

A common layout for the Tower of Babel: The Systems Biology Graphical Notation

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

Graphs are everywhere in biology

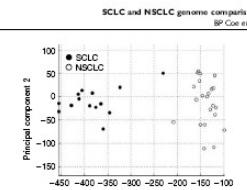
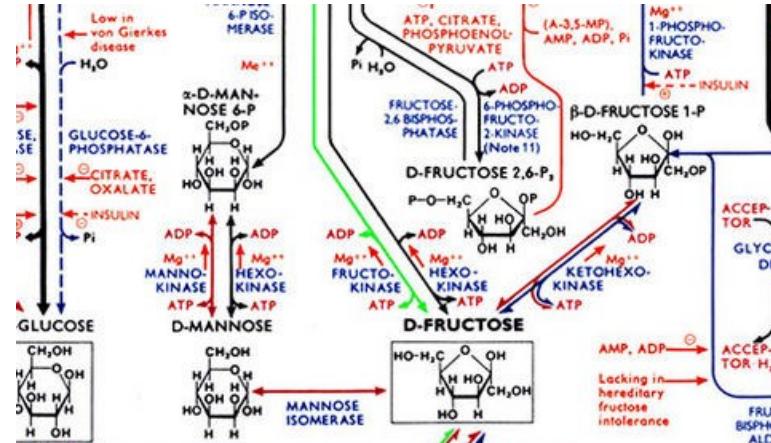
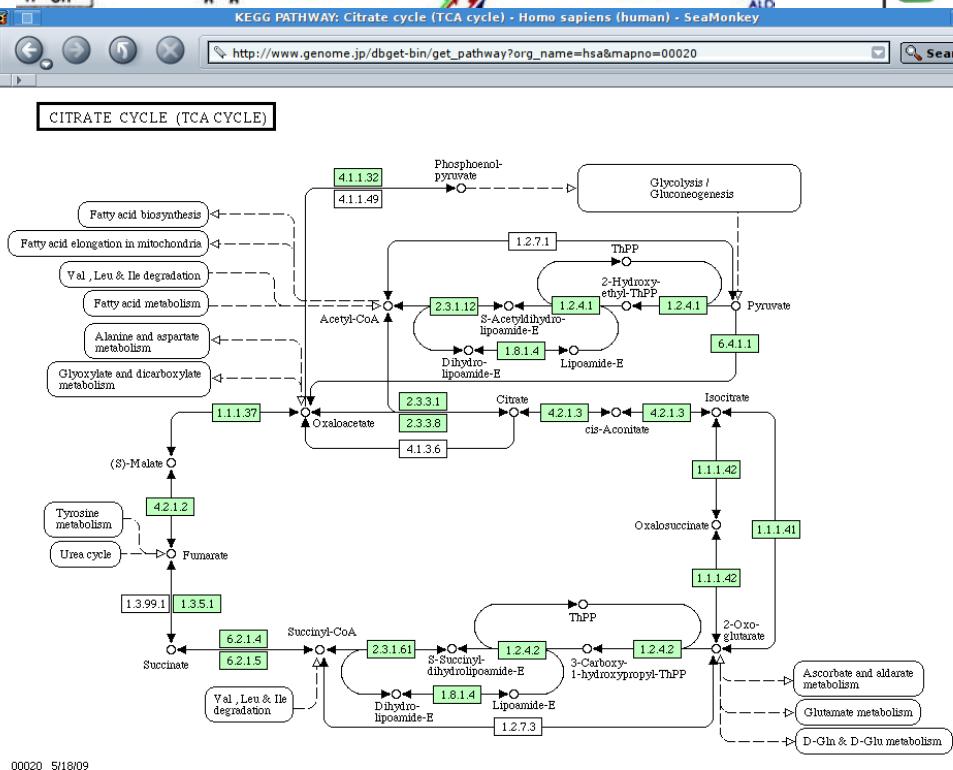
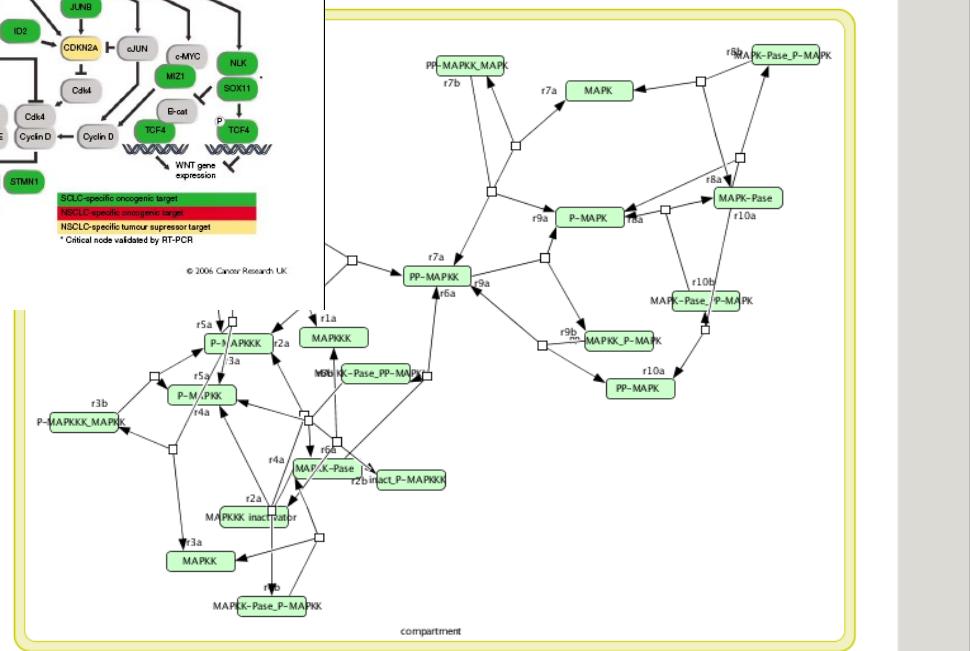
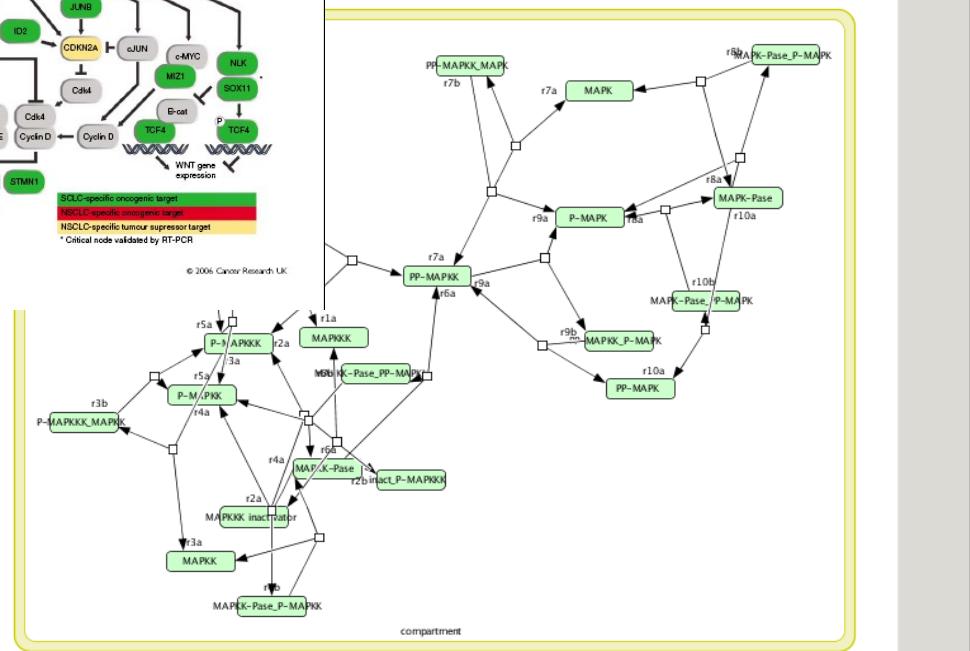
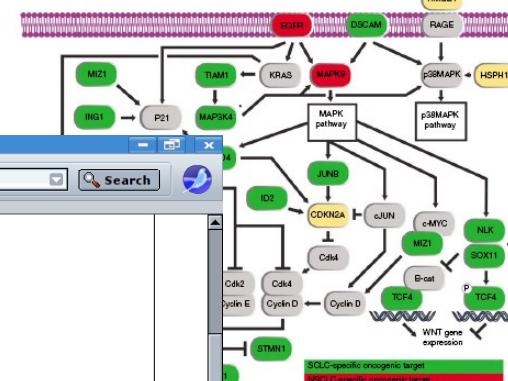
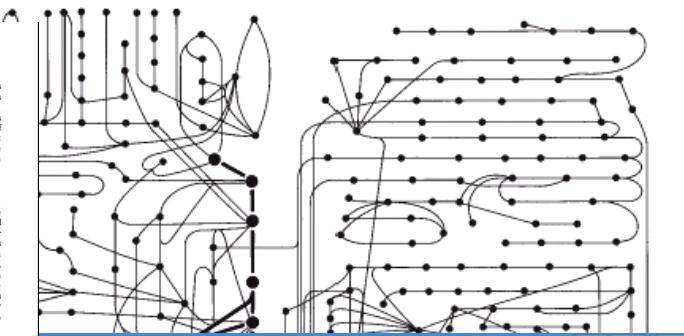
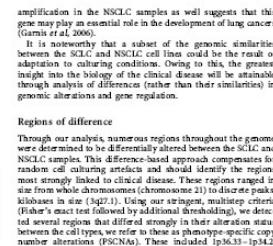


Figure 4 Contribution of copy number-induced gene expression differences to the SCLC and NSCLC phenotypes. Principal component analysis was performed utilising all 243 Affymetrix probe sets demonstrating expression differences as a result of copy number alterations. The SCLC samples are indicated by solid circles, while the NSCLC samples are indicated by open circles. Strong separation of the SCLC and NSCLC clusters along principal component 1 demonstrates the contribution of these genes to the differential phenotypes.



00020 5/18/09
(c) Kanehisa Laboratories



Ambiguity of usual representations

X → Y

X → 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

X → Y

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

X → Y

is transformed into

translocates (X "=" Y)

is degraded into

associates into

dissociates into

stimulates the activity of

stimulates the expression of

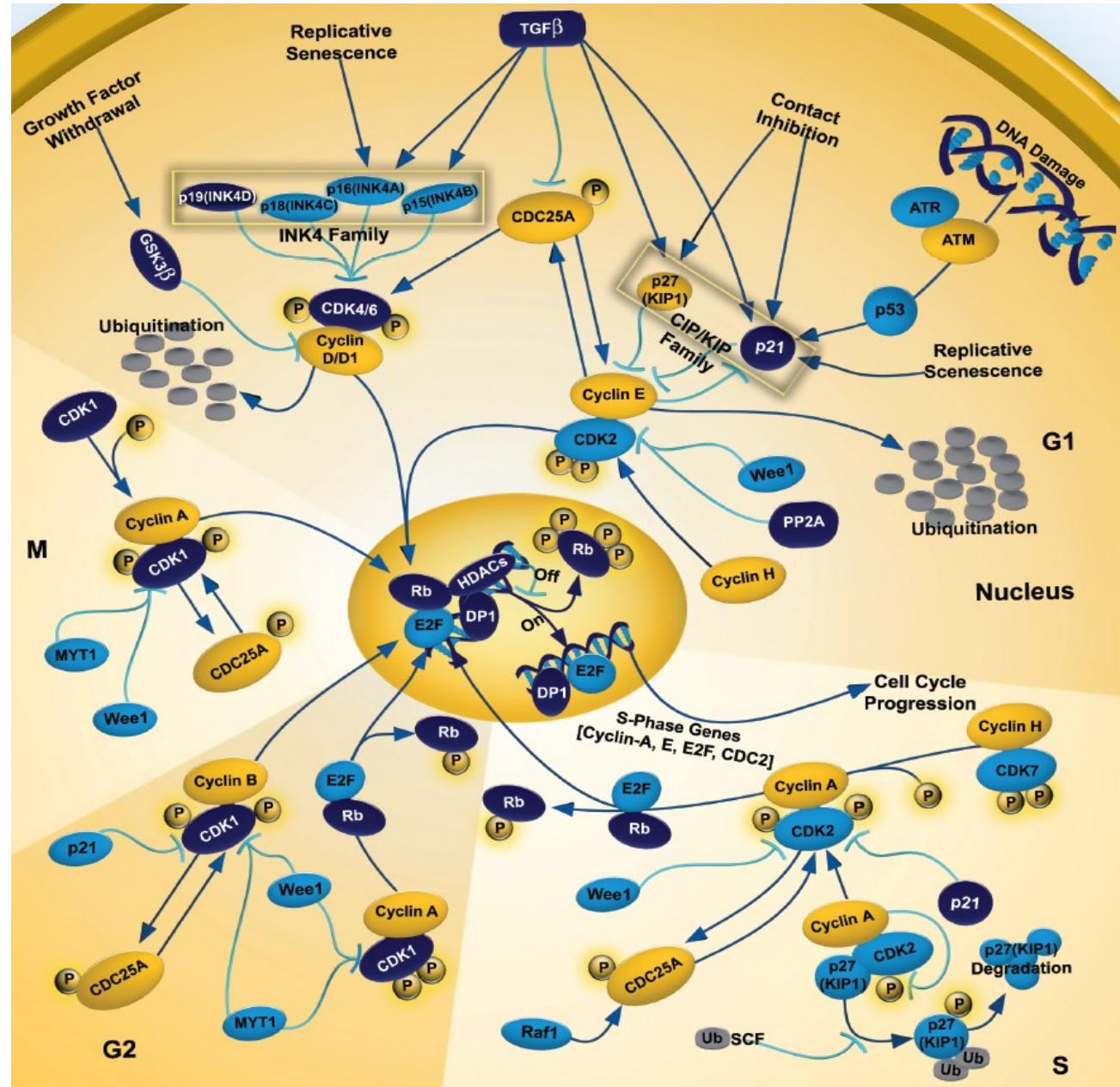
catalyses the formation of

X inhibits Y



inhibition

Can this be understood by biologists?



Can this be understood by biologists?

Stimulates? but ...
what exactly?

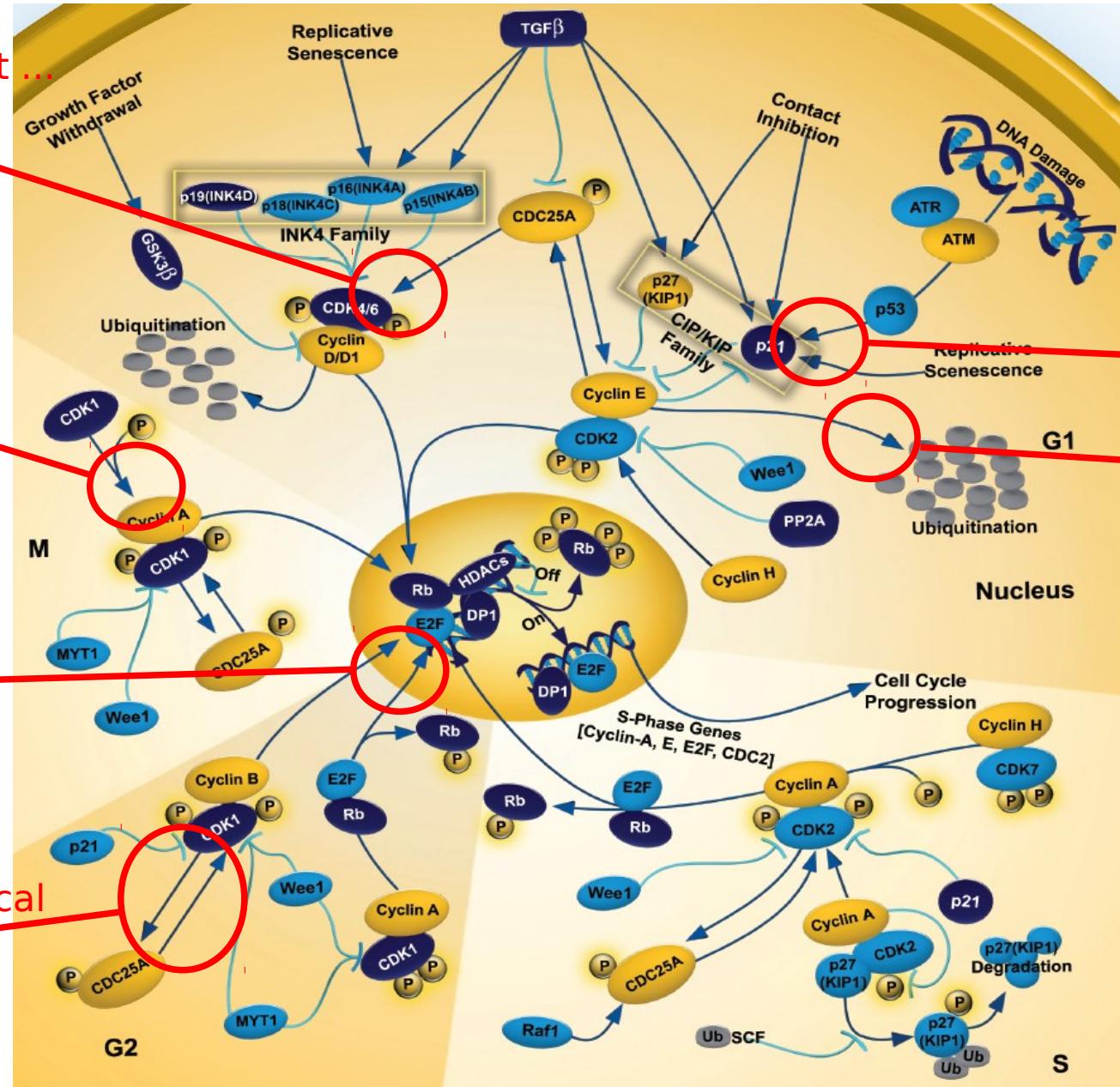
Associates into?

Translocates?

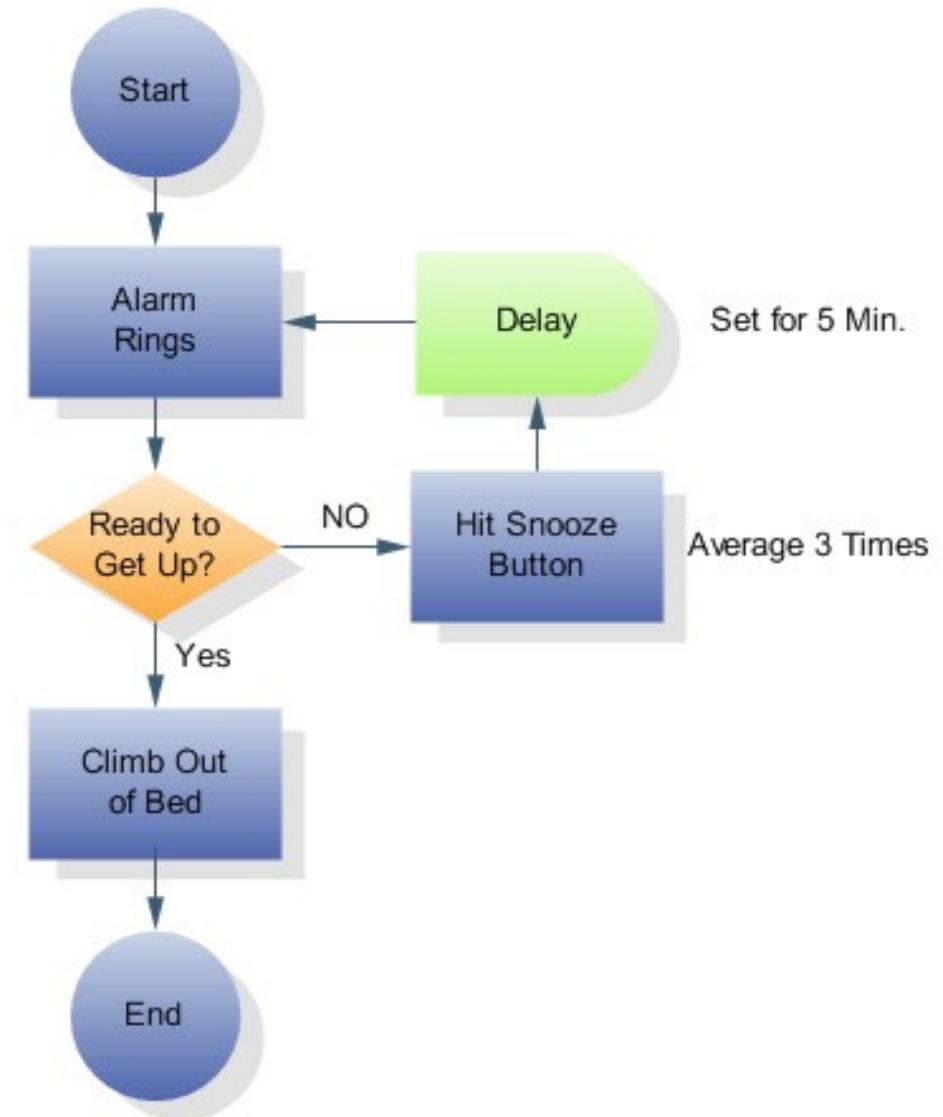
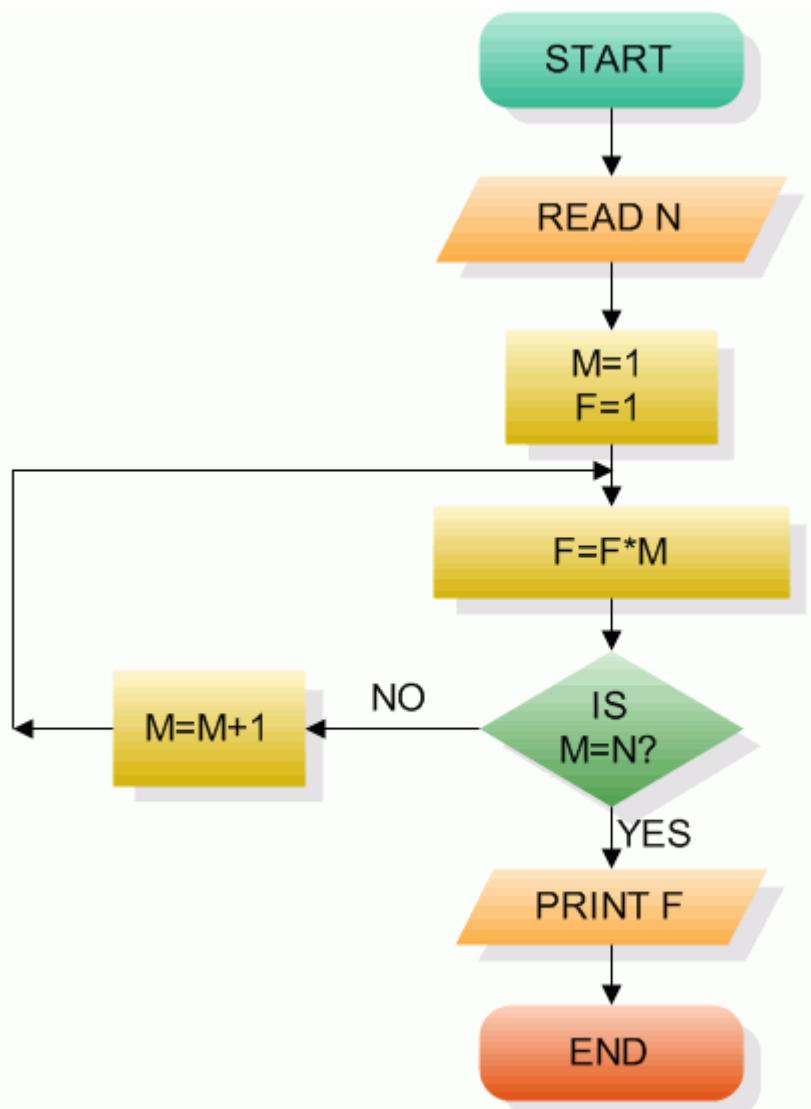
No idea. Reciprocal
stimulation?

