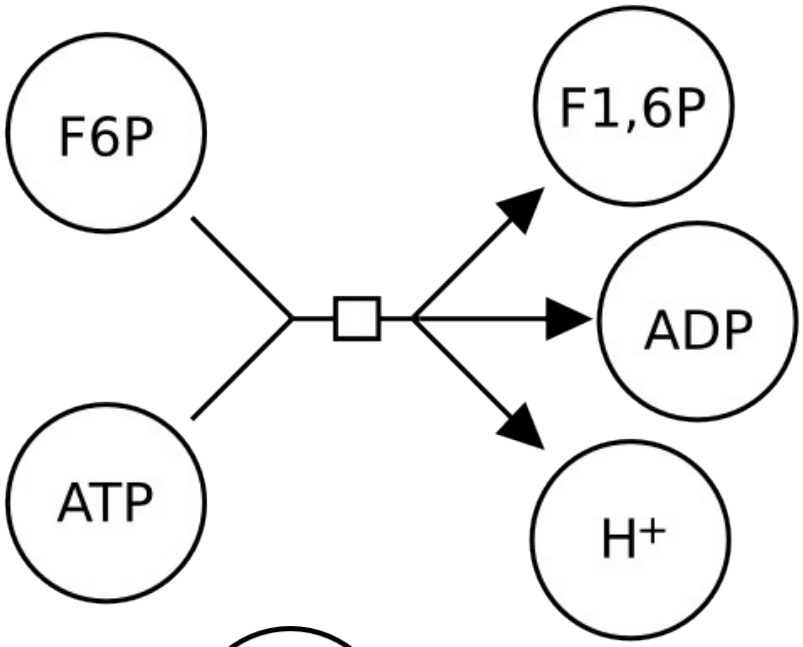
Different representations of a pathway

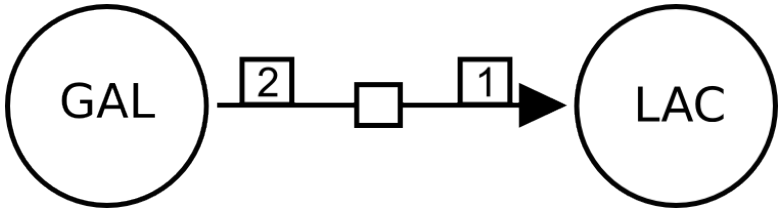


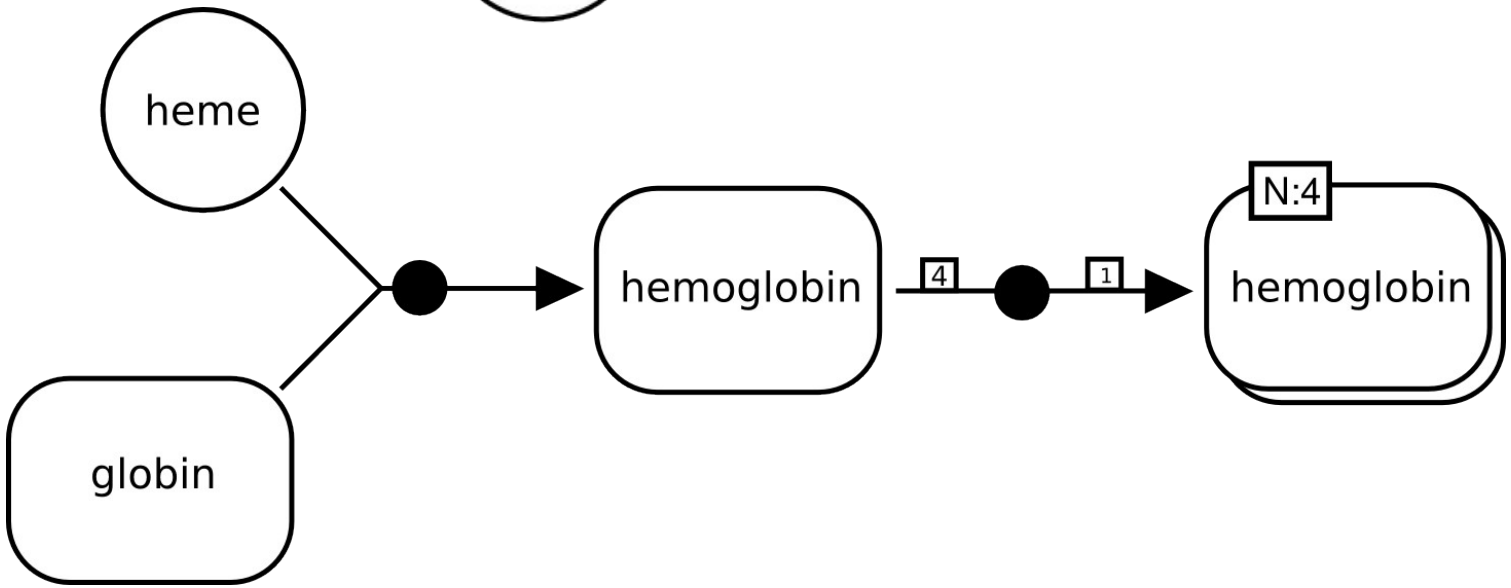
THREE *different* meanings for ONE symbol!