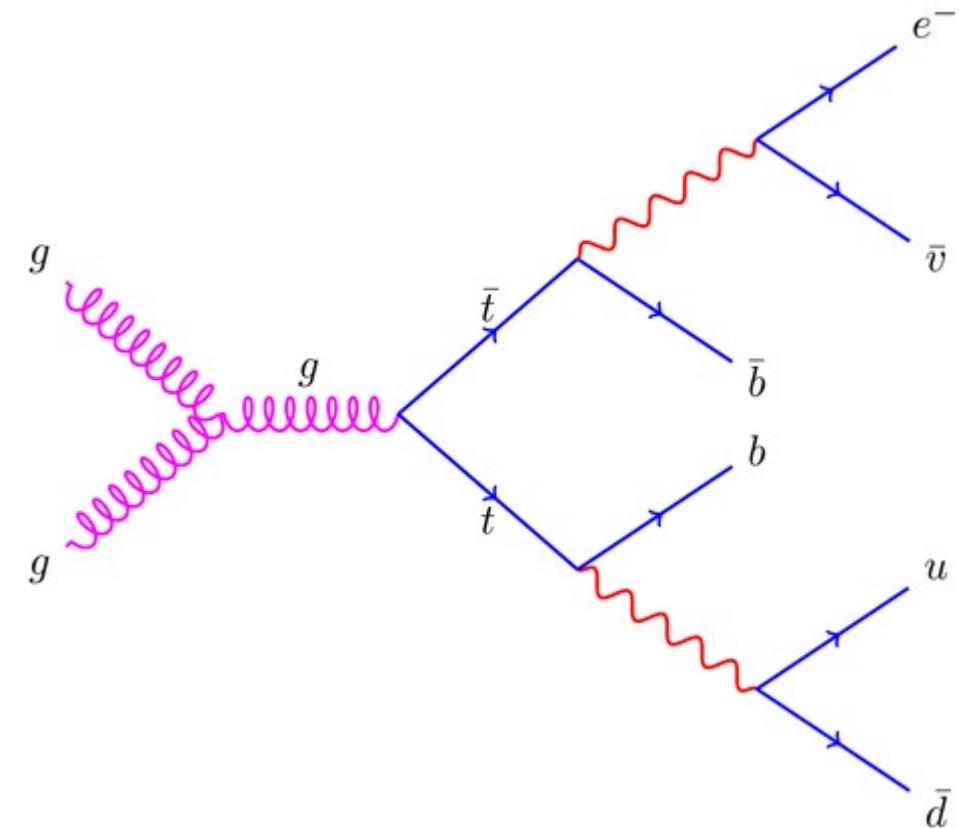
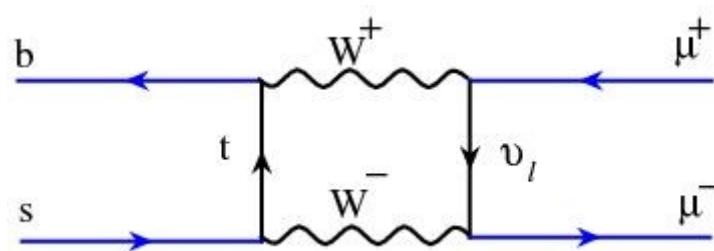
Stimulates gene
transcription?

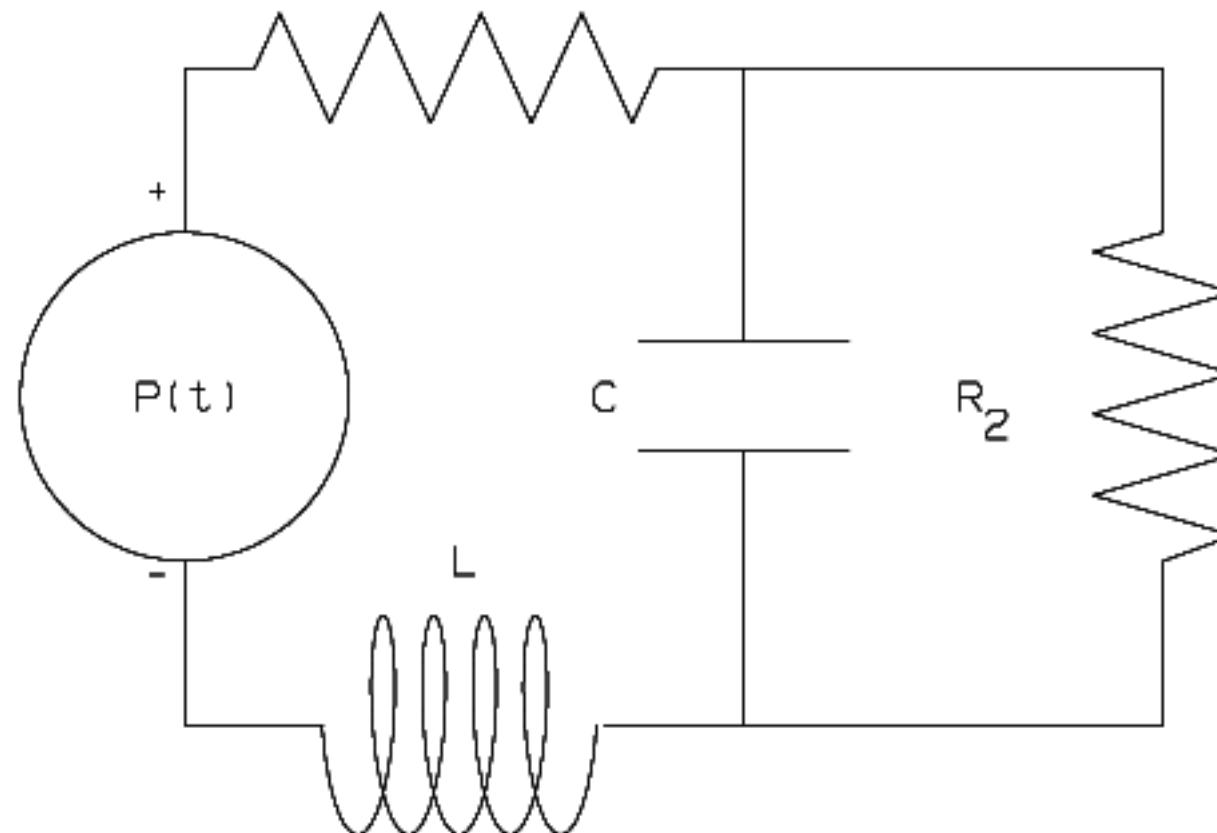
Is degraded?

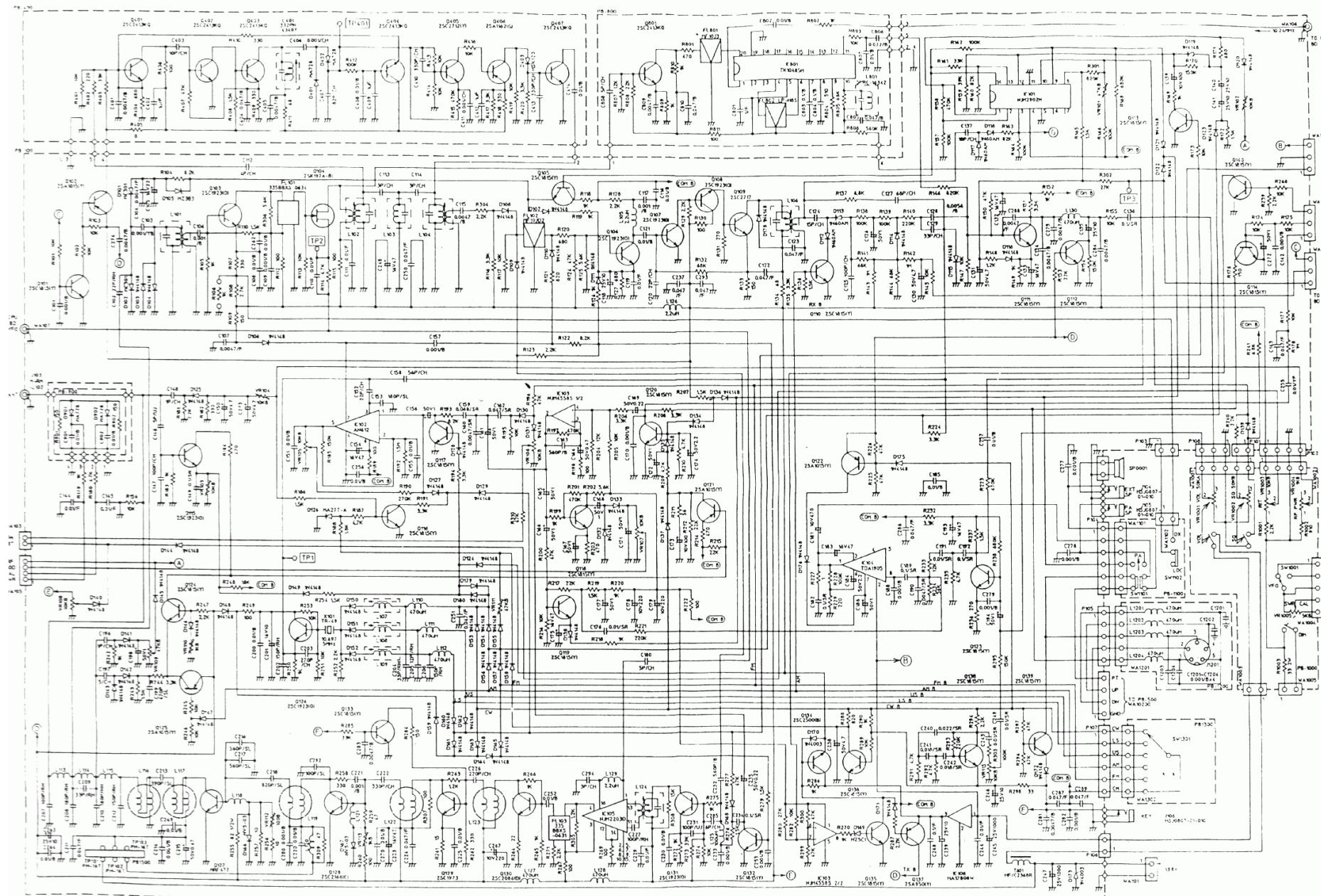


Every computer scientist understands those









What did-those diagram bring?



FOOD



WORK



COM



TRAVEL



FUN



Basic science

Systems Biology

BMBF PUBLIK

Technology



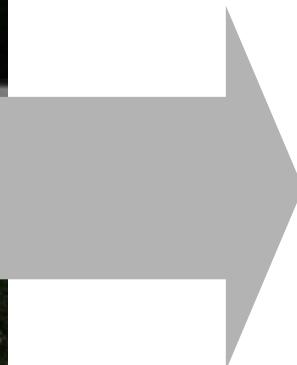
EXTREME GENETIC ENGINEERING

An Introduction to Synthetic Biology

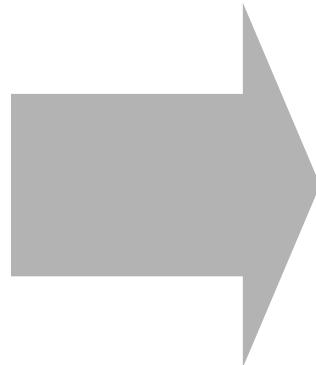
January 2007



What happens if one cannot read the blueprint

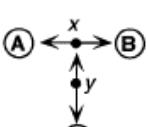
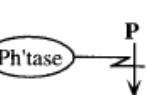
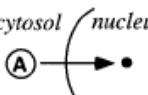


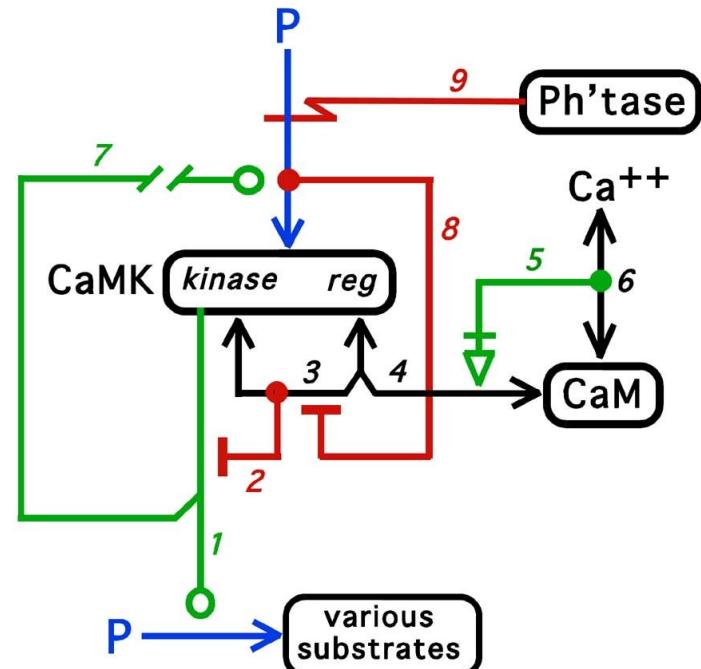
What happens if one cannot read the blueprint



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

	Non-covalent binding, for example between proteins A and B. A filled circle or "node" can be placed on the connecting line to represent the A:B complex itself.
	Asymmetric binding where protein A contributes a peptide that binds to a receptor site or pocket on protein B.
	Representation of multimolecular complexes: x is A:B; y is (A:B):C. This notation is extensible to any number of components in a complex.
	Covalent modification of protein A. The single-arrowed line indicates that A can exist in a phosphorylated state. The node represents the phosphorylated species.
	Cleavage of a covalent bond: dephosphorylation of A by a phosphatase.
	Proteolytic cleavage at a specific site within a protein.
	Stoichiometric conversion of A into B.
	Transport of A from cytosol to nucleus. The filled circle represents A after it has been transported into the nucleus (the node functions like a ditto mark).
	Formation of a homodimer. Filled circle on the right represents another copy of A. The filled circle on the binding line represents the homodimer A:A.
	z is the combination of states defined by x and y.
	Enzymatic stimulation of a reaction.
	General symbol for stimulation.
	General symbol for inhibition.
	Shorthand symbol for transcriptional activation.
	Shorthand symbol for transcriptional inhibition.
	Degradation products



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

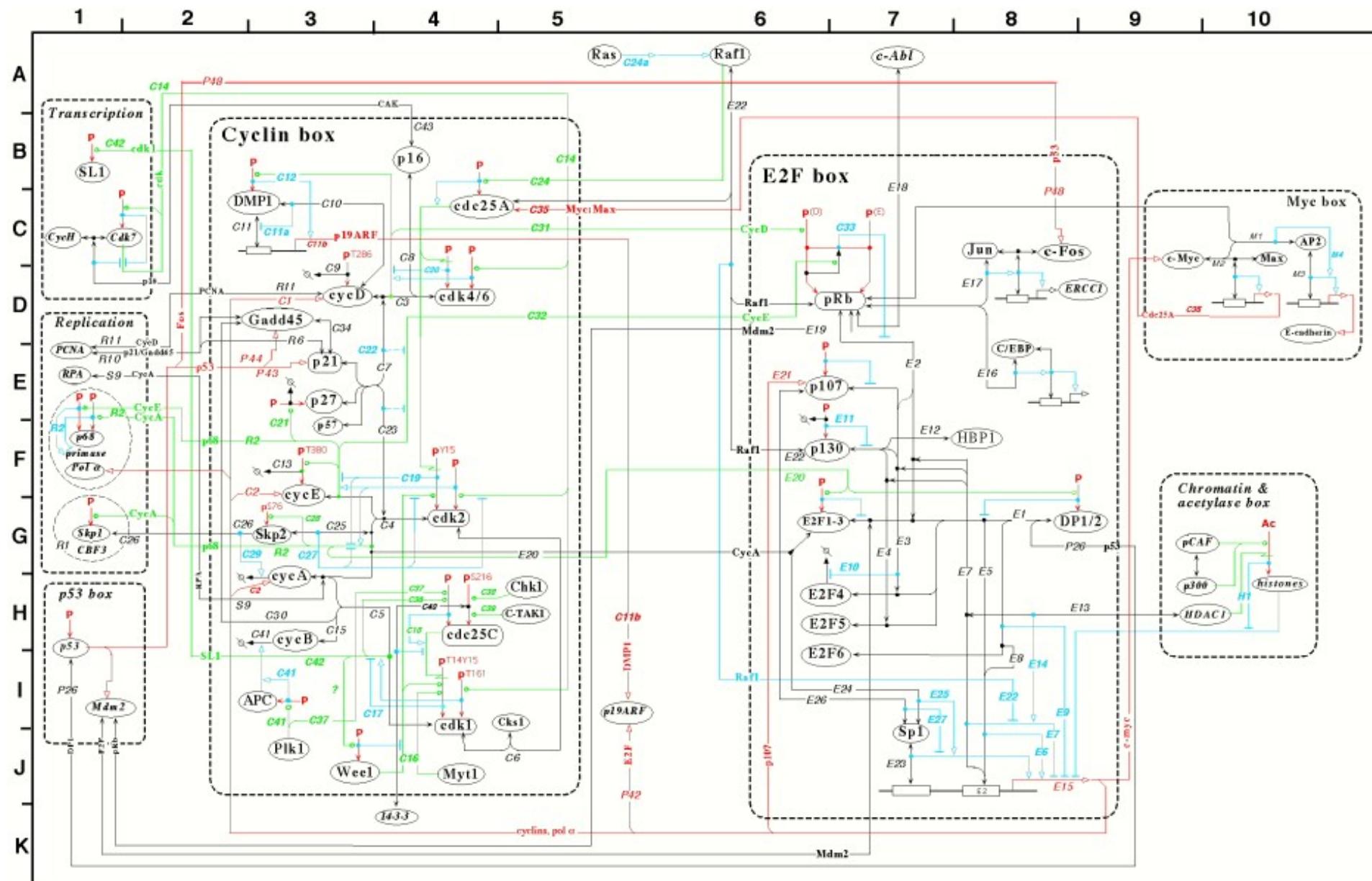
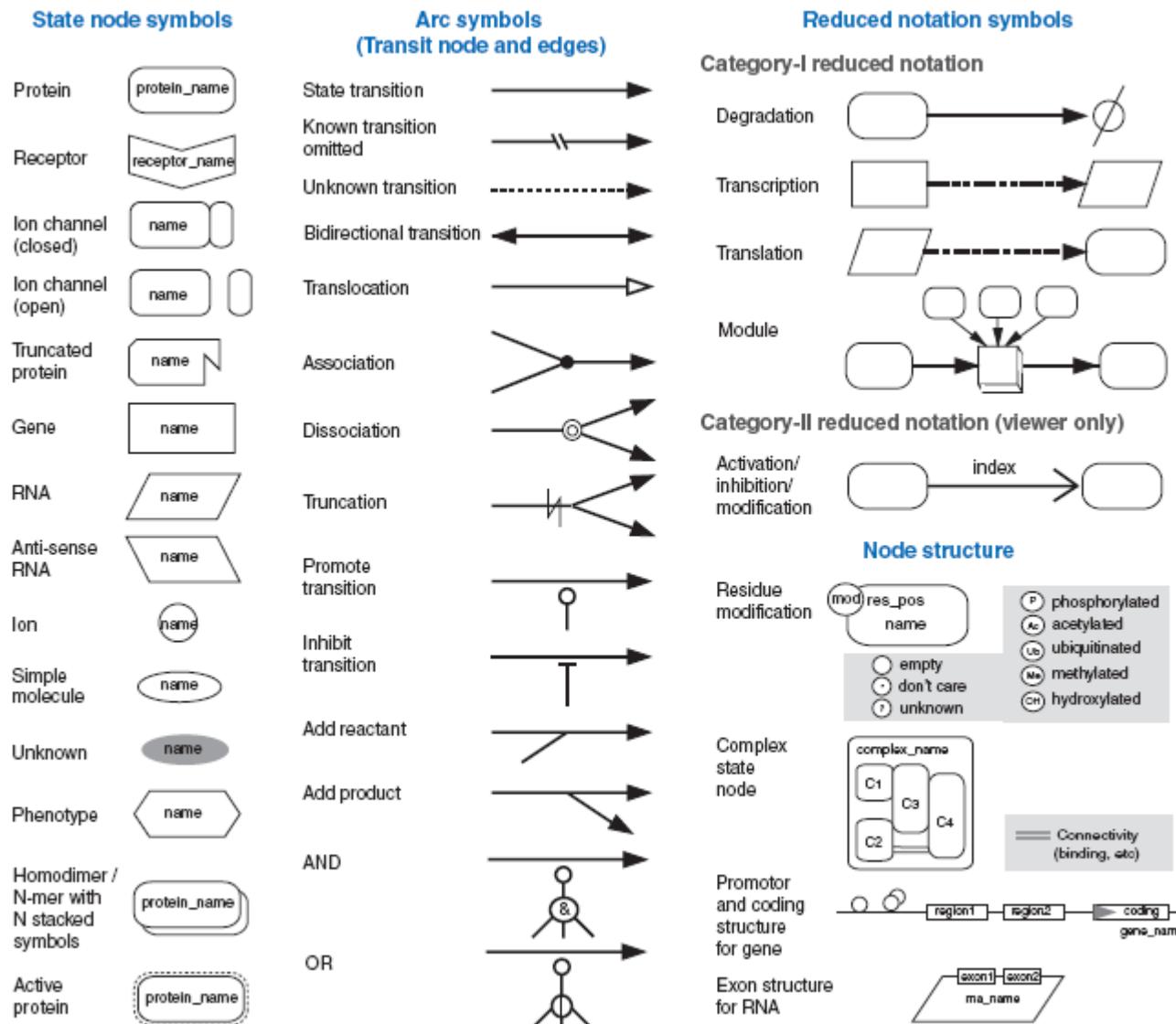


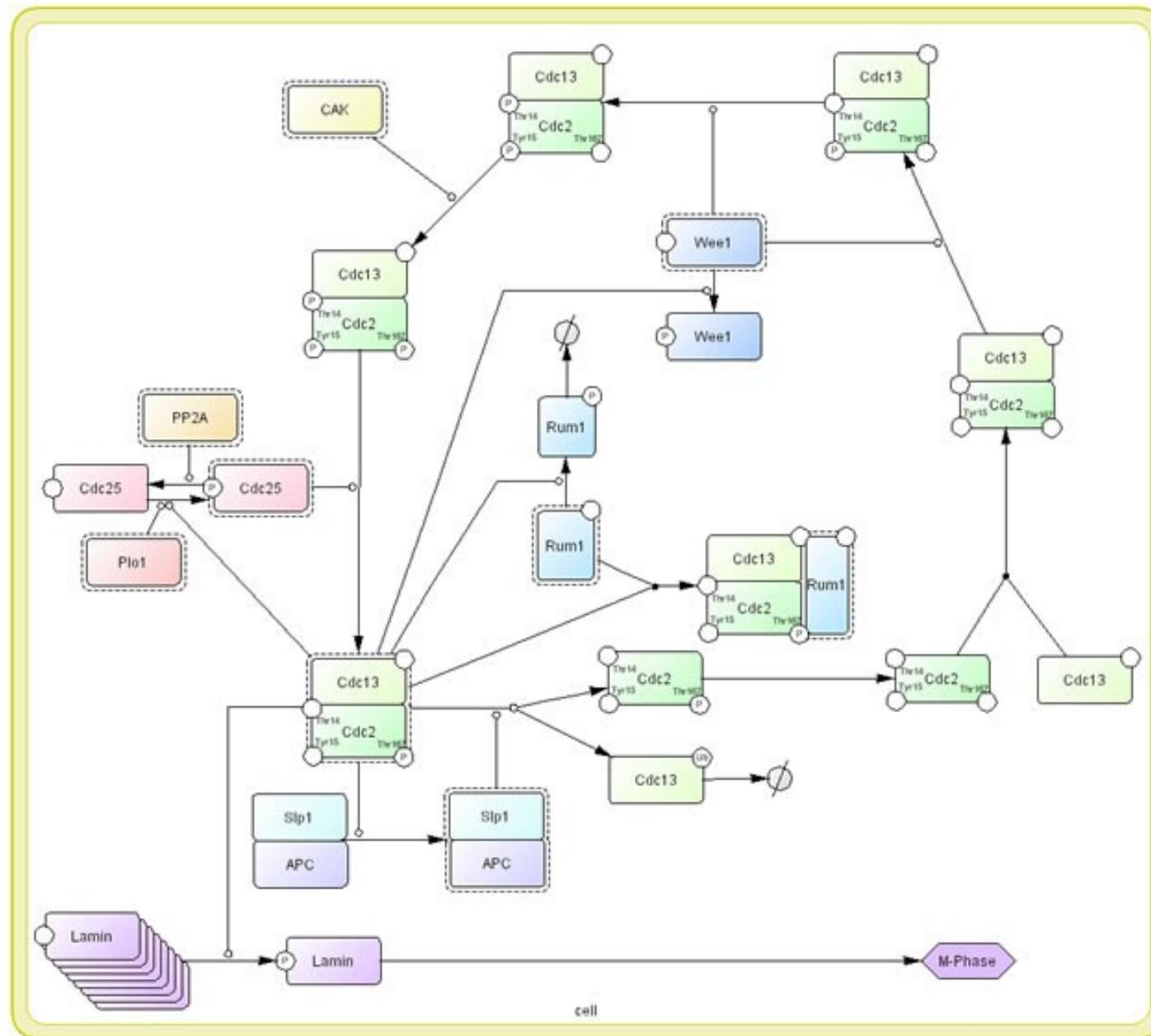
Figure 6A: The Cyclin - E2F cell cycle control system (version 3a - June 8, 1999)





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

Cell Cycle in Kitano's Process Diagram



- 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 Graphical Notation



<http://www.sbgn.org/>

- An unambiguous way of graphically describing and interpreting biochemical and cellular events
- Limited amount of symbols
Re-use existing symbols
 - Smooth learning curve
- Can represent logical or mechanistic models, biochemical pathways, at different levels of granularity
- Detailed technical specification, precise data-models and growing software support
- Initiated by Hiroaki Kitano. Developed over four years by a diverse community

The Systems Biology Graphical Notation

Nicolas Le Novère¹, Michael Hucka², Huaiyu Mi³, Stuart Moodie⁴, Falk Schreiber^{5,6}, Anatoly Sorokin⁷, Emek Demir⁸, Katja Wegner⁹, Mirit I Aladjem¹⁰, Sarala M Wimalaratne¹¹, Frank T Bergman¹², Ralph Gauges¹³, Peter Ghazal^{4,14}, Hideya Kawaiji¹⁵, Lu Li¹, Yukiko Matsuoka¹⁶, Alice Villéger^{17,18}, Sarah E Boyd¹⁹, Laurence Calzone²⁰, Melanie Courtof²¹, Ugur Dogrusoz²², Tom C Freeman^{14,23}, Akira Funahashi²⁴, Samik Ghosh¹⁶, Akiya Jouraku²⁴, Sohyoung Kim¹⁰, Fedor Kolpakov^{25,26}, Augustin Luna¹⁰, Sven Sahle¹³, Esther Schmidt¹, Steven Watterson^{4,22}, Guanming Wu²⁷, Igor Goryanin⁴, Douglas B Kell^{18,28}, Chris Sander⁸, Herbert Sauro¹², Jacky L Snoep²⁹, Kurt Kohn¹⁰ & Hiroaki Kitano^{16,30,31}

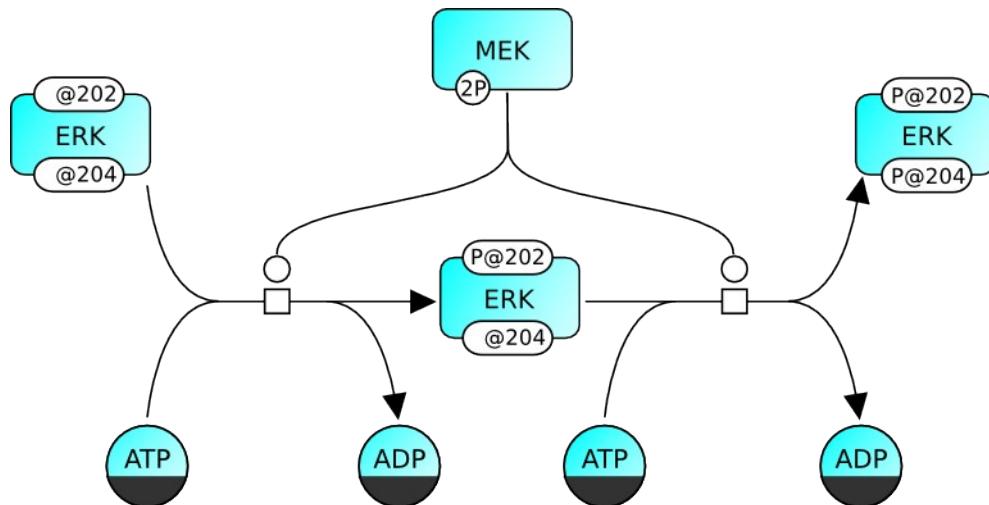
¹EMBL European Bioinformatics Institute, Hinxton, UK. ²Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA. ³SRI International, Menlo Park, California, USA. ⁴Centre for Systems Biology at Edinburgh, University of Edinburgh, Edinburgh, UK. ⁵Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. ⁶Institute of Computer Science, University of Halle, Halle, Germany. ⁷School of Informatics, University of Edinburgh, Edinburgh, UK. ⁸Memorial Sloan Kettering Cancer Center - Computational Biology Center, New York, NY, USA. ⁹Science and Technology Research Institute, University of Hertfordshire, Hatfield, UK. ¹⁰National Cancer Institute, Bethesda, Maryland, USA. ¹¹Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. ¹²Department of Bioengineering, University of Washington, Seattle, Washington, USA. ¹³BIOQUANT, University of Heidelberg, Heidelberg, Germany. ¹⁴Division of Pathway Medicine, University of Edinburgh Medical School, Edinburgh, UK. ¹⁵Riken OMICS Science Center, Yokohama City, Kanagawa, Japan. ¹⁶The Systems Biology Institute, Tokyo, Japan. ¹⁷School of Computer Science, University of Manchester, Manchester, UK. ¹⁸Manchester Interdisciplinary Biocentre, Manchester, UK. ¹⁹Clayton School of Information Technology, Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia. ²⁰U900 INSERM, Paris Mines Tech, Institut Curie, Paris, France. ²¹Terry Fox Laboratory, British Columbia Cancer Research Center, Vancouver, British Columbia, Canada. ²²Bilkent Center for Bioinformatics, Bilkent University, Ankara, Turkey. ²³The Roslin Institute, University of Edinburgh, Midlothian, UK. ²⁴Department of Biosciences and Informatics, Keio University, Hiyoshi, Kouhoku-ku, Yokohama, Japan. ²⁵Institute of Systems Biology, Novosibirsk, Russia. ²⁶Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia. ²⁷Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ²⁸School of Chemistry, University of Manchester, Manchester, UK. ²⁹Department of Biochemistry, Stellenbosch University, Matieland, South Africa. ³⁰Sony Computer Science Laboratories, Tokyo, Japan. ³¹Okinawa Institute of Science and Technology, Okinawa, Japan. Correspondence should be addressed to N.L.N. (lenov@ebi.ac.uk).

39 authors, 31 affiliations



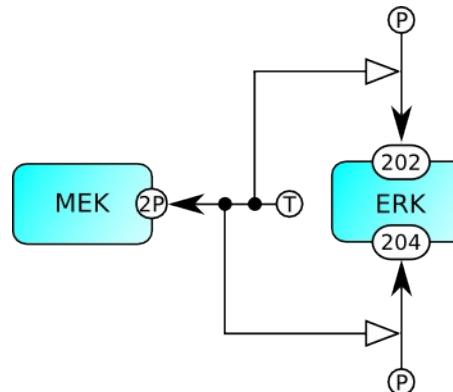
Graph trinity: three languages in one notation

Process Descriptions



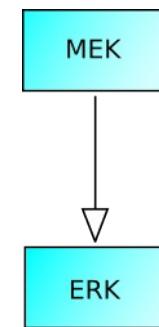
- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships



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

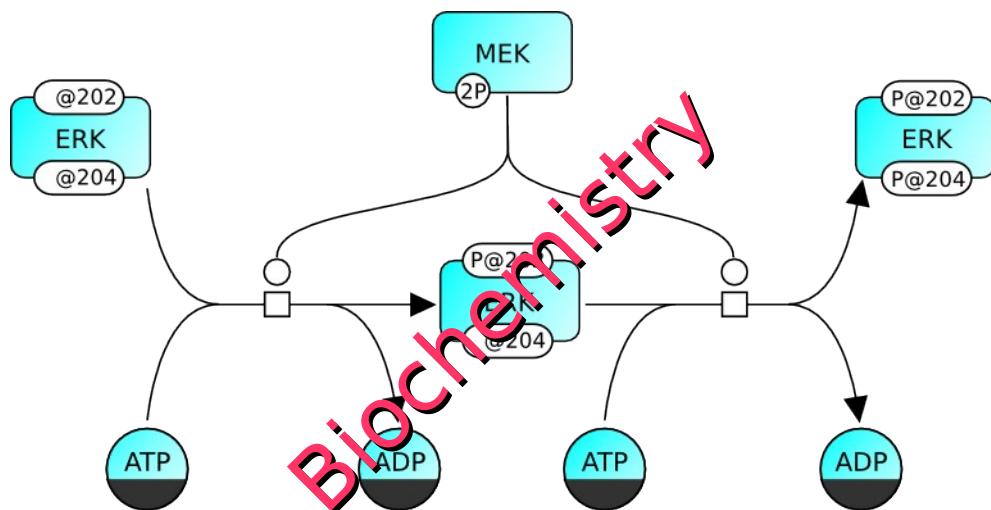
Activity Flows



- Ambiguous
- Conceptual
- Sequential

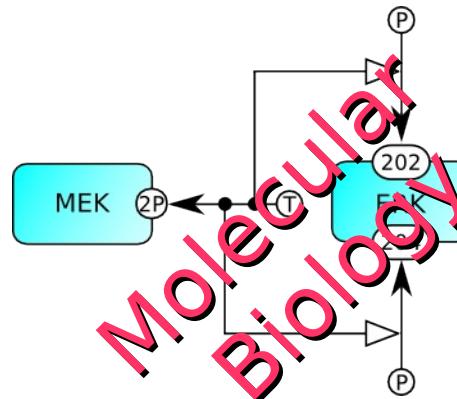
Graph trinity: three languages in one notation

Process Descriptions



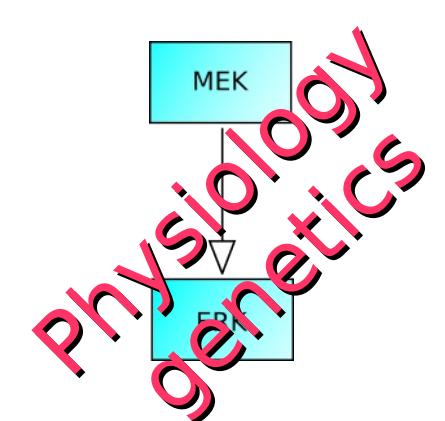
- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships



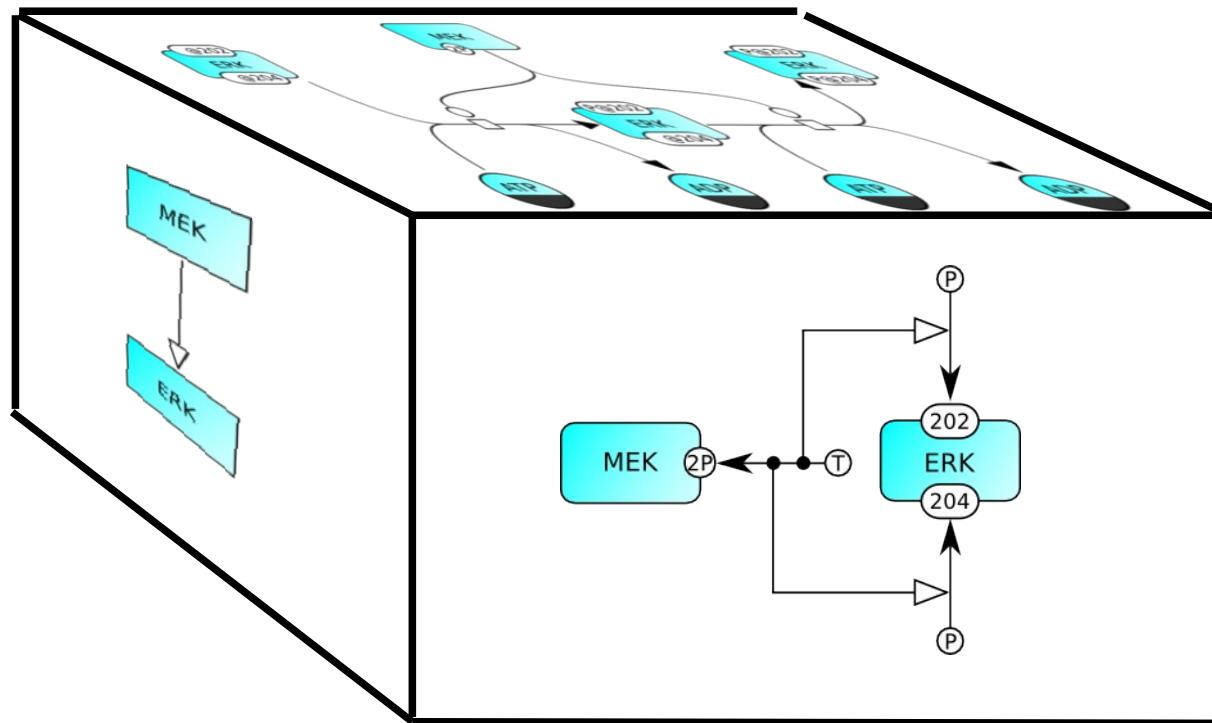
- Unambiguous
- Mechanistic
- Non-sequential
- Independence of relationships