- Transition
- catalysis
- inhibition

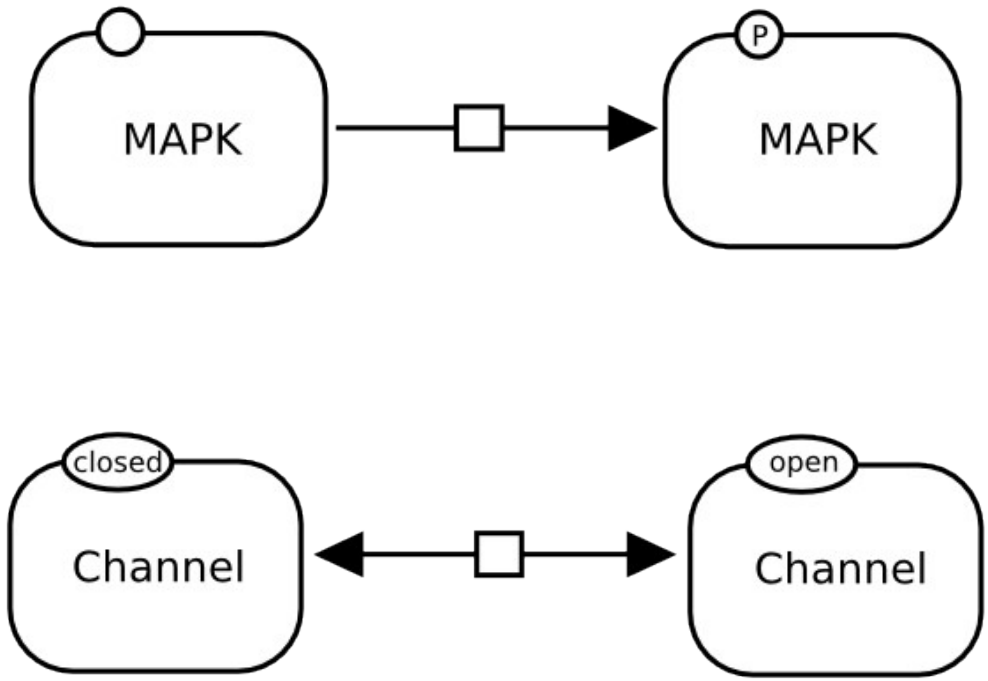
Examples of transitions