Activity Flows

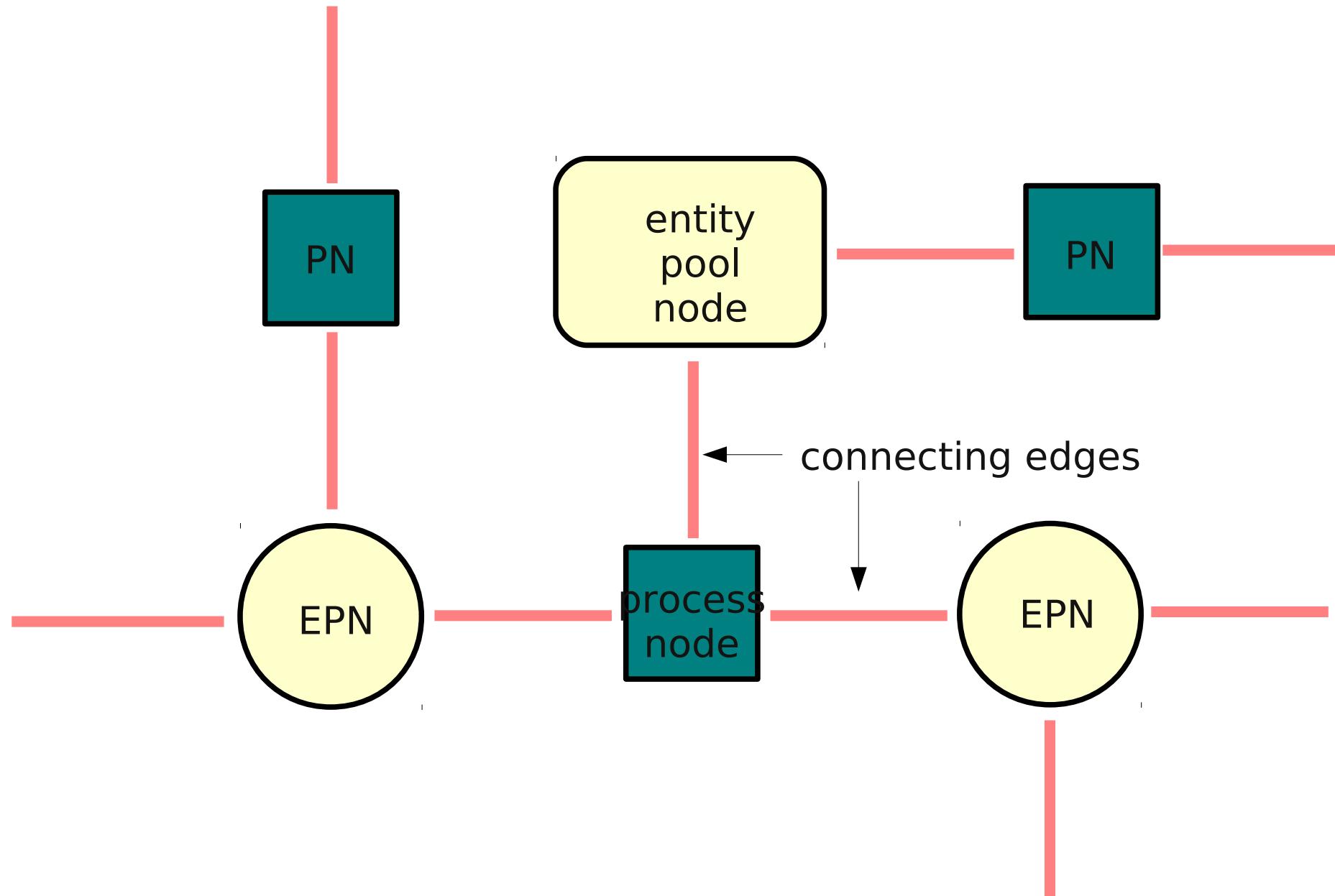


- Ambiguous
- Conceptual
- Sequential

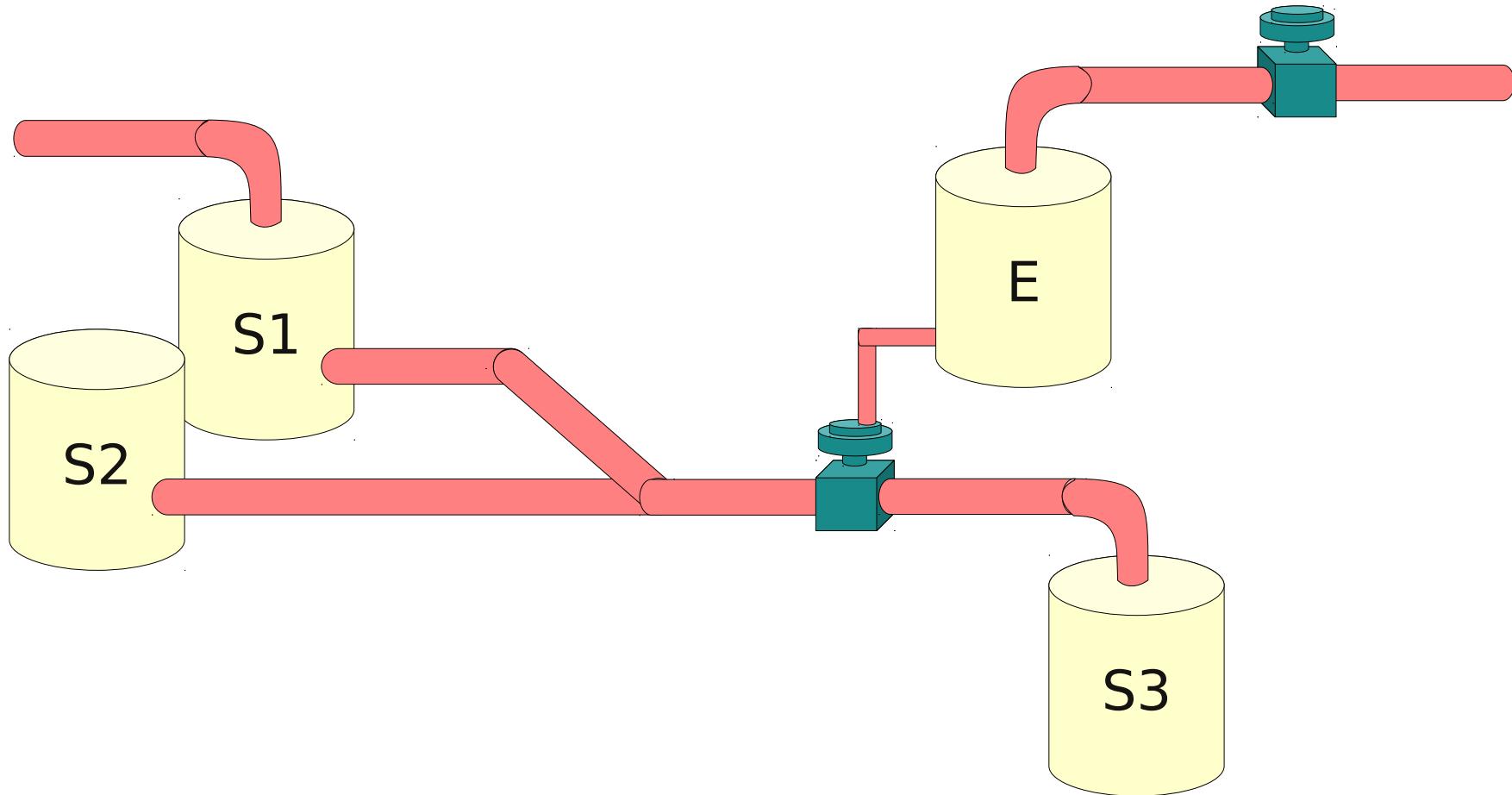
Three orthogonal projections of biology

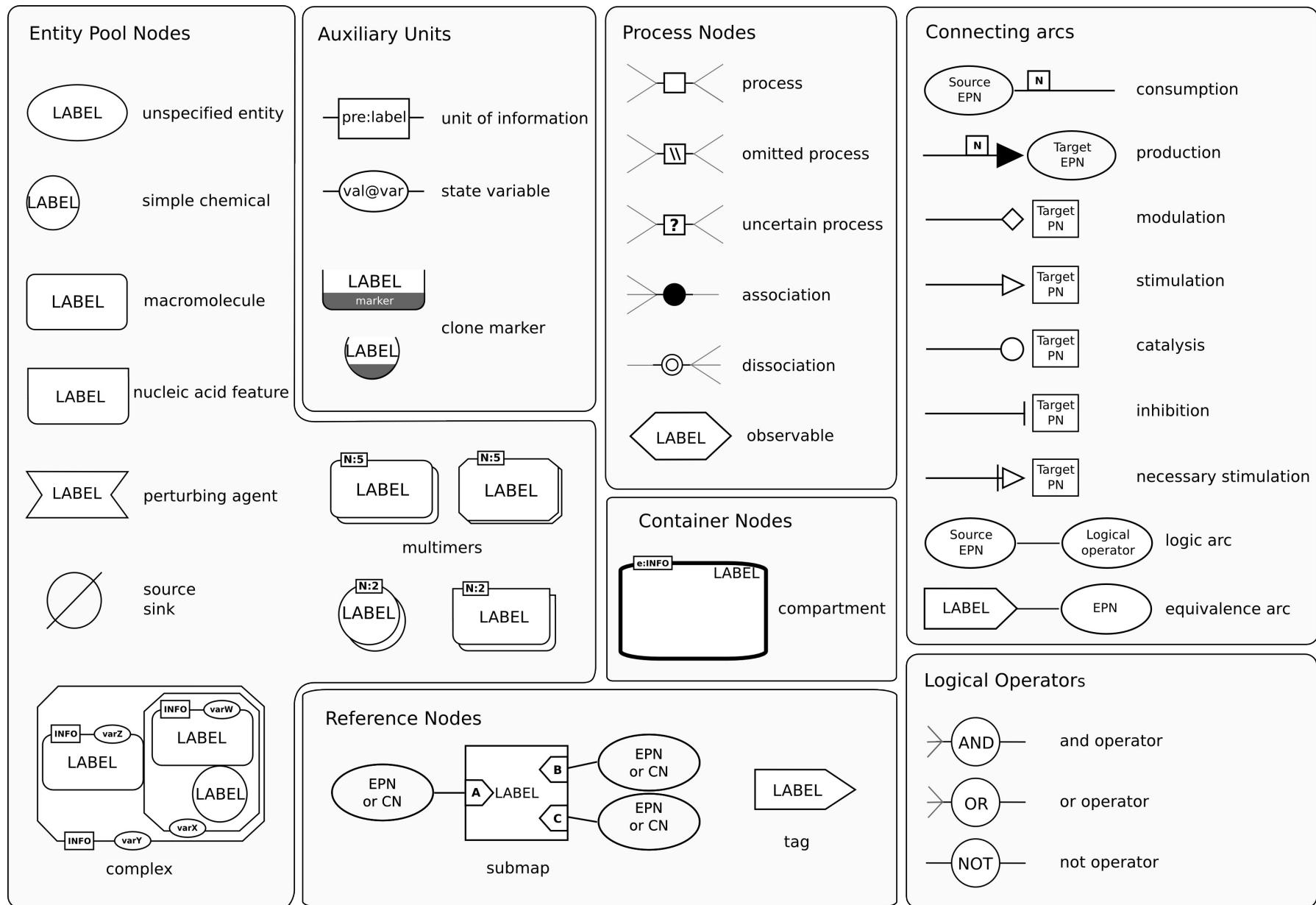


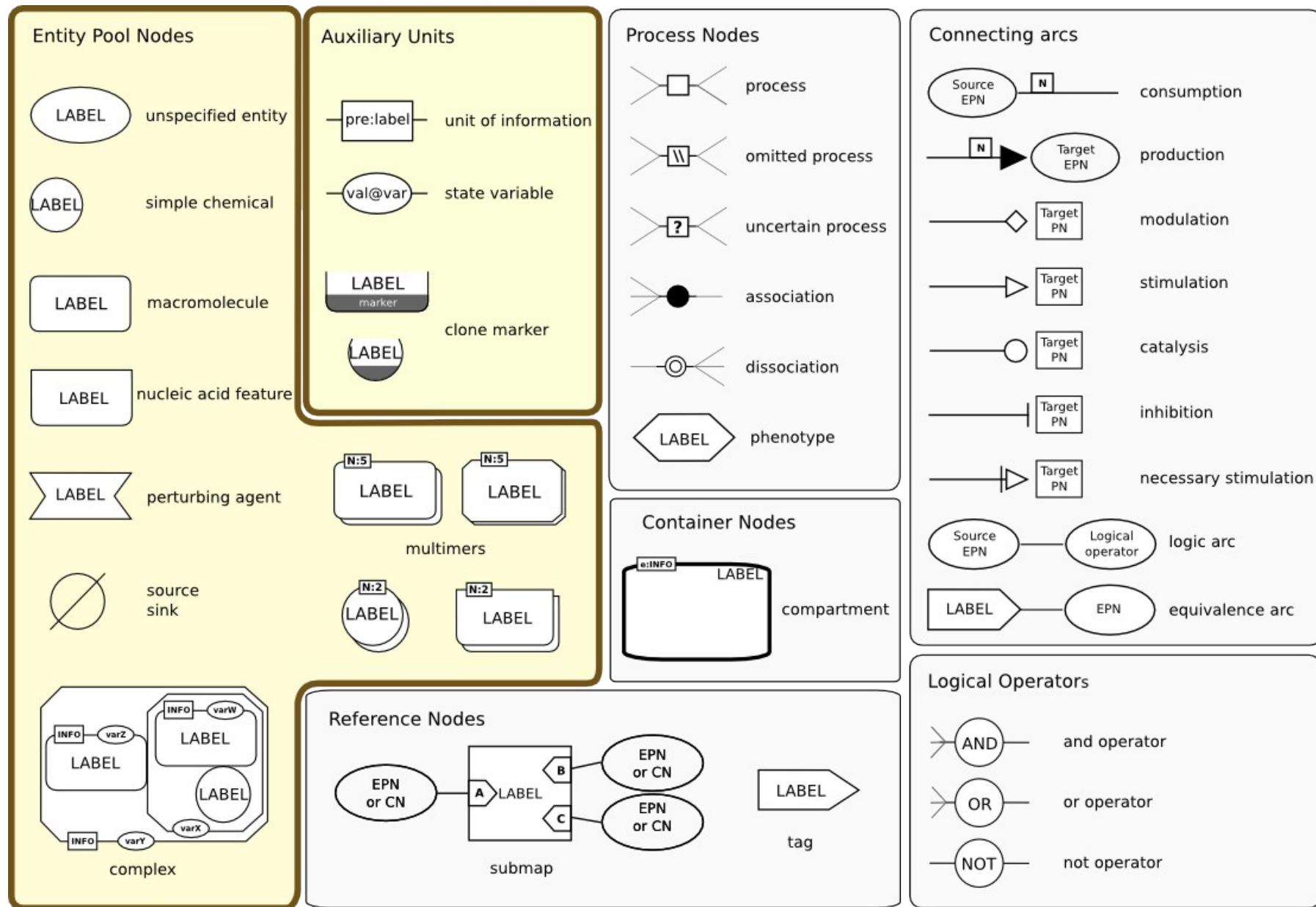
Process Descriptions are bipartite graphs

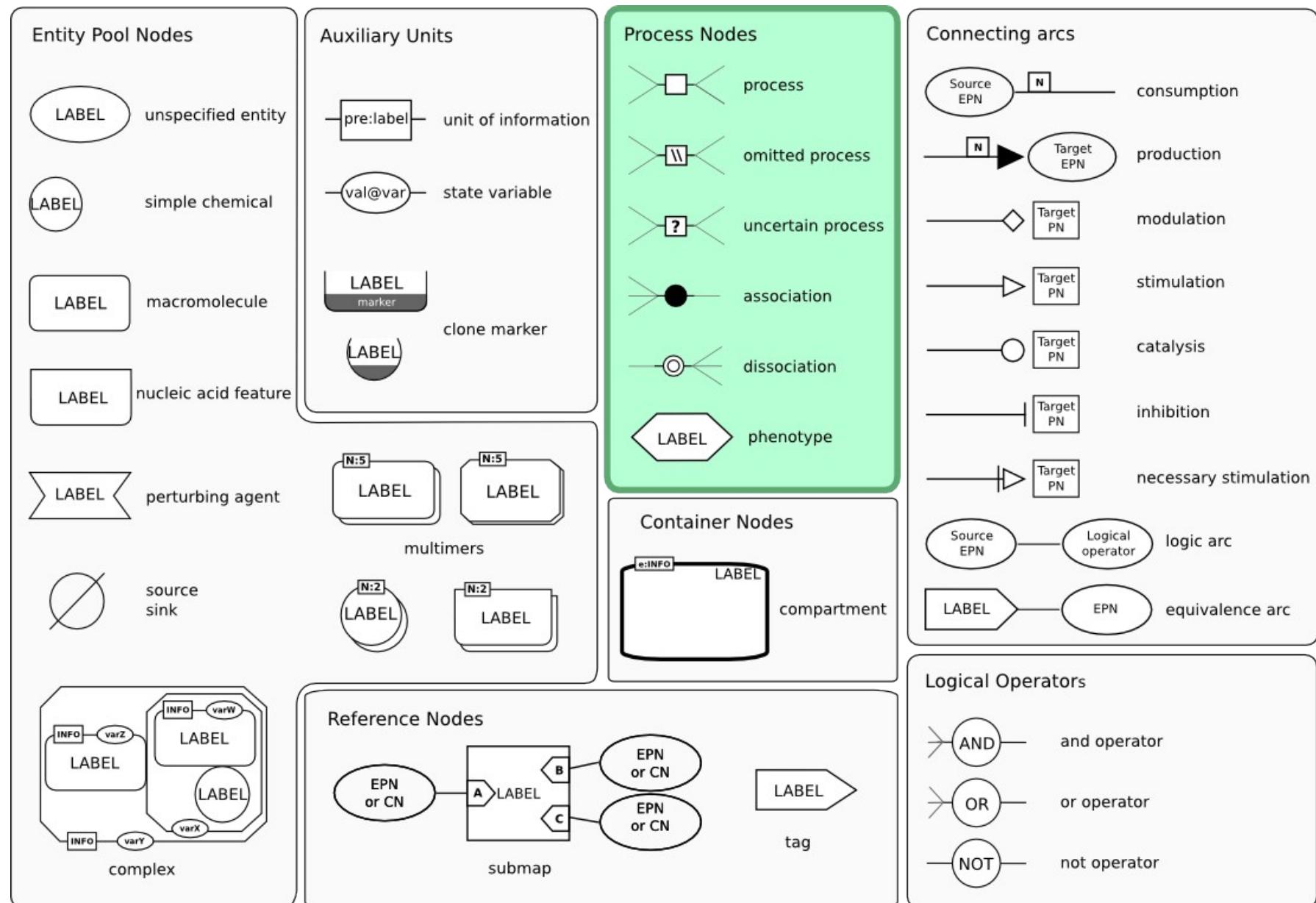


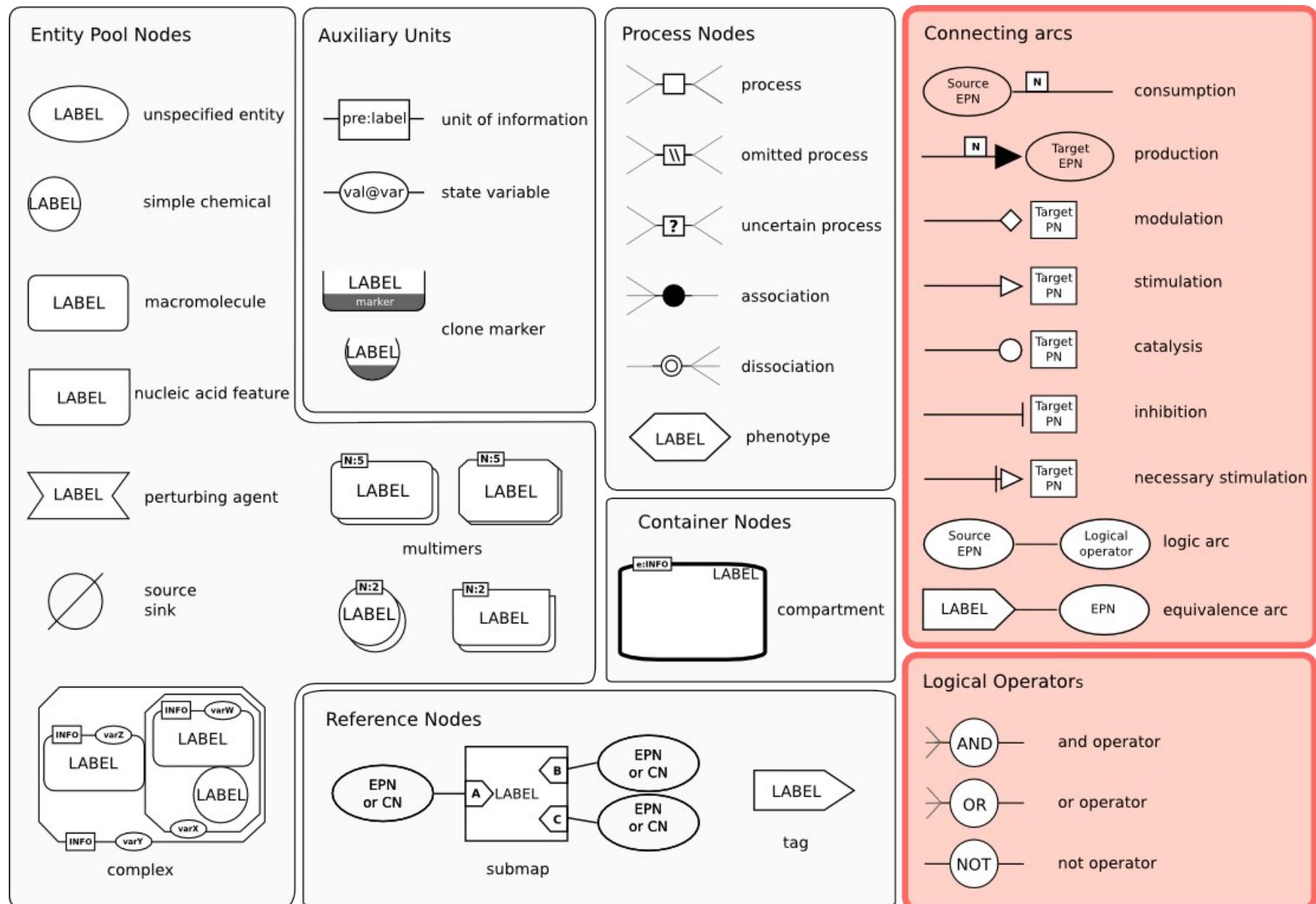
Process Descriptions can be viewed as pipelines



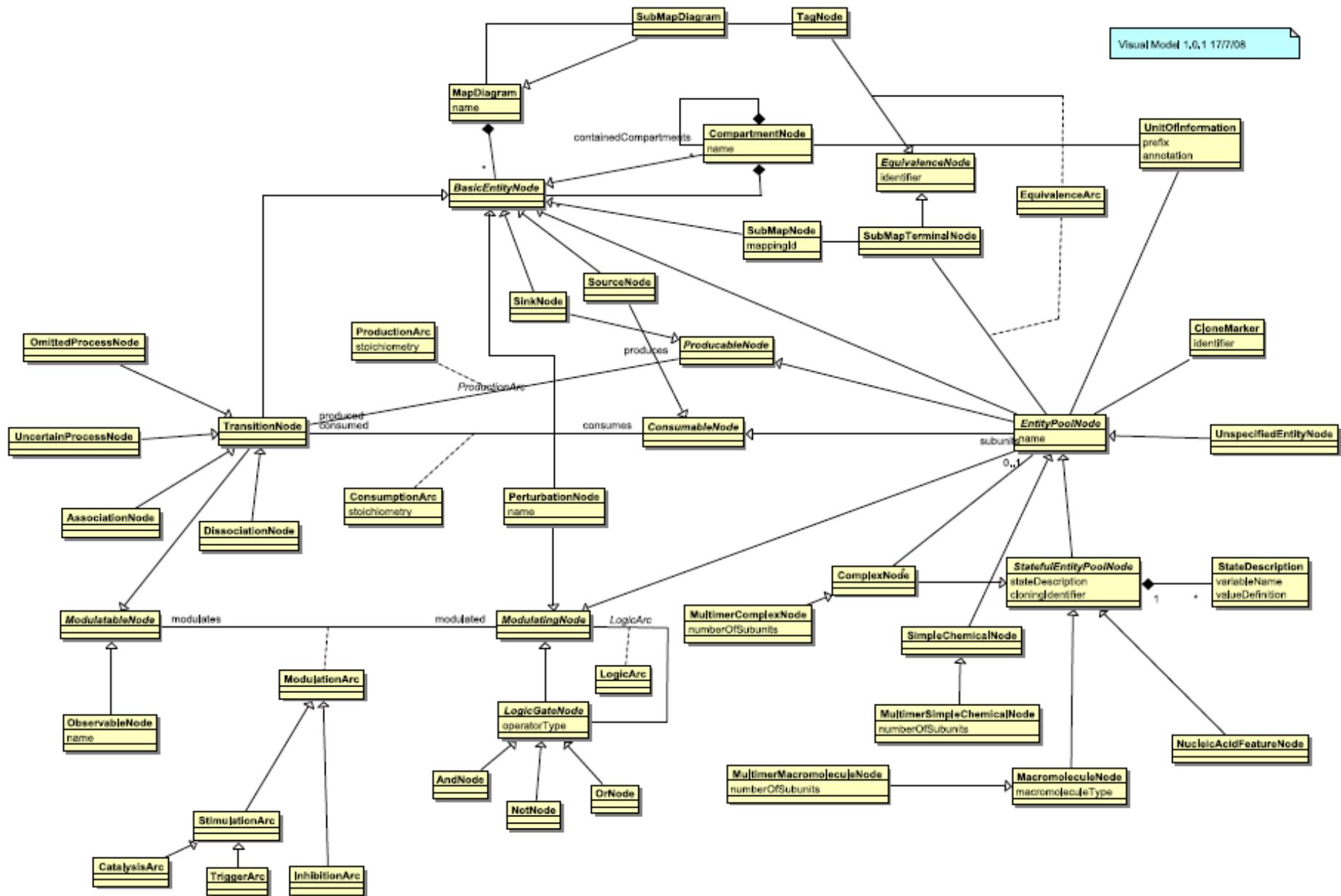








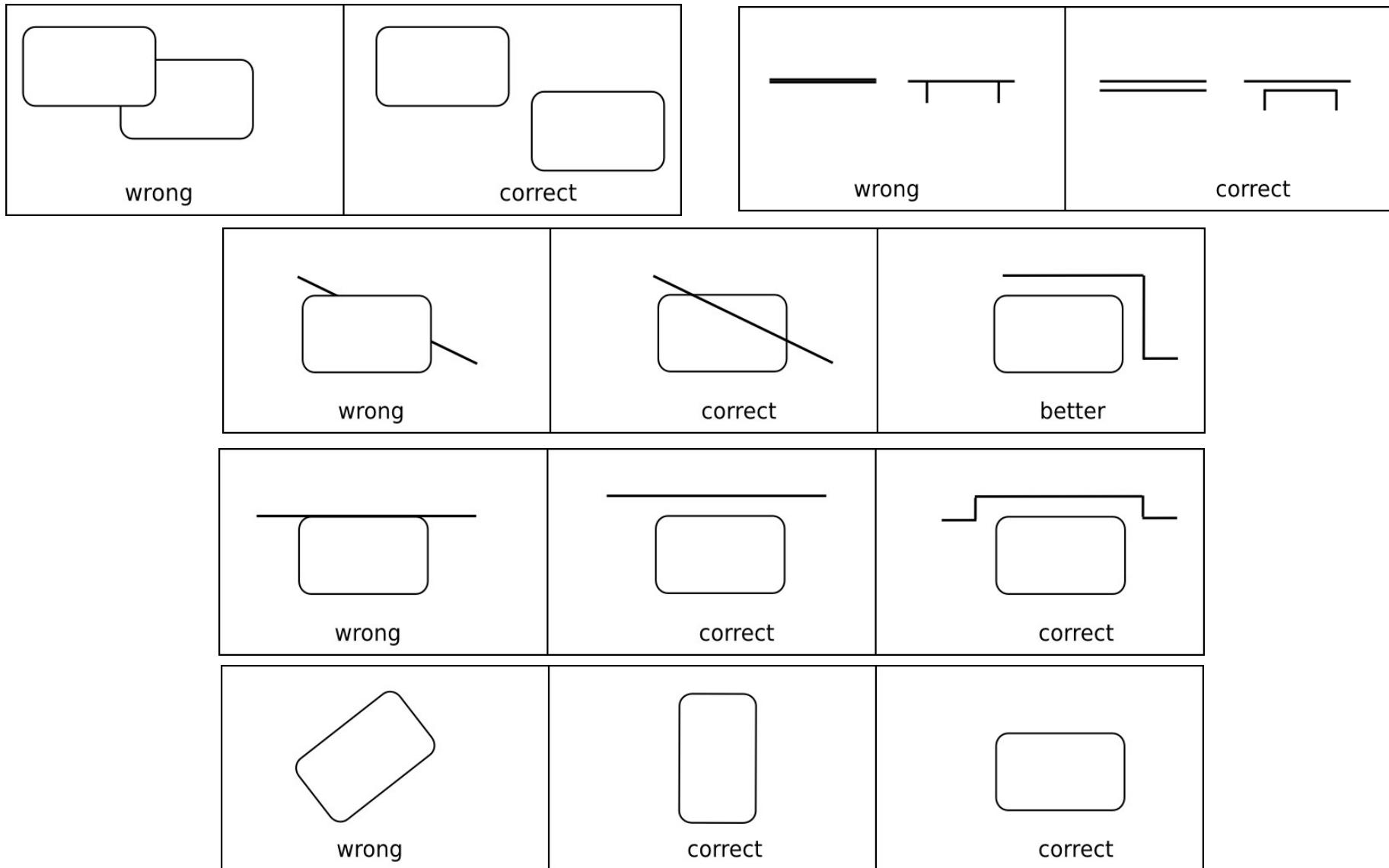
Process Descriptions data model

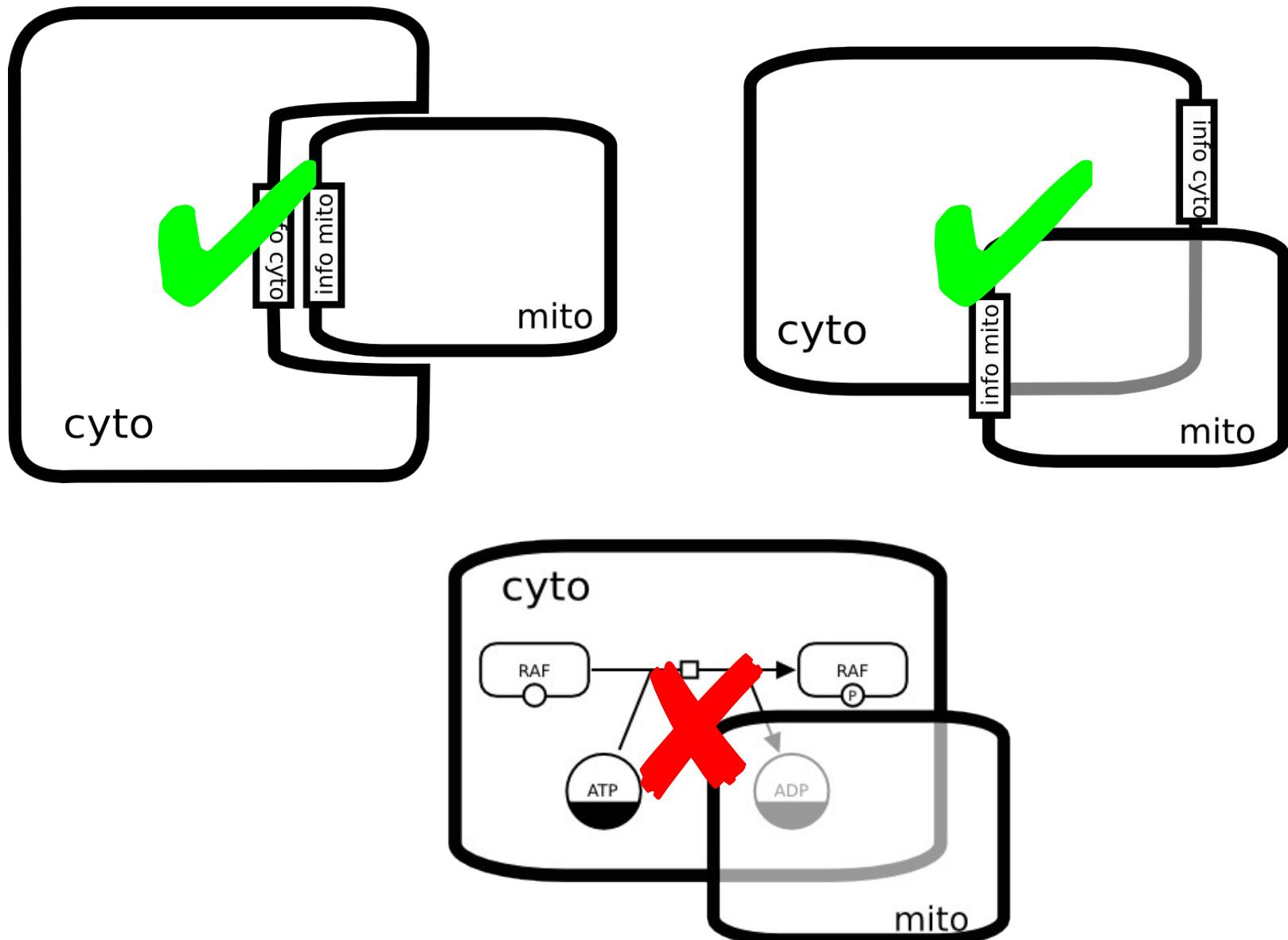


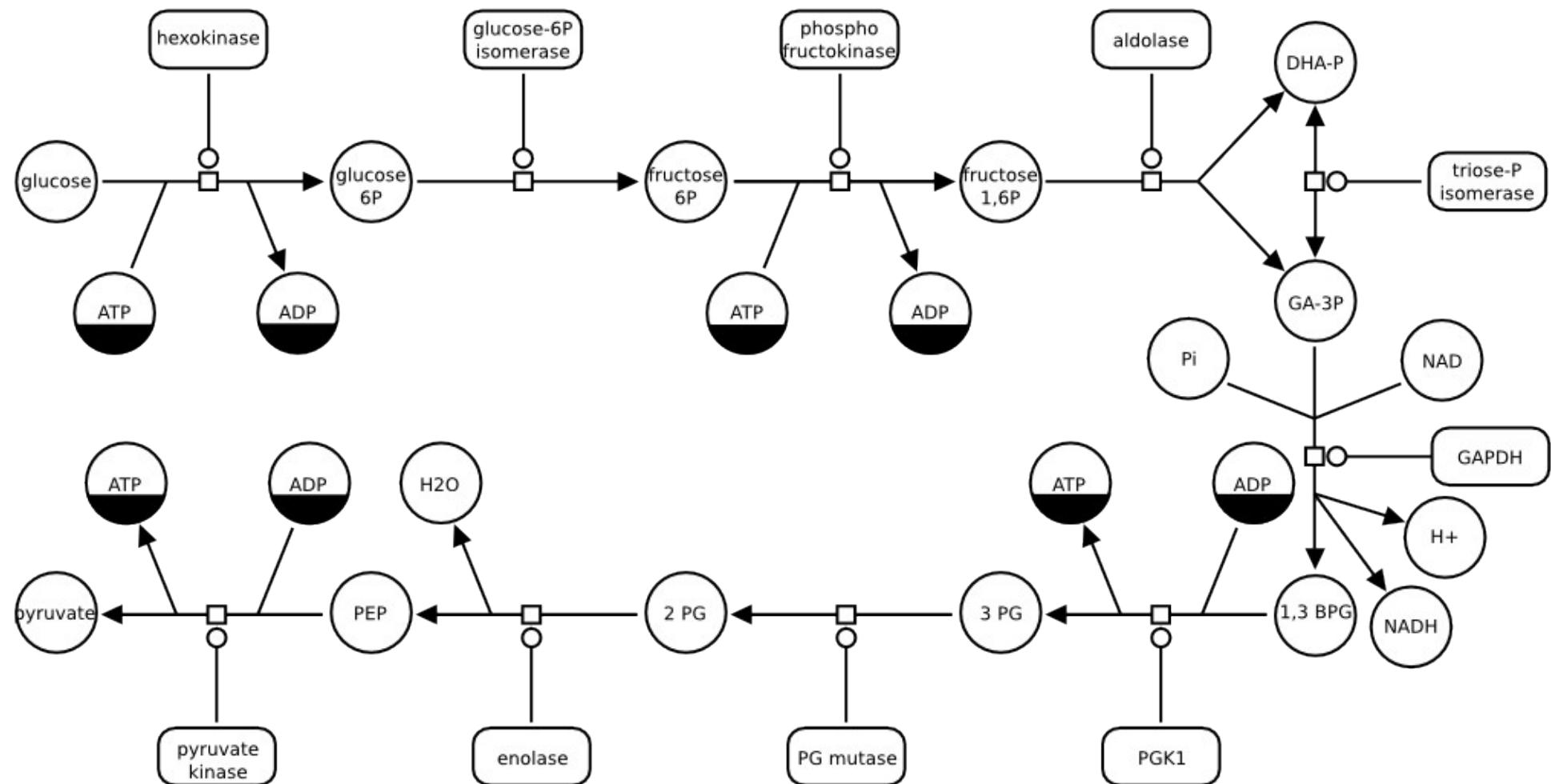
Process Descriptions syntax definition

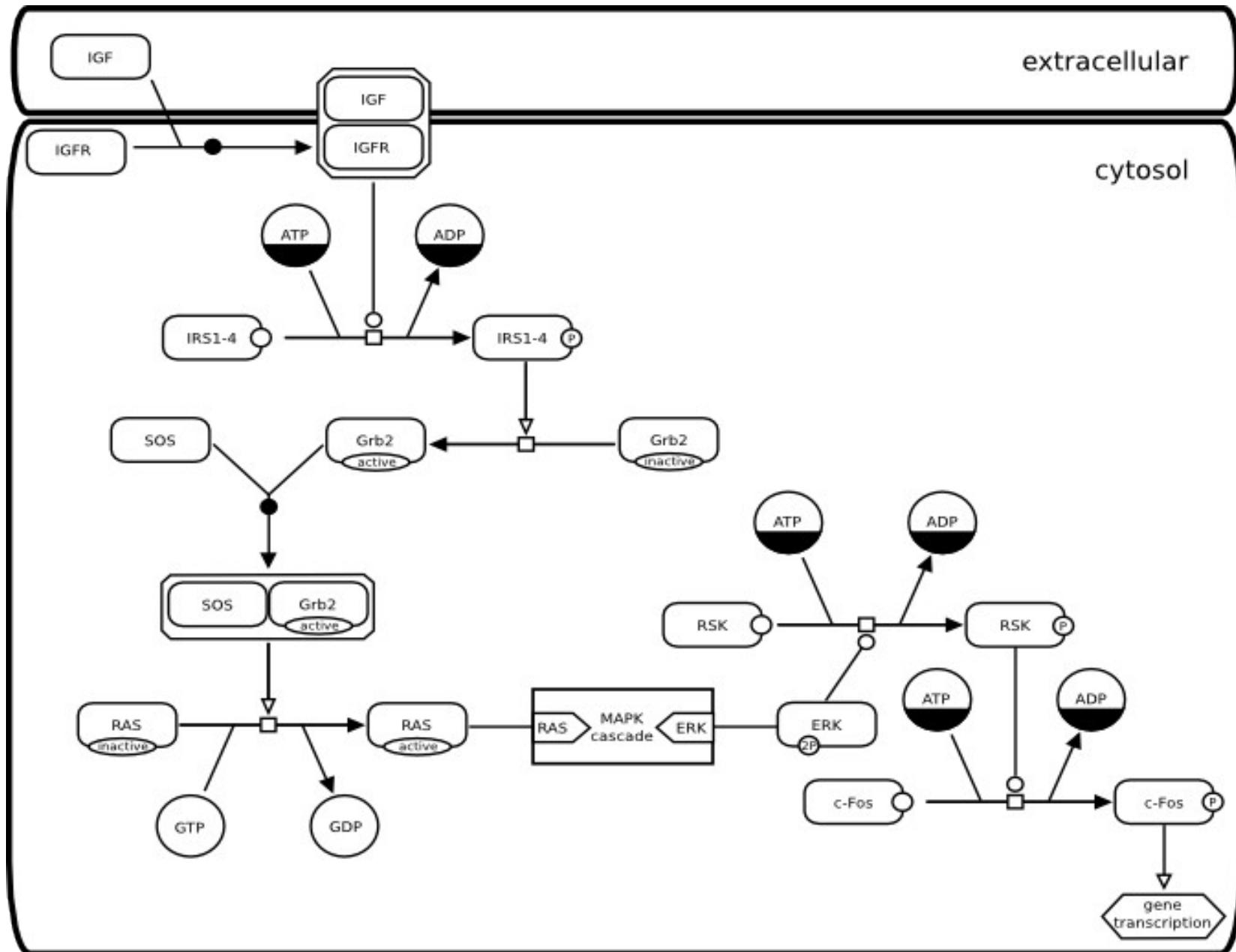
	<i>Arc\EPN</i>						
	<i>macromolecule</i>	<i>simple chemical</i>	<i>unspecified entity</i>	<i>multimer</i>	<i>complex</i>	<i>nucleic acid feature</i>	
<i>consumption</i>	I	I	I	I	I	I	
<i>production</i>	O	O	O	O	O	O	
<i>modulation</i>	I	I	I	I	I	I	O
<i>stimulation</i>	I	I	I	I	I	I	O
<i>catalysis</i>	I	I	I	I	I	I	O
<i>inhibition</i>	I	I	I	I	I	I	O
<i>trigger</i>	I	I	I	I	I	I	O
<i>logic arc</i>	I	I	I	I	I	I	
<i>equivalence arc</i>	I	I	I	I	I	I	O

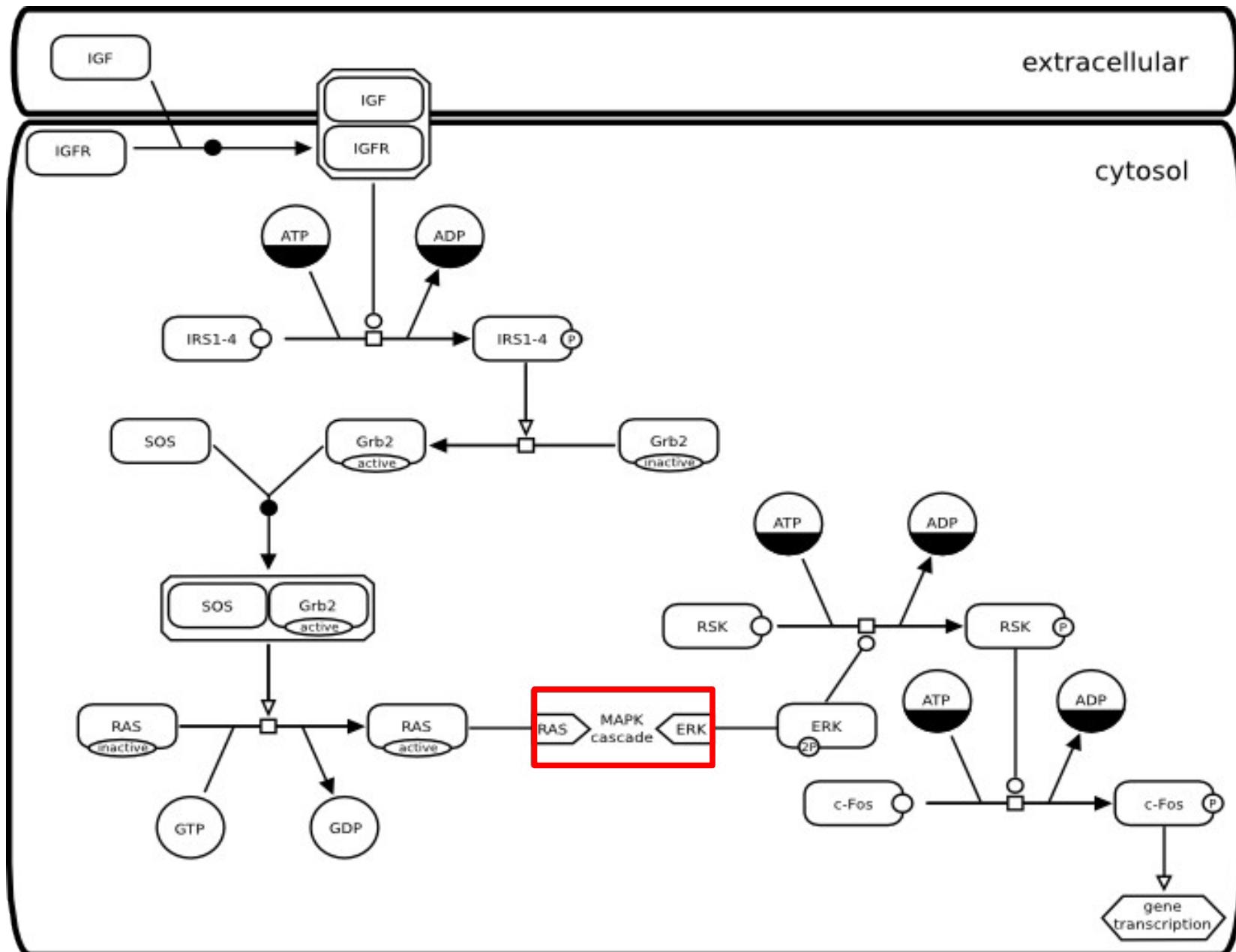
	<i>Arc\PN</i>							
	<i>transition</i>	<i>omitted process</i>	<i>uncertain process</i>	<i>association</i>	<i>dissociation</i>	<i>and</i>	<i>or</i>	<i>not</i>
<i>consumption</i>	O	O	O	O	O(1)			
<i>production</i>	I	I	I	I(1)	I			
<i>modulation</i>	O	O	O			I(1)	I(1)	I(1)
<i>stimulation</i>	O	O	O			I(1)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

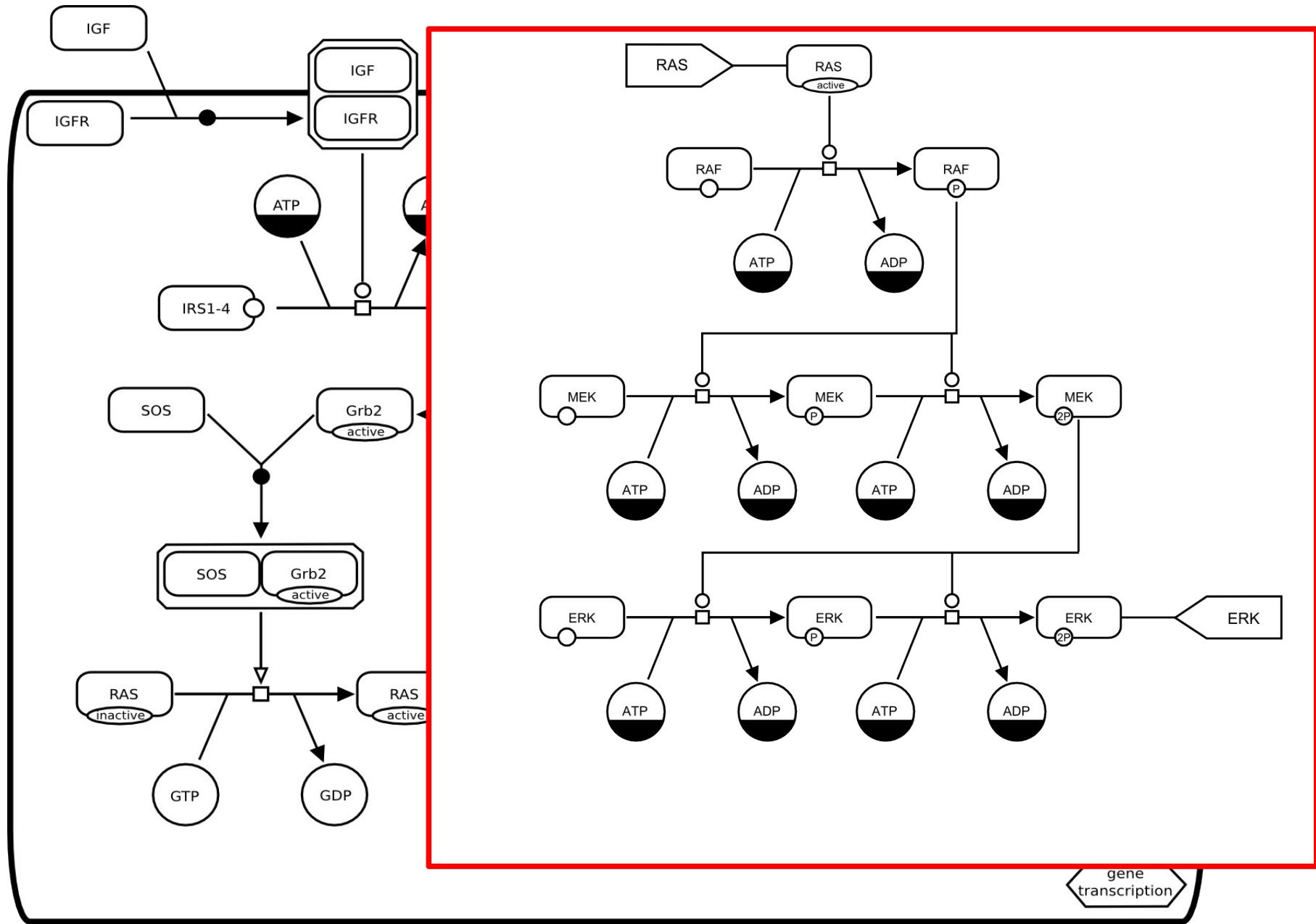


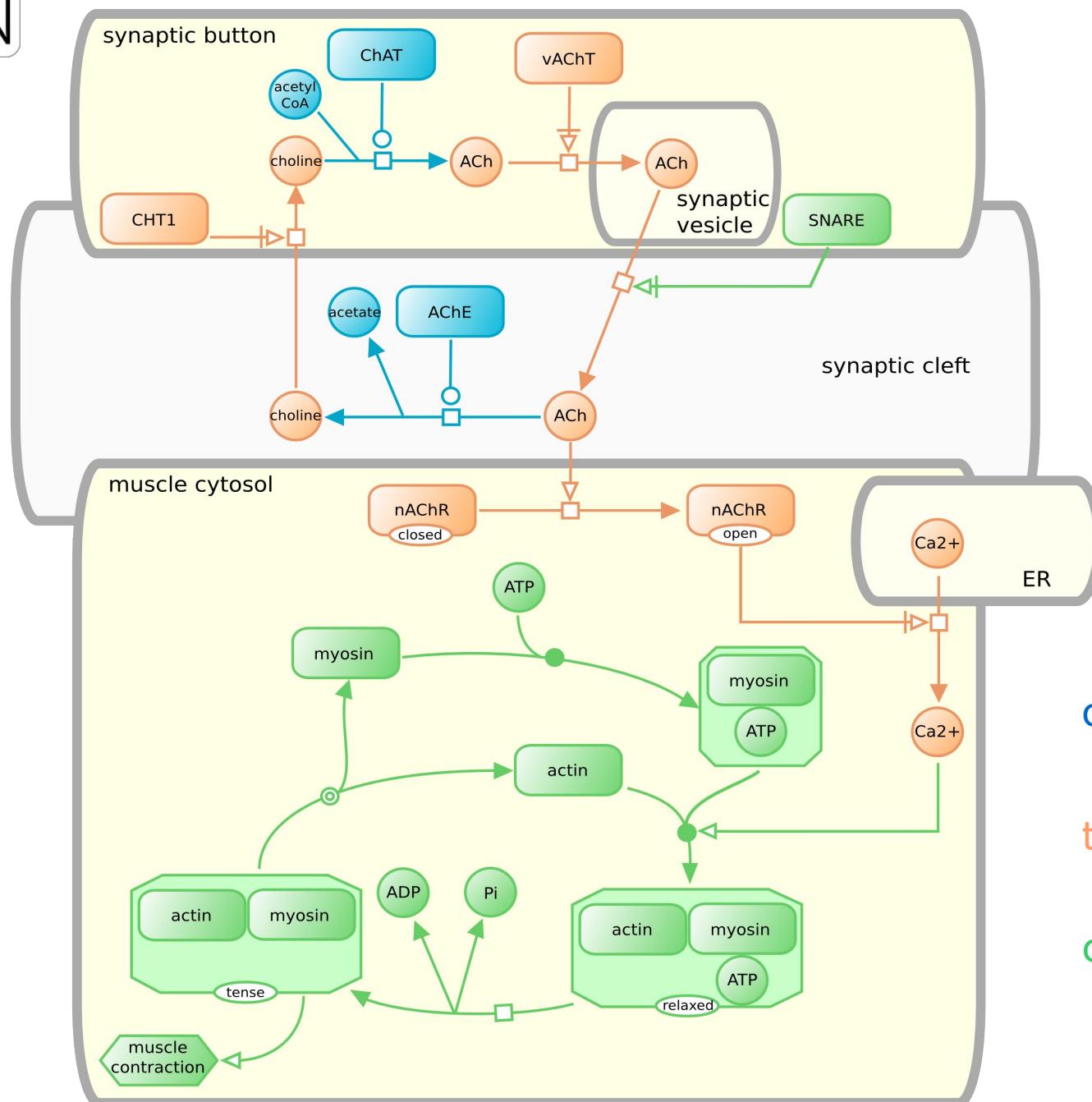






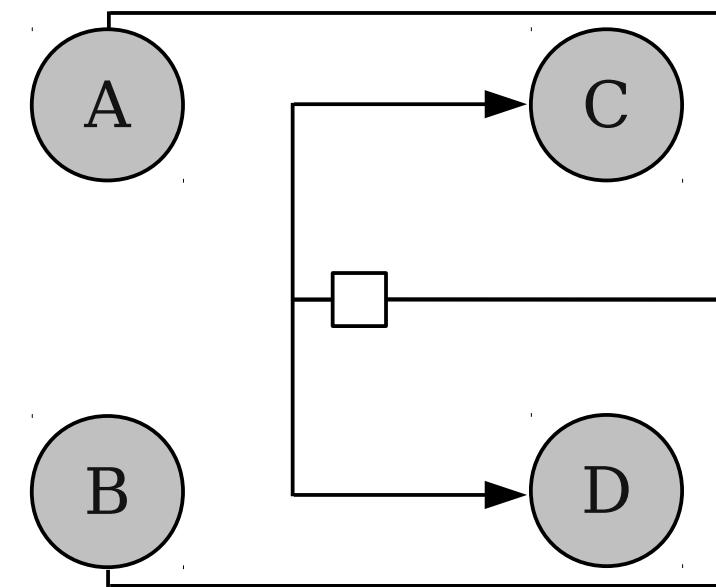
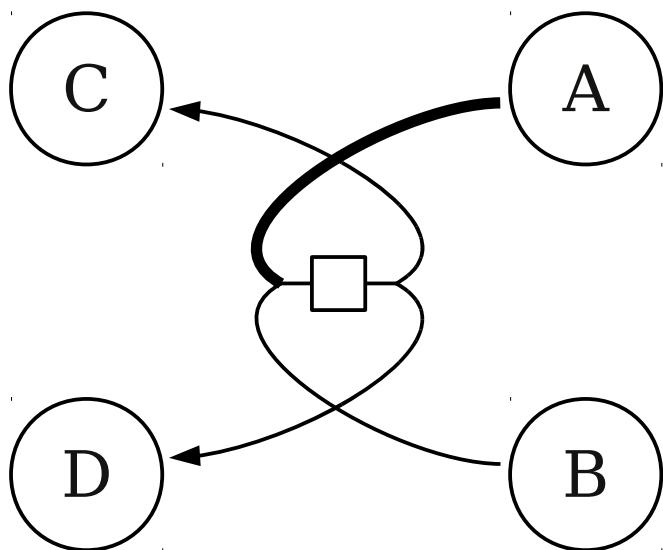
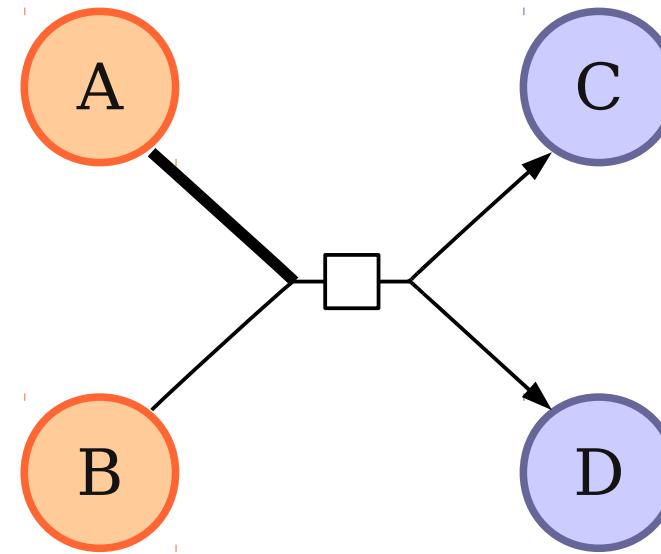
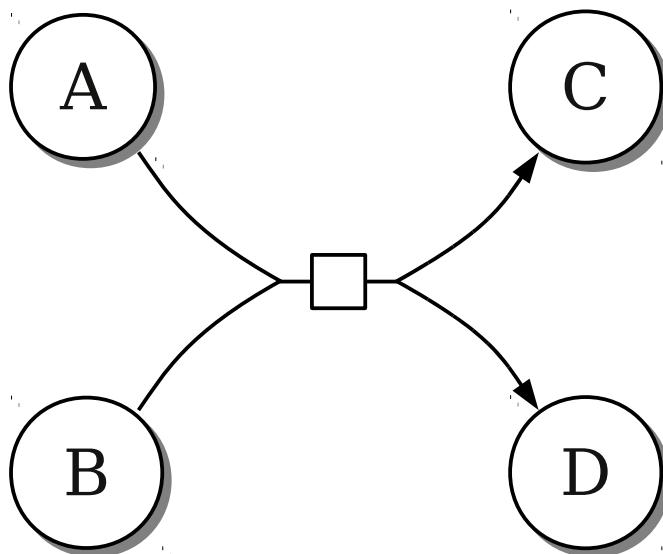