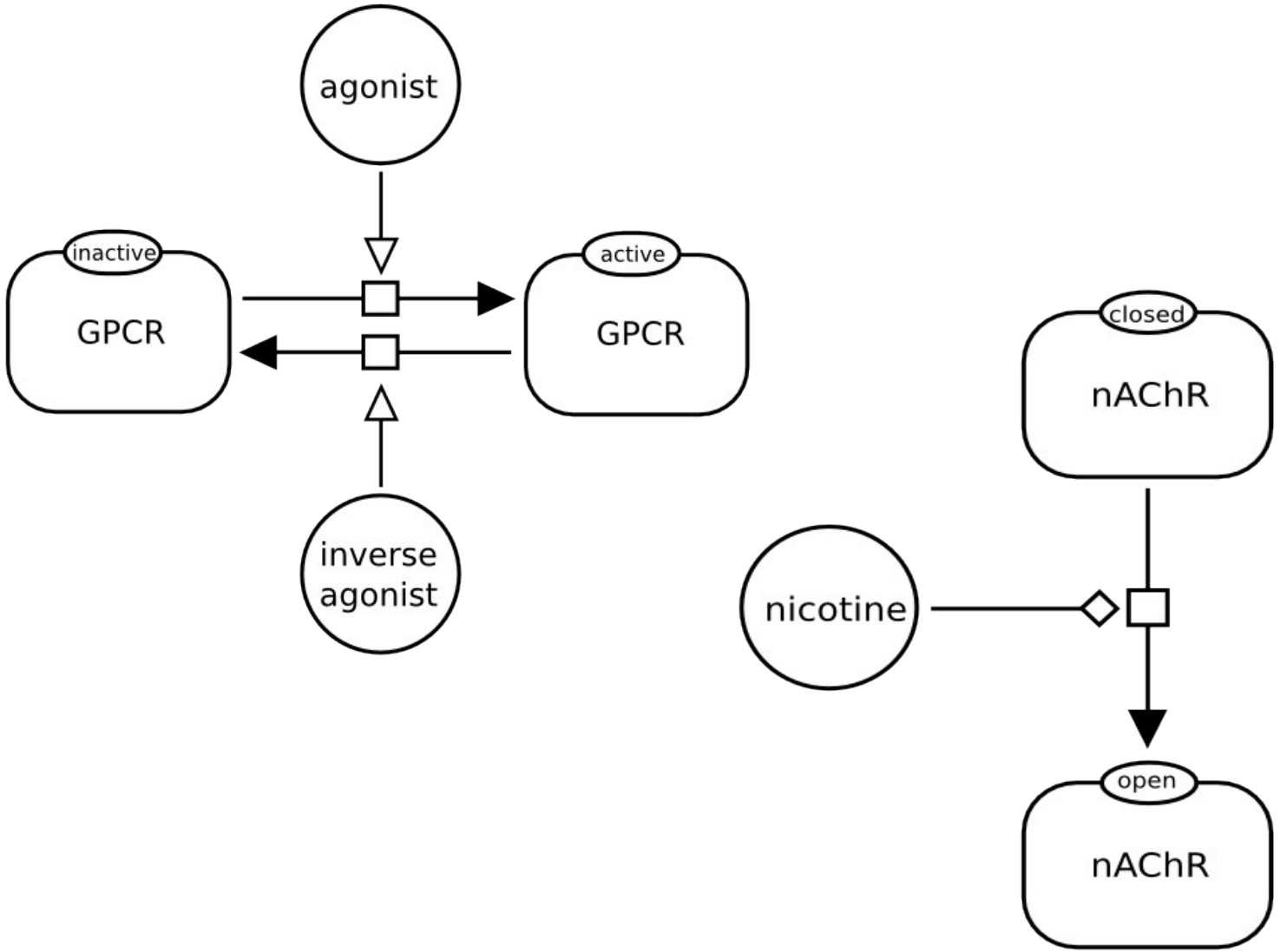




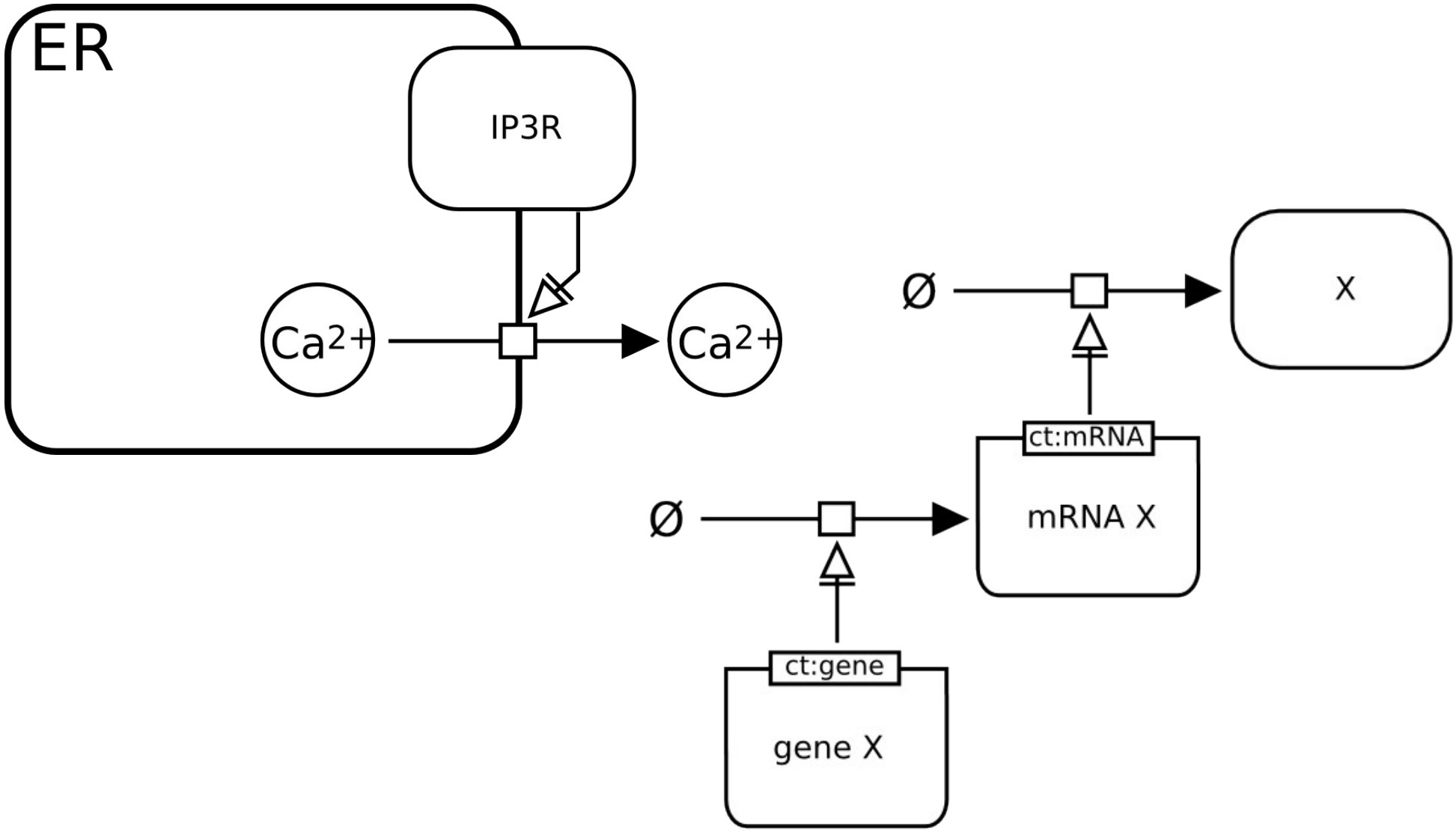
Transitions with state variables



Transitions with modulations



Necessary stimulations



Use of logical operators