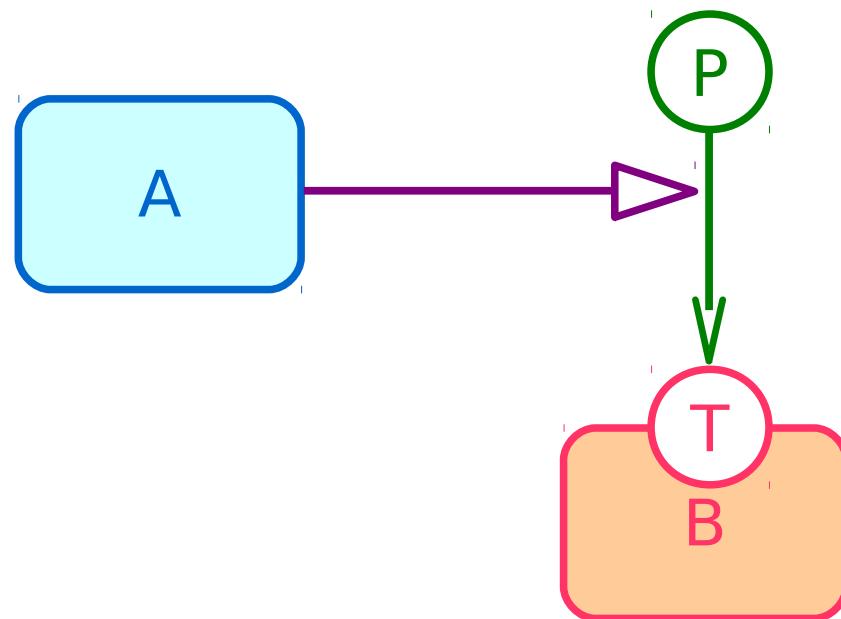
catalytic processes
transport processes
contractile proteins

All those diagrams are identical

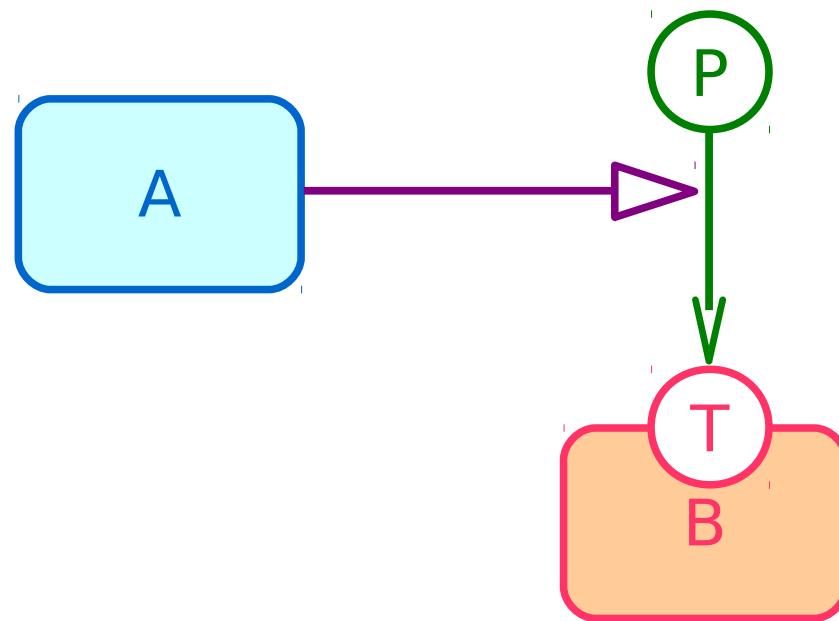


- Generics, i.e, entity pools representing several possible biochemical types. (e.g. MAPK instead of ERK1 and ERK2).
- Trans-compartment (e.g. transmembrane) structures
- Logical combination of state-variable values (and close-world/open-world position)
- Moving and transforming compartments
- Non-chemical entity pool nodes (“voltage”, “pH” ...)

Entity Relationships can be viewed as rules

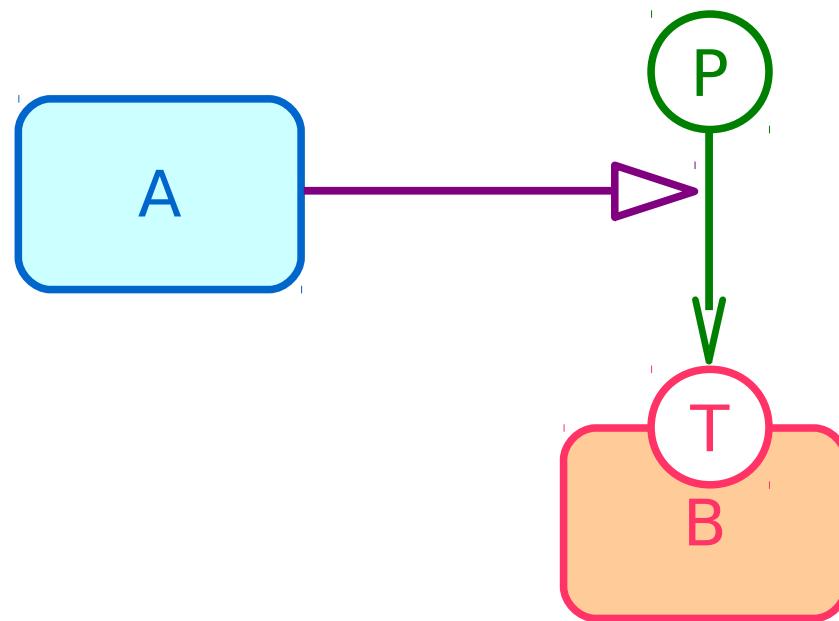


Entity Relationships can be viewed as rules



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

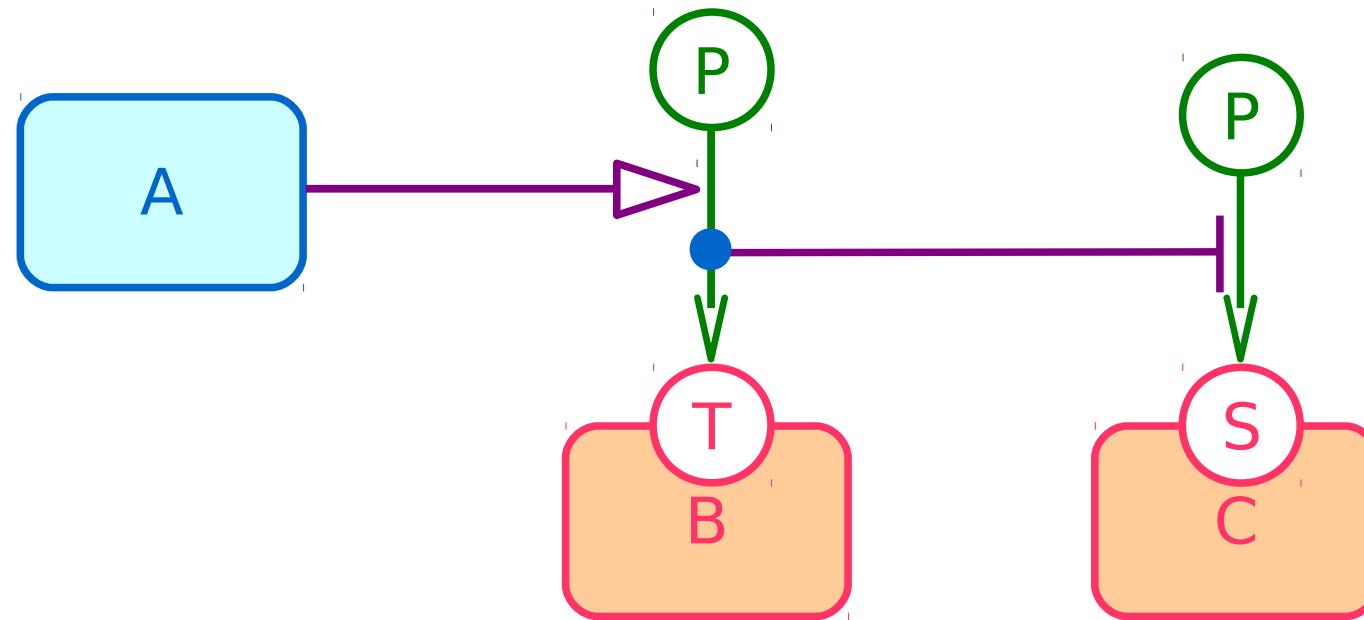
Entity Relationships can be viewed as rules



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

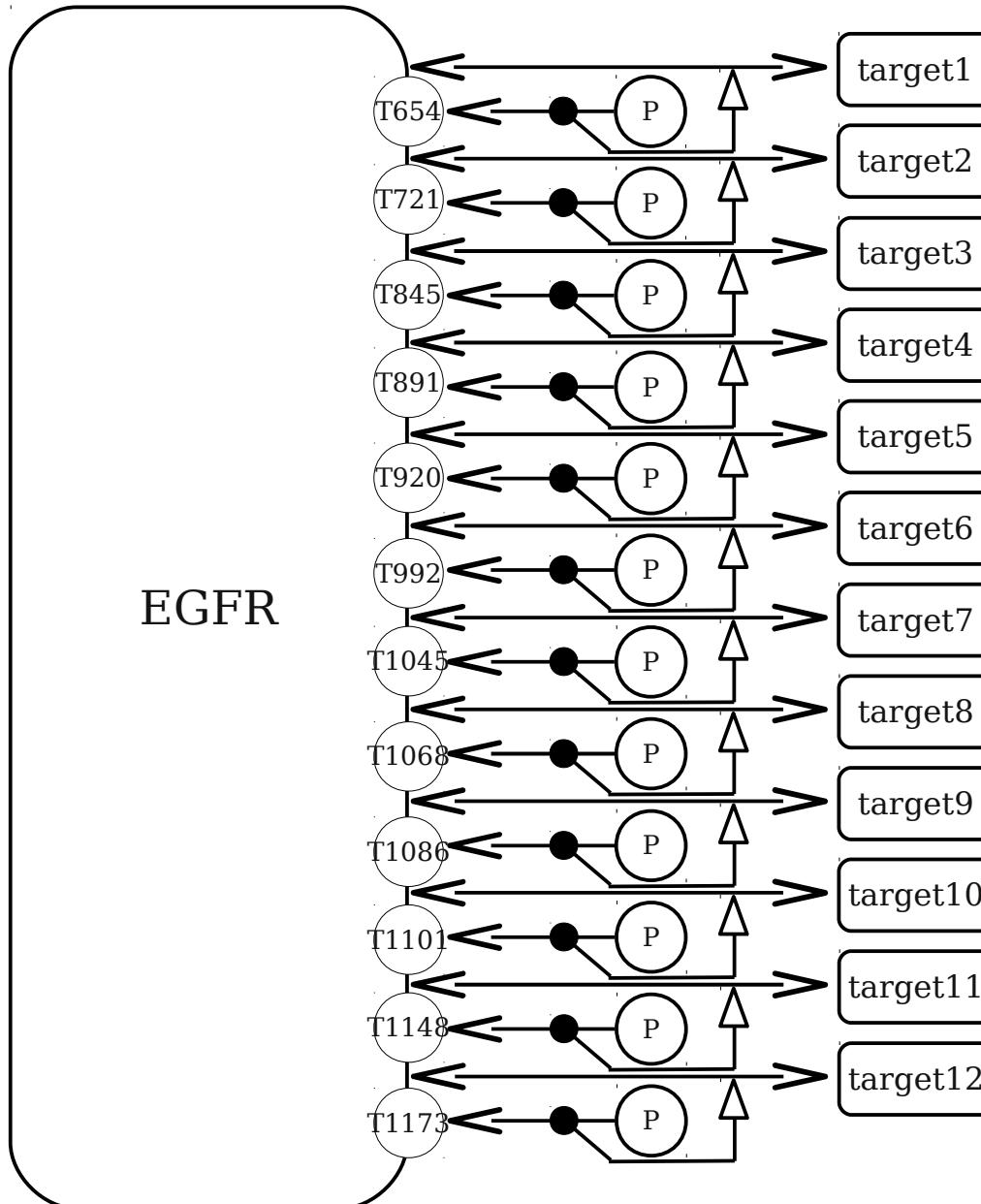


If A exists, the assignment of the value P to the state variable T of B is increased

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

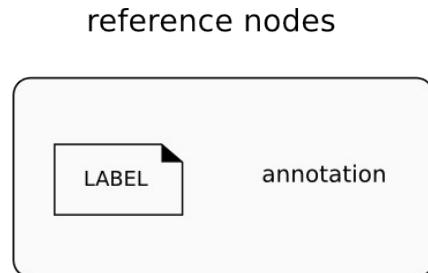
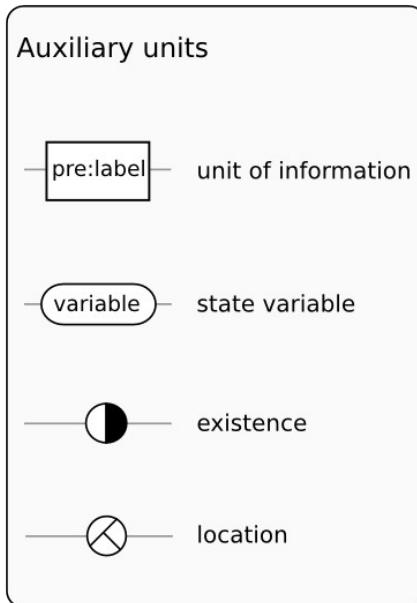
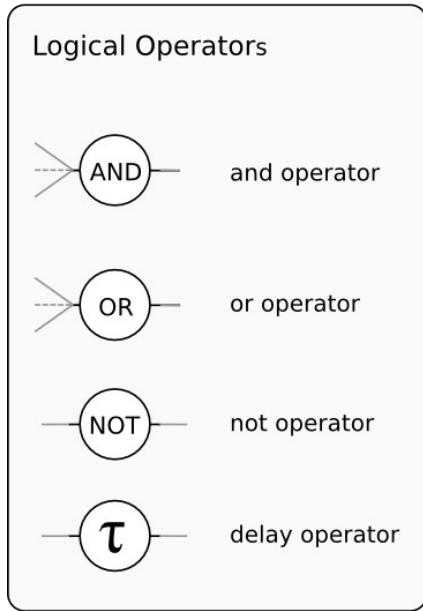
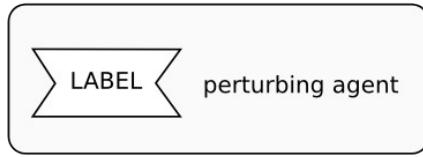
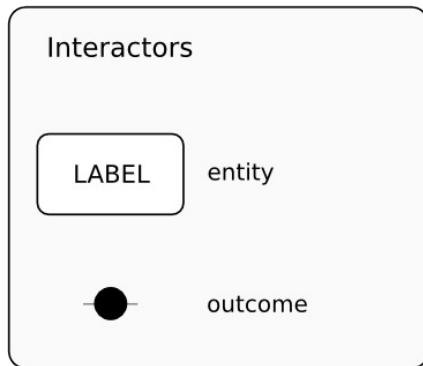


Process Diagram:
“once a state variable value,
always a state variable value”

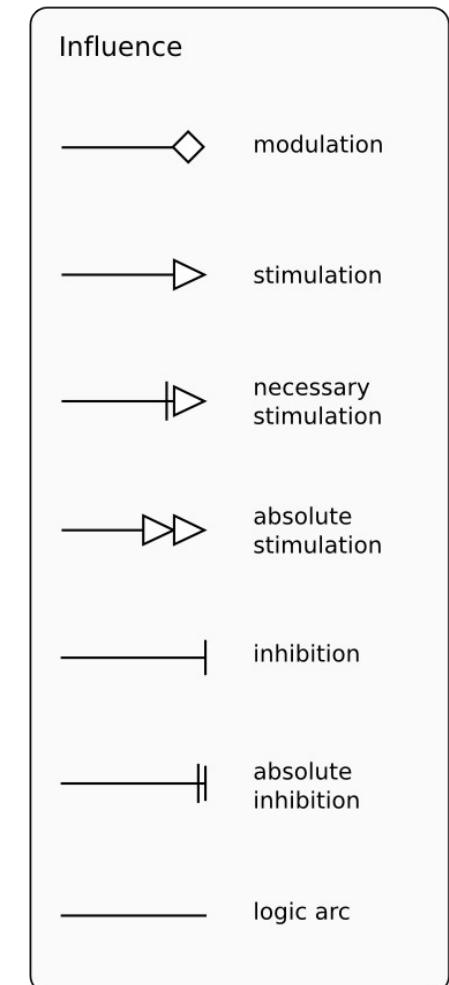
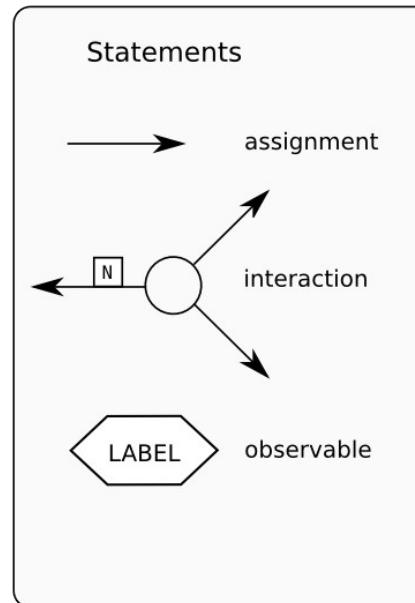
$2^{12} = 4096$ states
(i.e. EPN glyphs) for EGFR
and 4096 complexes between
EGFR and targets

SBGN Entity Relationships L1 reference card

Entity Nodes

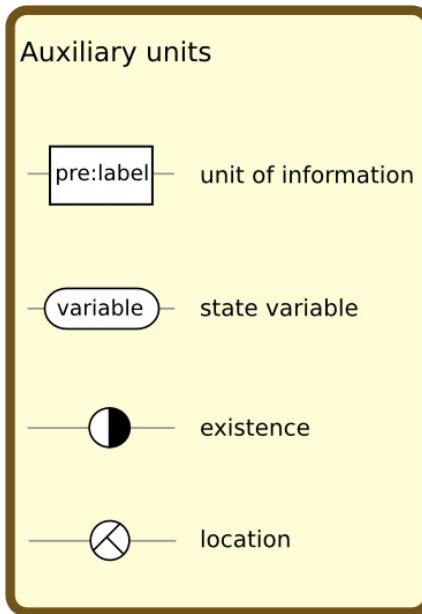
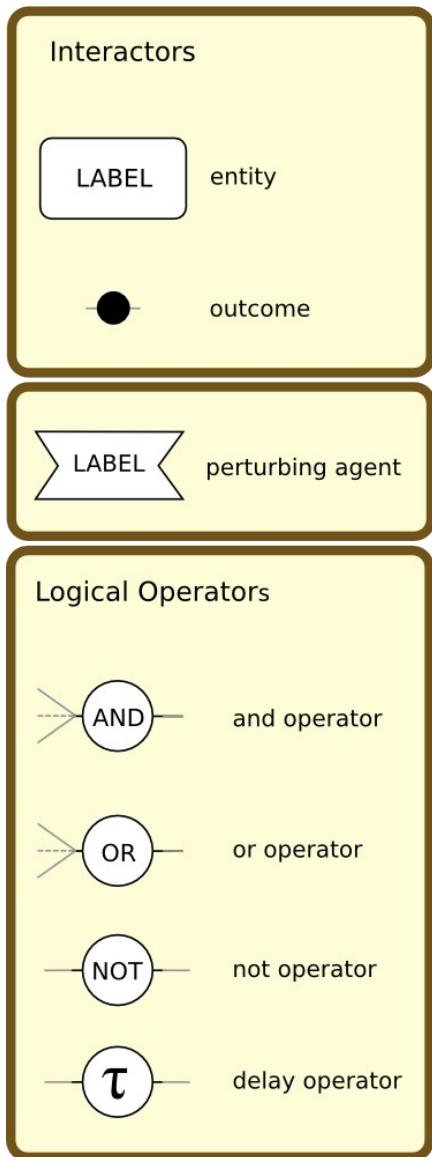


Relationship Nodes

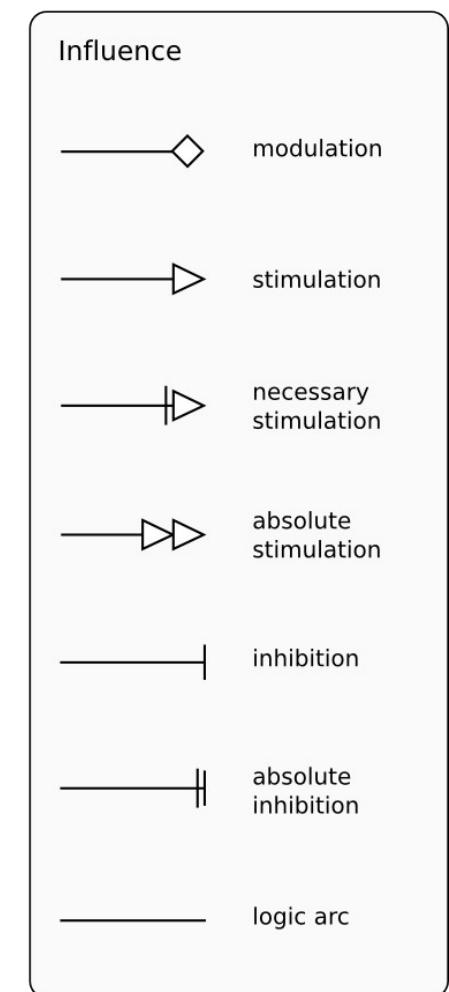
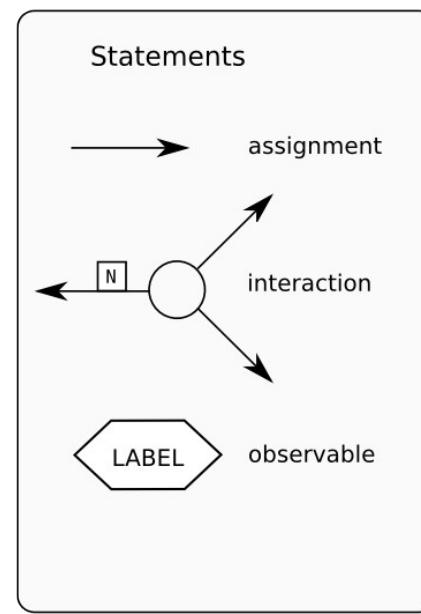


SBGN Entity Relationships L1 reference card

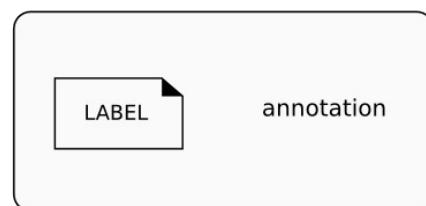
Entity Nodes



Relationship Nodes

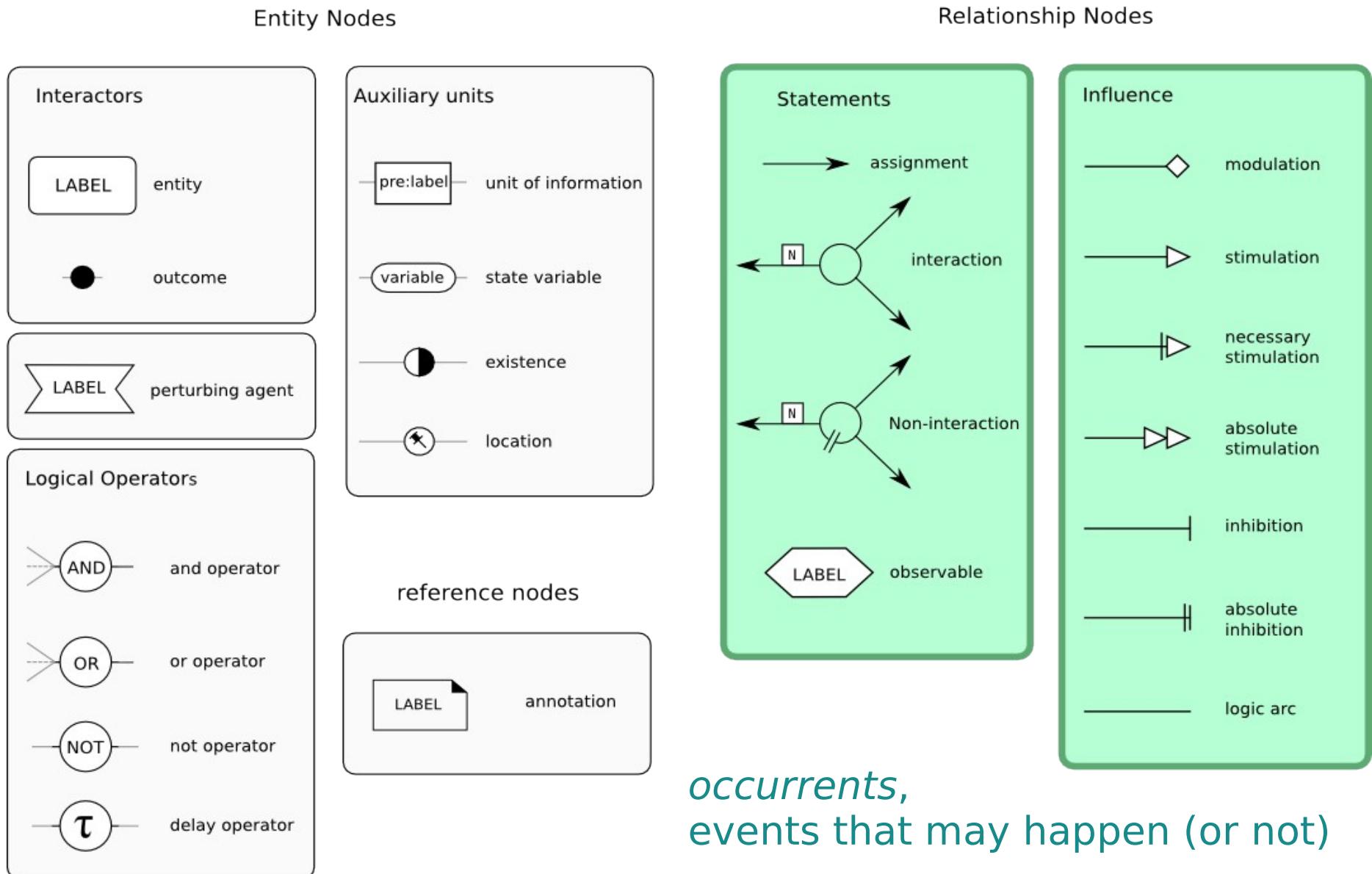


reference nodes



*continuants,
things that exists (or not)*

SBGN Entity Relationships L1 reference card



symbols \ Arc	<i>assignment</i>	<i>interaction</i>	<i>modulation</i>	<i>stimulation</i>	<i>inhibition</i>	<i>necessary stimulation</i>	<i>absolute stimulation</i>	<i>absolute inhibition</i>	<i>logic arc</i>
<i>entity</i>		IO	I	I	I	I	I	I	I
<i>outcome</i>		I(1)O(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)
<i>and</i>			I(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)O
<i>or</i>			I(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)O
<i>not</i>			I(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)O(1)
<i>delay</i>			I(1)	I(1)	I(1)	I(1)	I(1)	I(1)	I(1)O(1)
<i>perturbing agent</i>			I	I	I	I	I	I	I
<i>unit of information</i>		IO							
<i>state variable</i>	I(1)O(1)								
<i>modulation</i>			O	O	O	O	O	O	
<i>stimulation</i>			O	O	O	O	O	O	
<i>inhibition</i>			O	O	O	O	O	O	
<i>necessary stimulation</i>			O	O	O	O	O	O	
<i>absolute stimulation</i>			O	O	O	O	O	O	
<i>absolute inhibition</i>			O	O	O	O	O	O	
<i>assignment</i>			O	O	O	O	O	O	
<i>interaction</i>			O	O	O	O	O	O	
<i>phenotype</i>			O	O	O	O	O	O	

3.4.2 Influences

A *modulation* (Section 2.4.3.1) linking an *entity node* E and a relationship R means: “If E exists then R is either reinforced or weakened”.

A *stimulation* (Section 2.4.3.2) linking an *entity node* E and a relationship R means: “If E exists then R is reinforced” or “If E exists then the probability of R is increased”.

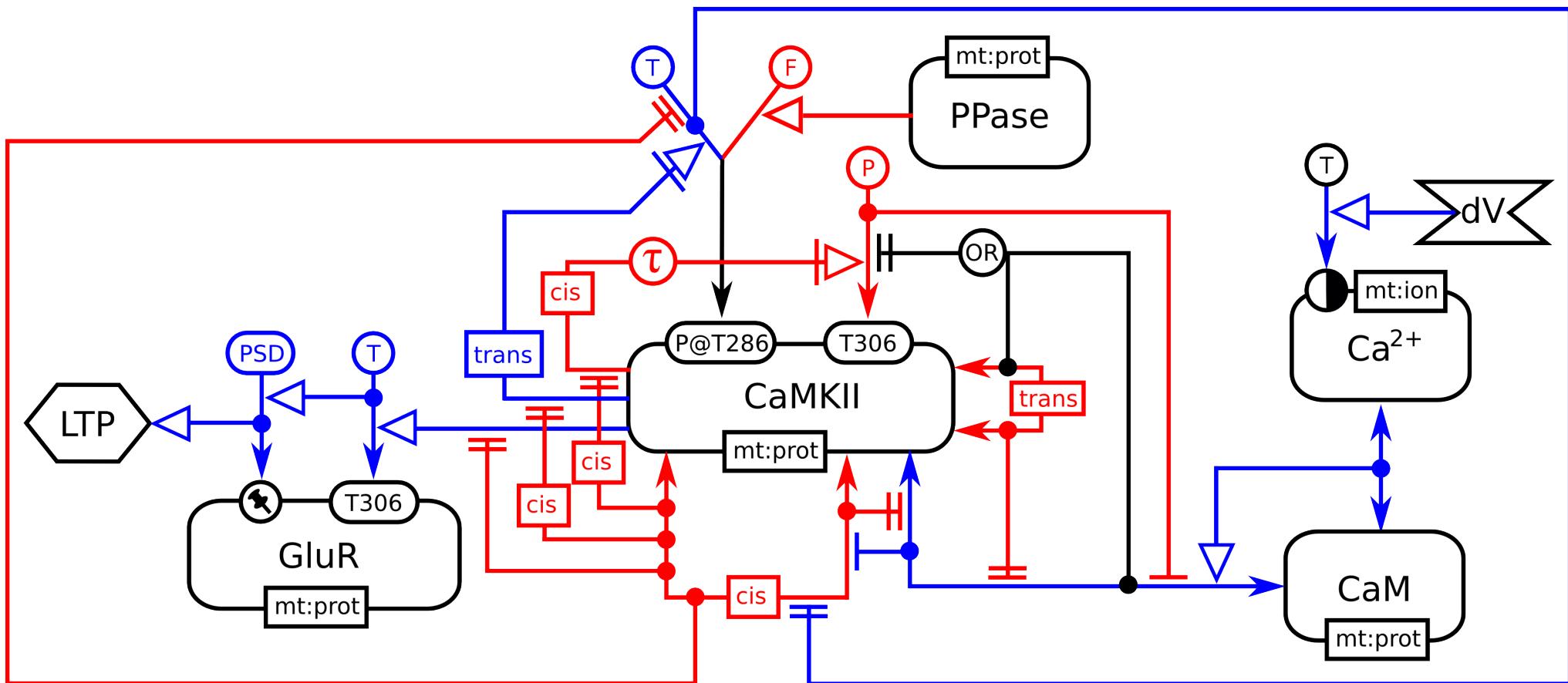
An *absolute stimulation* (Section 2.4.3.6) linking an *entity node* E and a relationship R means: “If E exists then R always takes place”.

A *necessary stimulation* (Section 2.4.3.4) linking an *entity node* E and a relationship R means: “ R only takes place if E exists”.

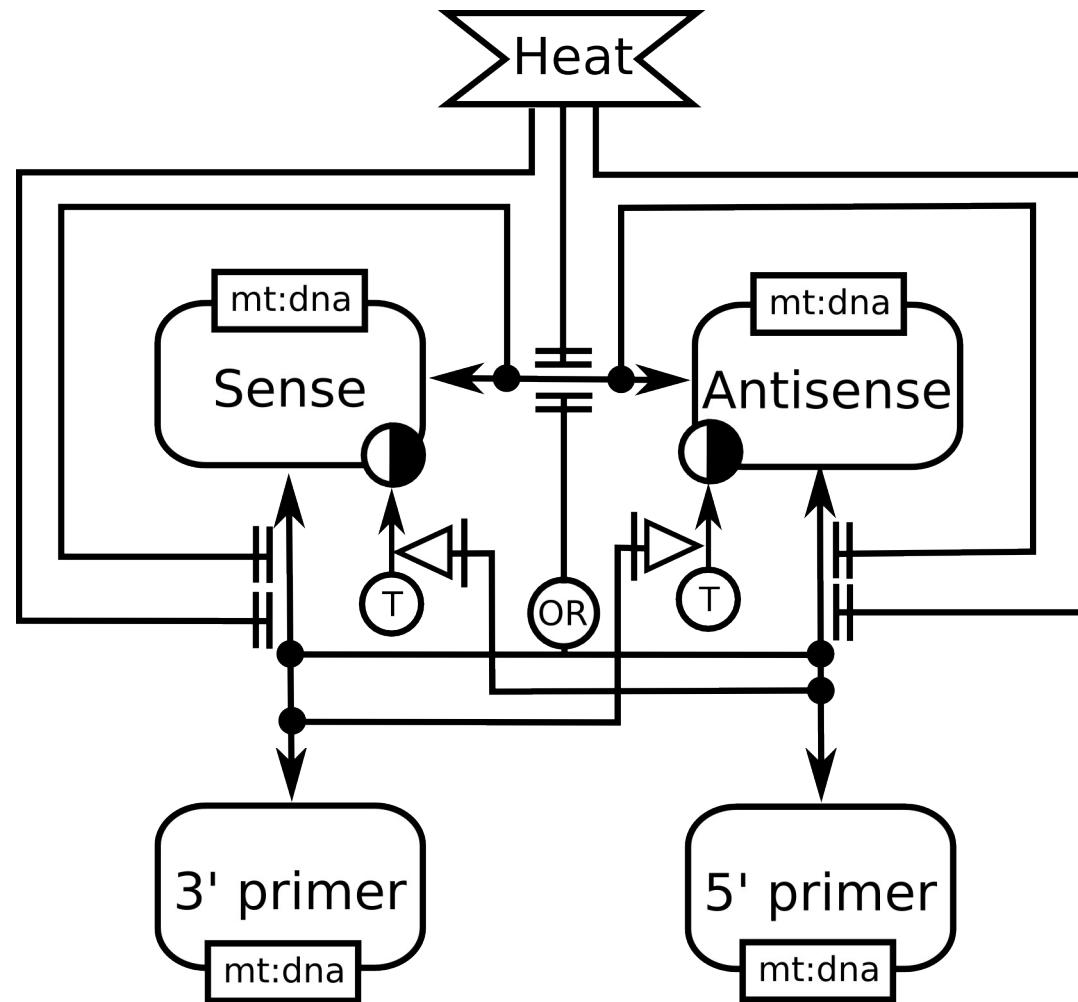
An *inhibition* (Section 2.4.3.3) linking an *entity node* E and a relationship R means: “If E exists then R is weakened” or “If E exists then the probability of R is lowered”.

An *absolute inhibition* (Section 2.4.3.5) linking an *entity node* E and a relationship R means: “If E exists then R never takes place”.

ER map of calcium-regulated synaptic plasticity

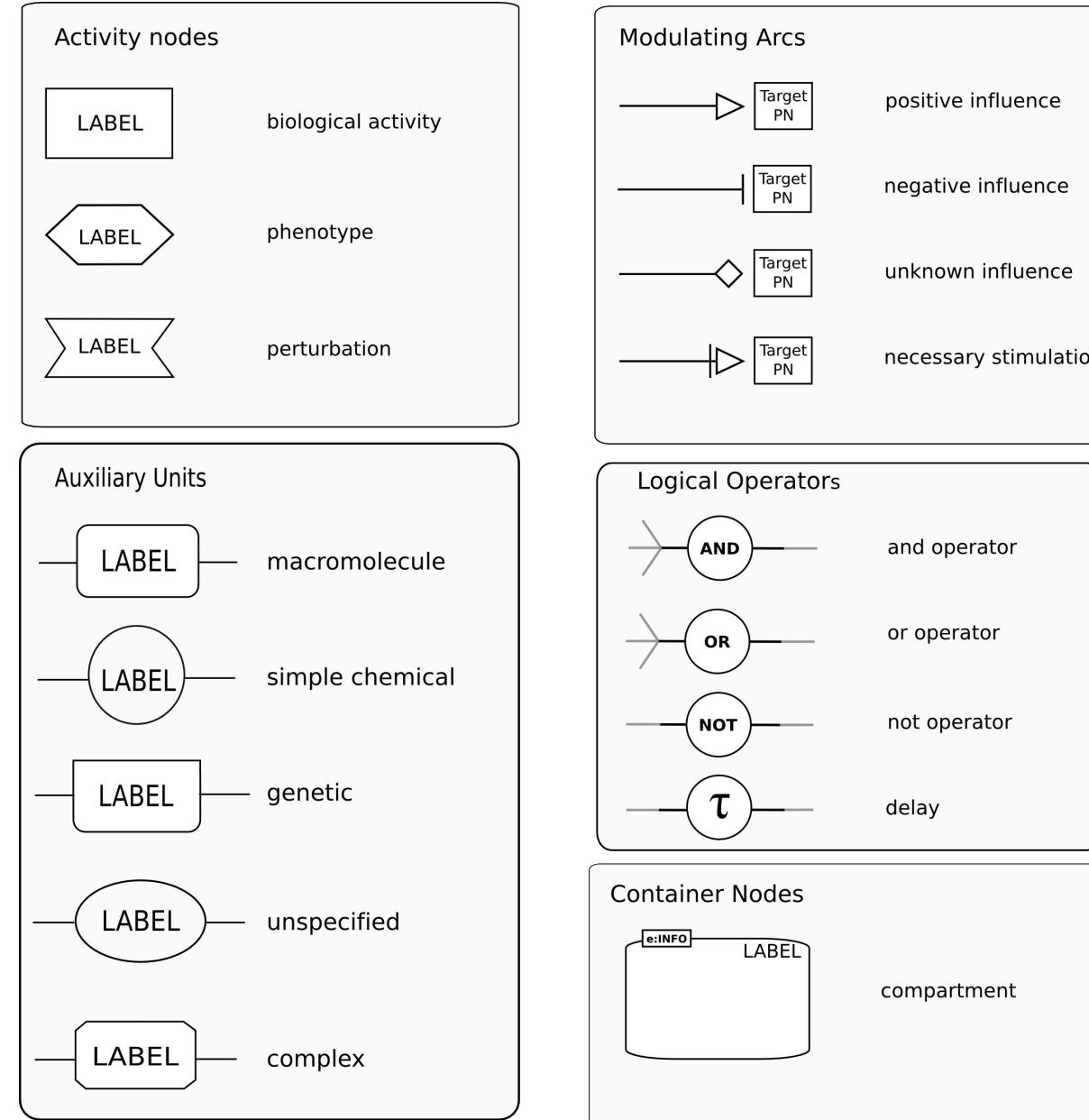


increases synaptic weight
decreases synaptic weight

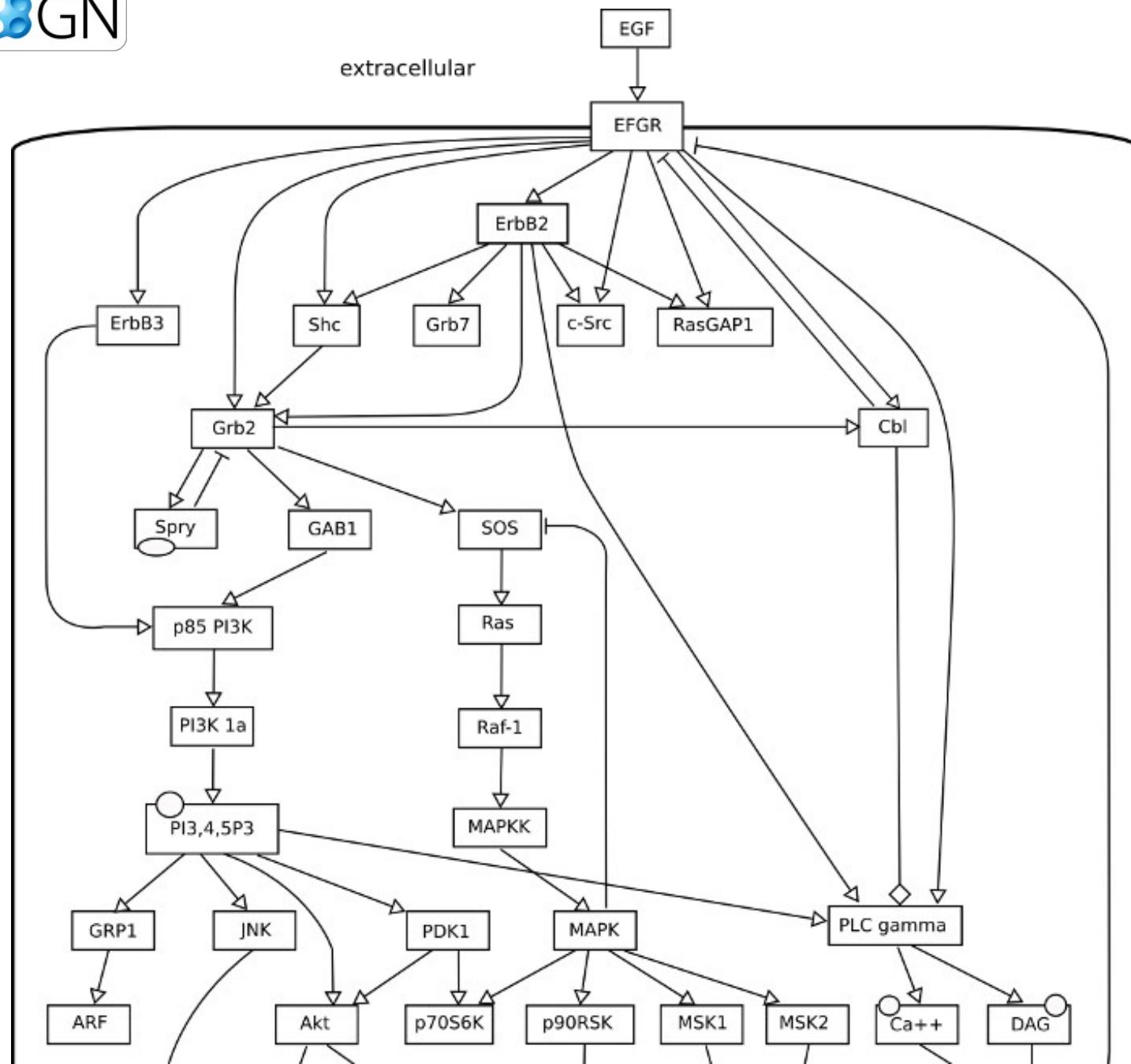


- Internal structure of entities, such as domains, sites, features. (NB: This has been agreed upon for the next version. We will be able to “nest” entities)
- Identification of instances: How to differentiate between several instances of the same entity, differentially involved in a relationships (e.g. trans-phosphorylation)?
- Identification of generics: How to lump together several entities for a given relationships? (e.g. MAPK instead of ERK1 and ERK2).

SBGN Activity Flows L1 reference card

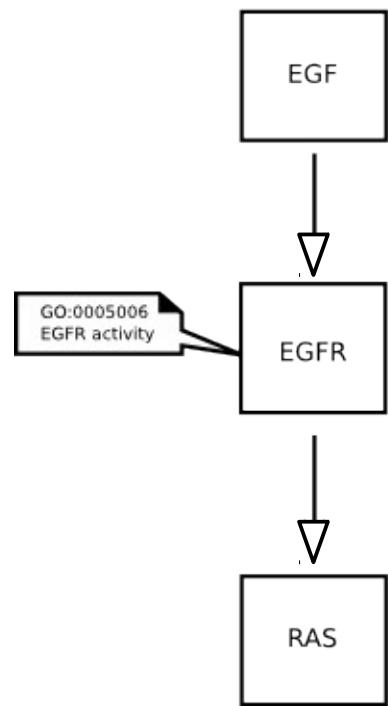


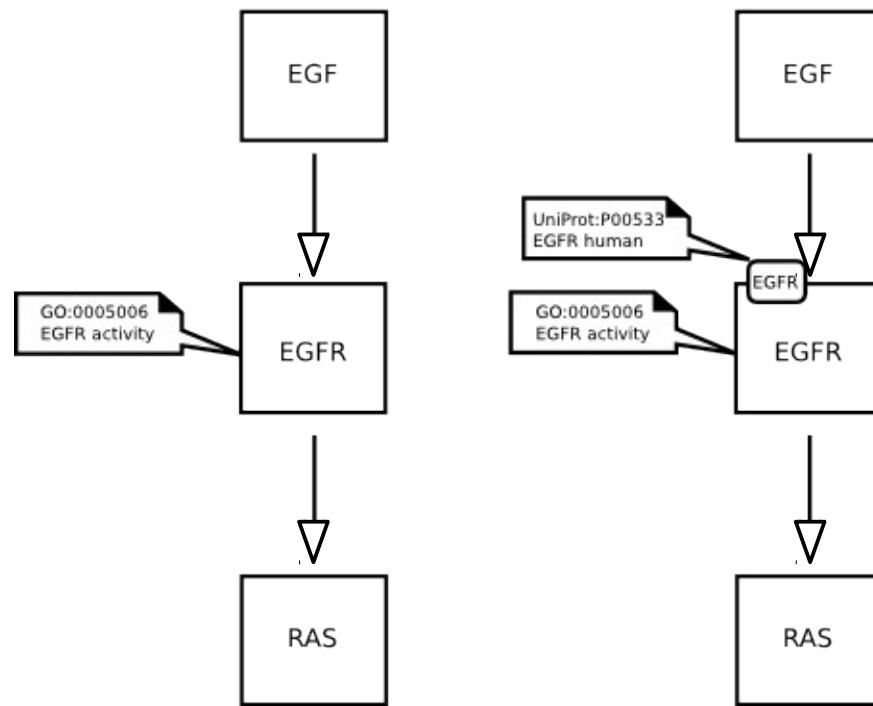
Example of Activity Flow map

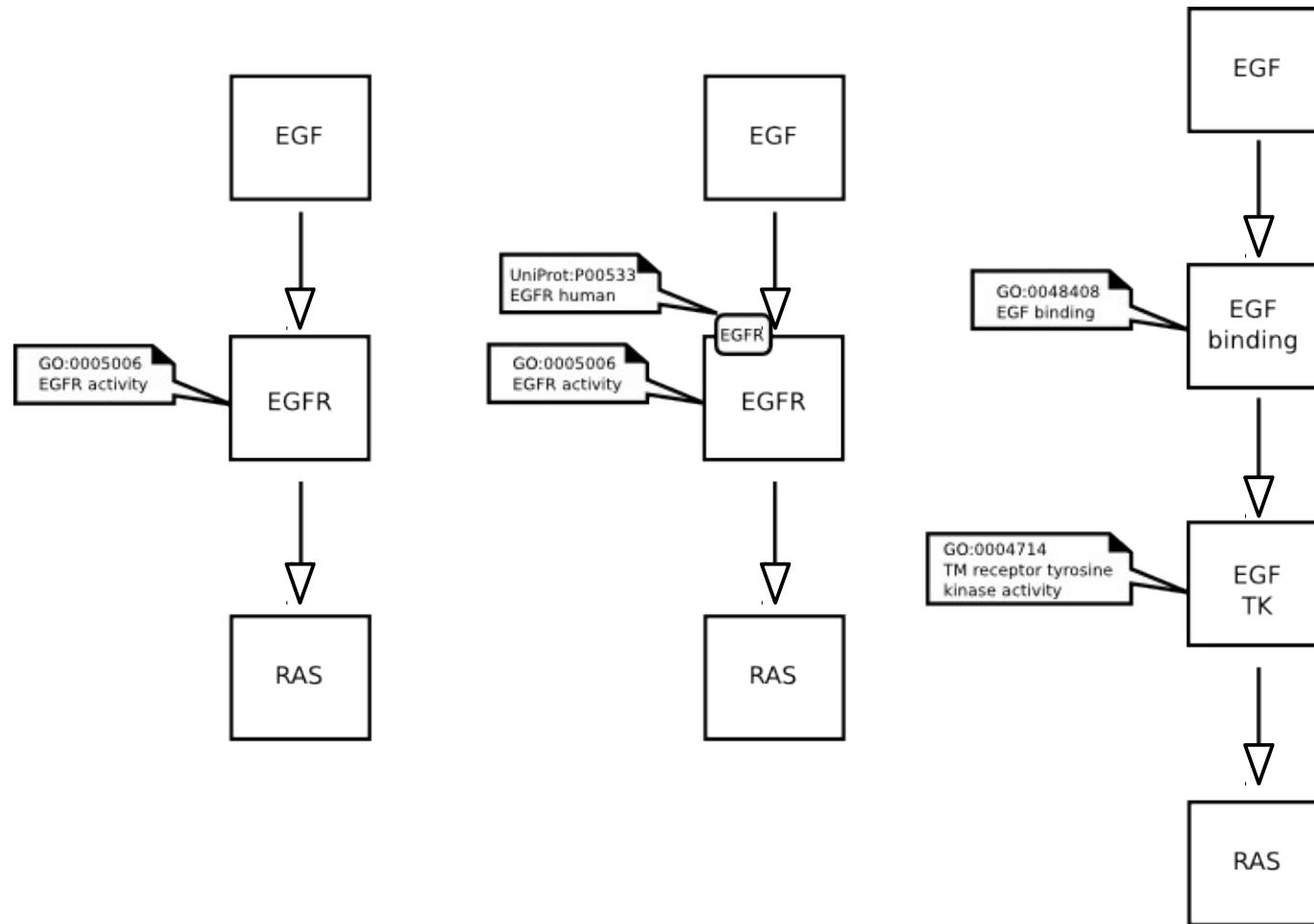


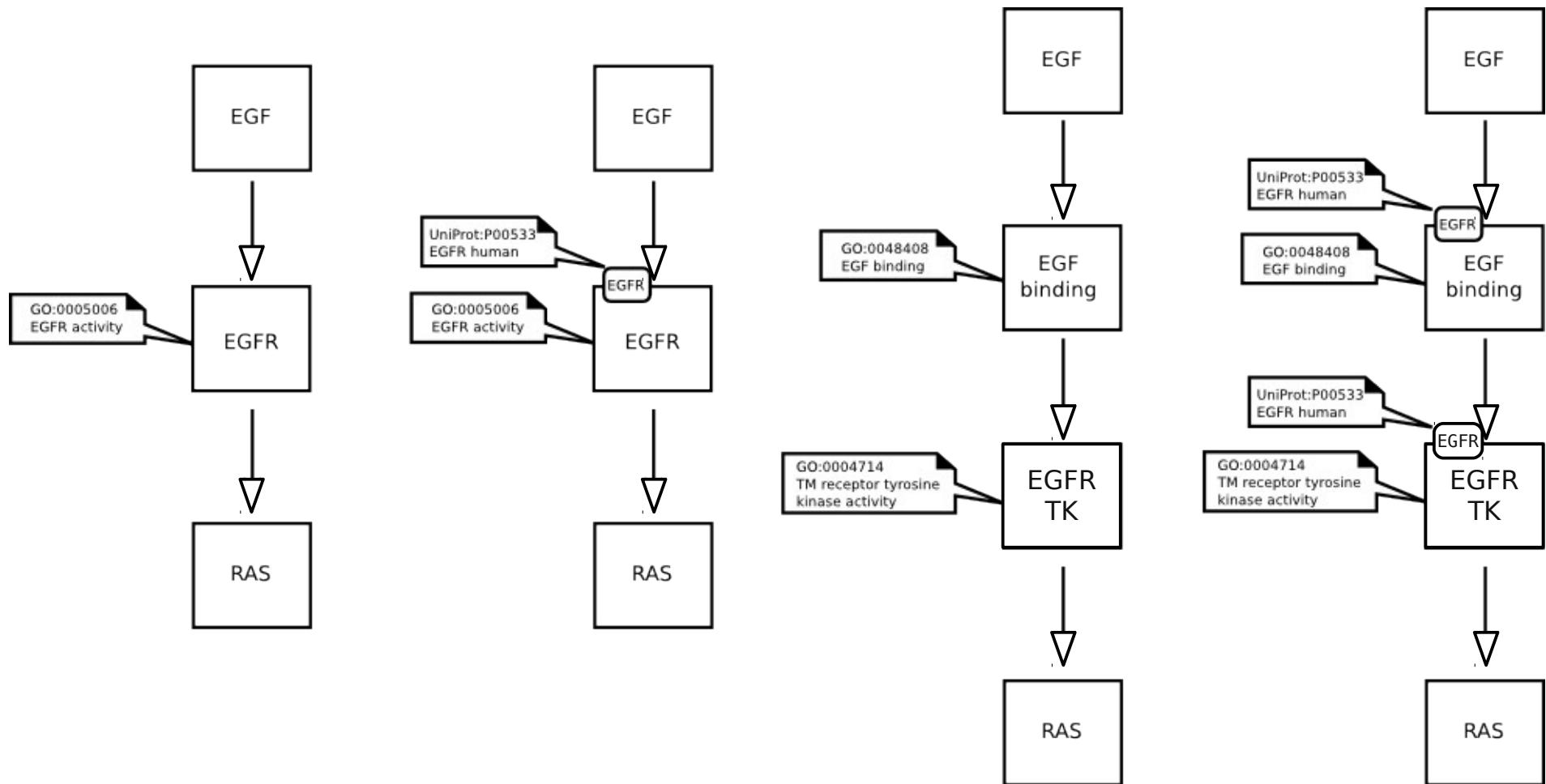


There Is More Than One Way To Do It (TIMTOWDI)

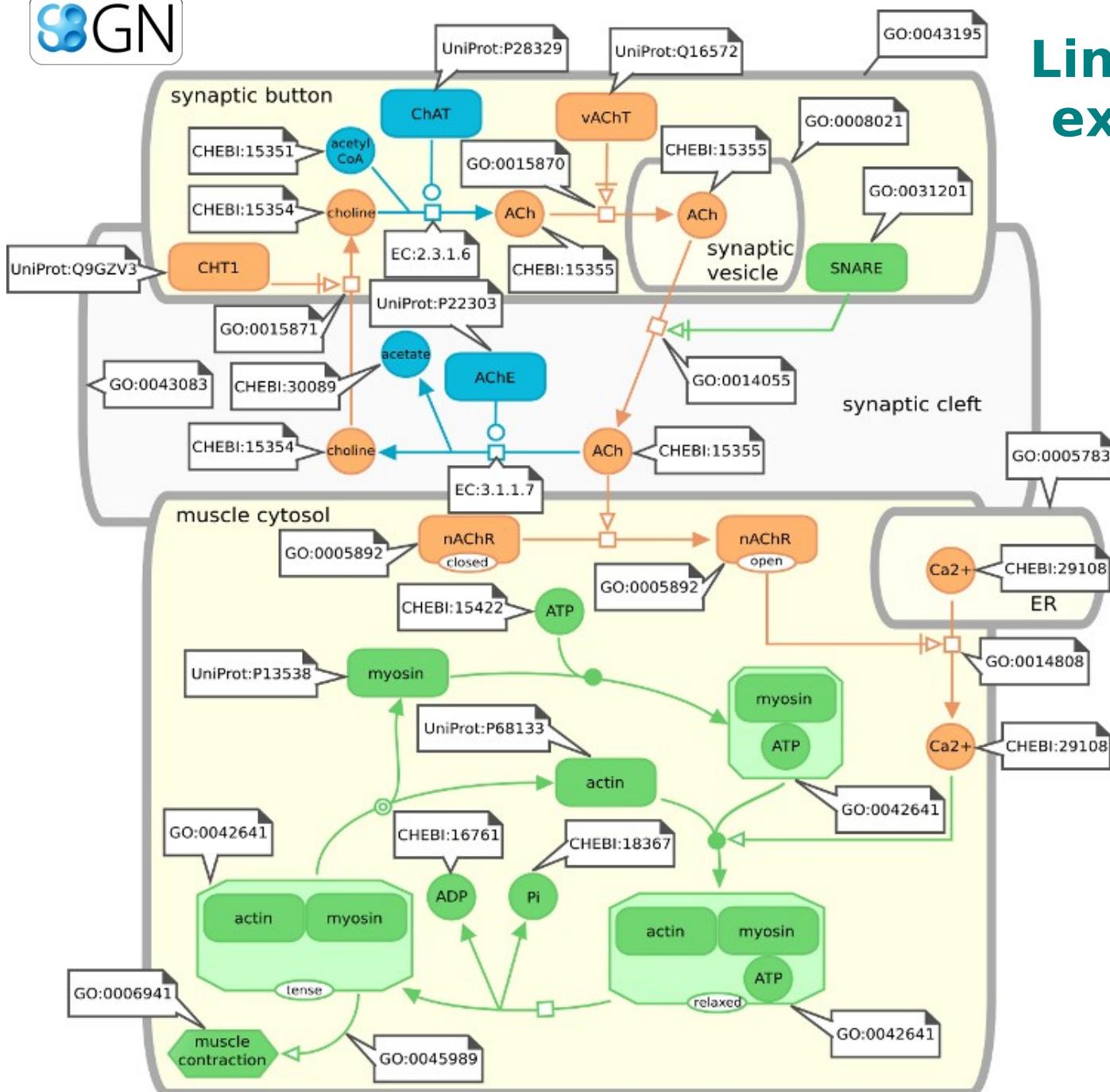


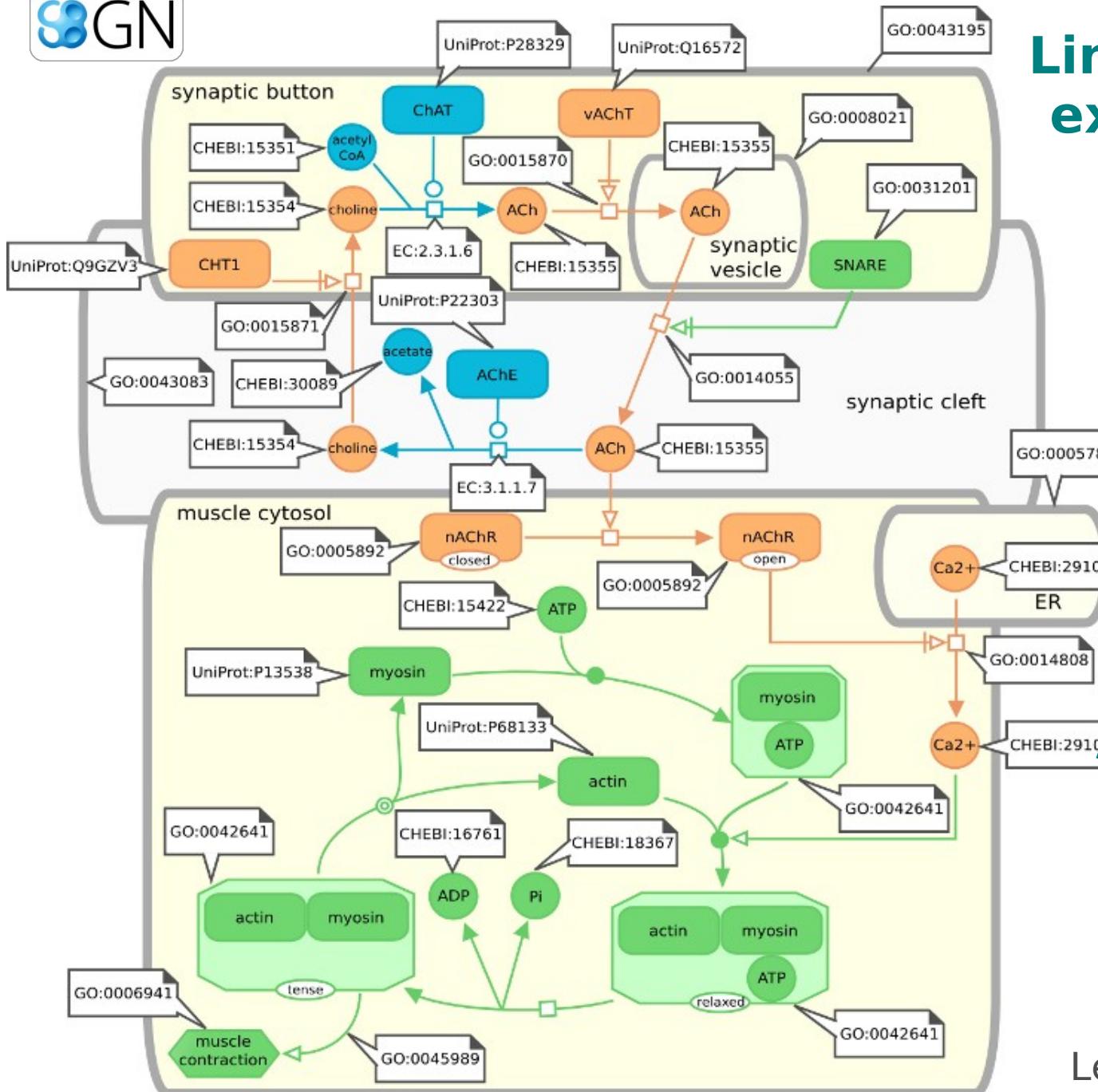






Linking SBGN maps to external information





Linking SBGN maps to external information

EBI-EBI Search All Databases Enter Text Here Go Research Advanced

DBpedia Search ChEBI

Search ChEBI Help

calcium(2+) (CHEBI:29108)

Main Automatic Xrefs

ChEBI Name calcium(2+) ChEBI ID CHEBI:29108 Last Modified 10 March 2008

Image Applet

Molfile InChI InChI=1/Calq+2

InChIKey InChIKey=BHQYQMZTOCNFJ-UHFFFAOYAP

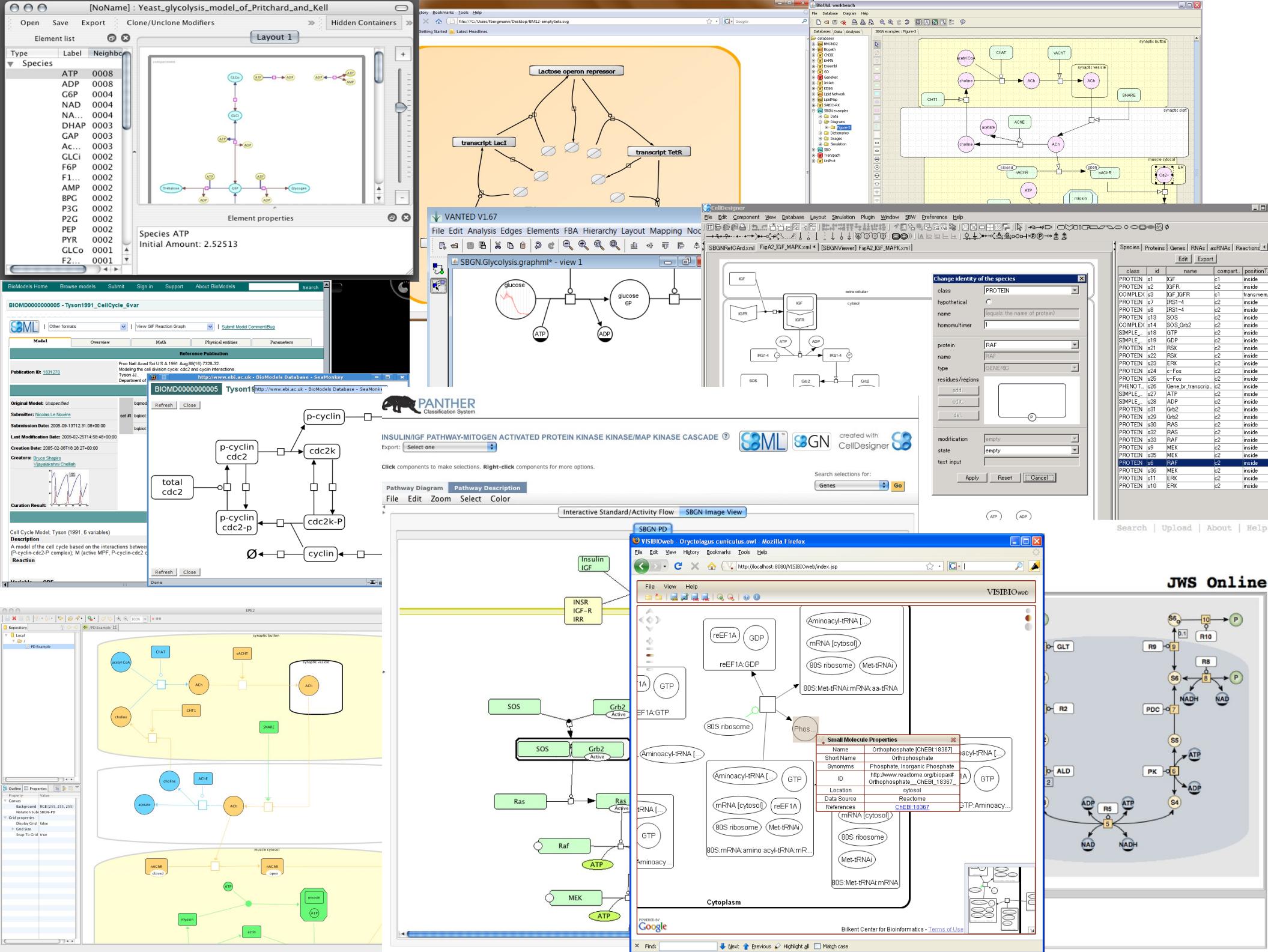
SMILES [Ca++]

Ca²⁺



<http://www.ebi.ac.uk/miriam/>

Le Novère et al (2005)
Nat Biotechnol, 23: 1509-1515.



- SBGN Process Descriptions
 - Level 1 Version 1.0 released on August 23rd 2008
 - Level 1 Version 1.1 released on September 1st 2009
 - Level 1 Version 2 under discussion
- SBGN Entity Relationships
 - Level 1 Version 1.0 released on September 1st 2009
- SBGN Activity Flows
 - Level 1 Version 1.0 released on September 1st 2009

- 4rd SBGN hackathon
(SBGN 5.5)

- 21-23 April 2010
 - Wittenberg

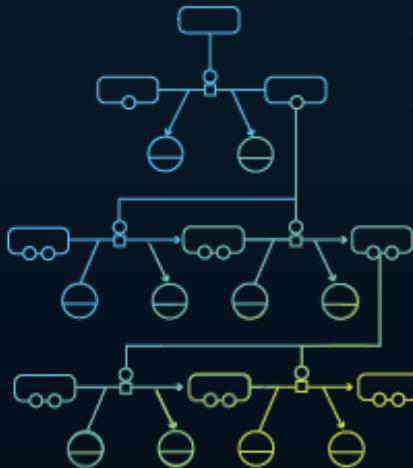
- 6th SBGN forum
(provisional)

- October 2010
 - Edinburgh
 - Satellite of ICSB 2010



- Future hackathons and forums

- part of HARMONY and COMBINE meetings
 - SBML (model), SED-ML (simulation), SBO (semantics), MIRIAM (metadata), SBGN (graphical), BioModels DB (distribution), ?



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 maps of biochemical and cellular processes studied in systems biology.

Standardizing the visual representation is crucial for more efficient and accurate transmission of biological knowledge between different communities in research, education, publishing, and more. When biologists are as familiar with the notation as electronics engineers are familiar with the notation of circuit schematics, they can save the time and effort required to familiarize themselves with different notations, and instead spend more time thinking about the biology being depicted.

SBGN is made up of [[three orthogonal languages](#)], representing different visions of biological systems. Each language defines a comprehensive set of symbols with precise semantics, together with detailed syntactic rules how maps are to be interpreted.

On this site, you can browse some [example maps](#) to get a feeling for SBGN, read the SBGN [specification documents](#), join [online discussions](#), see current working documents and software support 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.

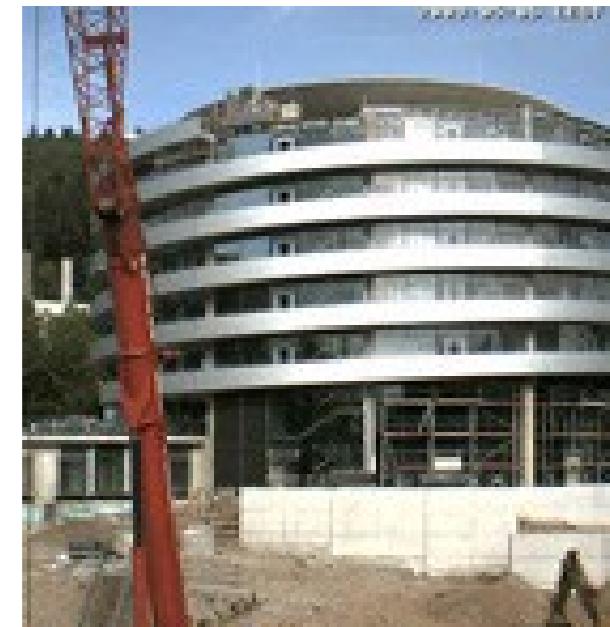
To quote SBGN as a whole, please use:

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H. The Systems Biology Graphical Notation. *Nat Biotechnol.* 2009 27(8):735-41.

SBGN News

(02 Sep. 09) The first specifications for [SBGN Entity Relationships](#) and [SBGN Activity Flows](#) are out.

What happens if one can read the blueprint



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

