

# **Report on the status of SBGN ER and proposed extensions**

# Graphs are everywhere in biology

Nature Precedings - doi:10.1038/npre.2010.4963.1 - Posted 8 Oct 2010

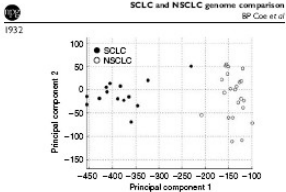
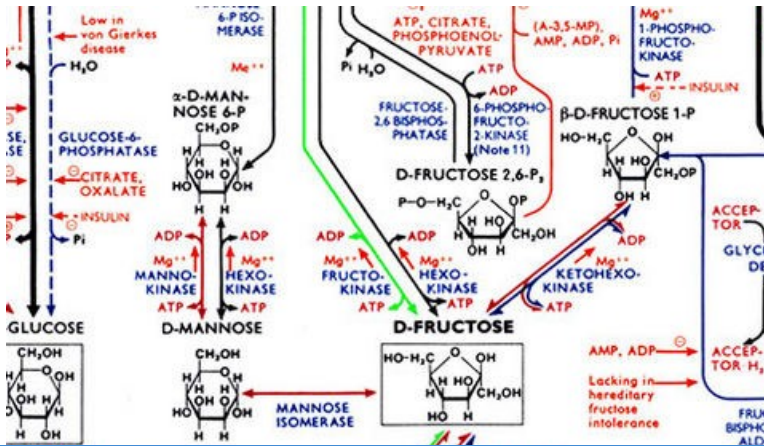
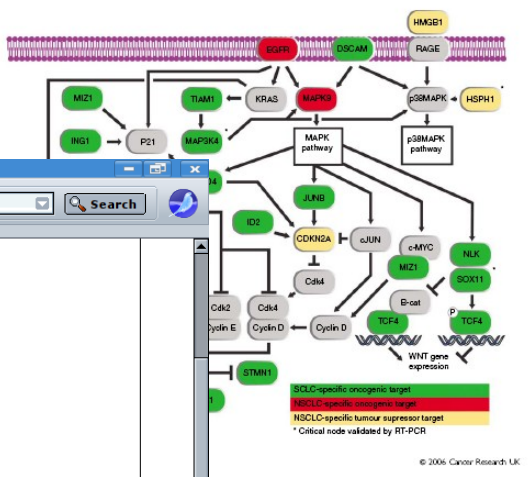
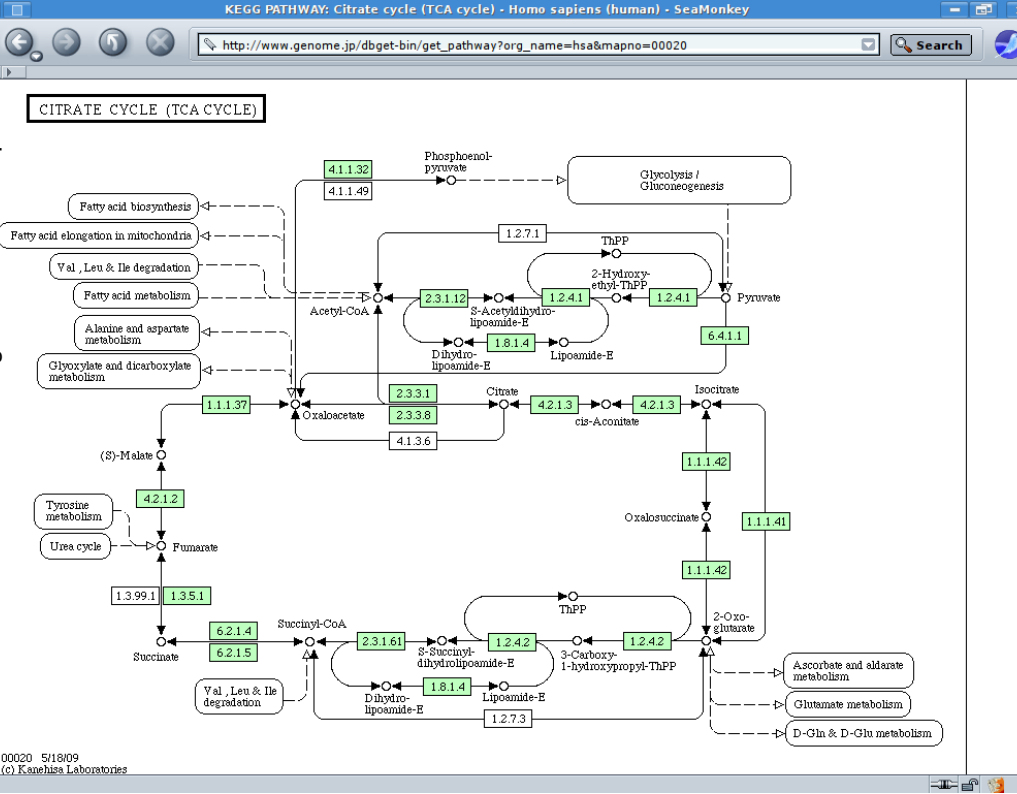
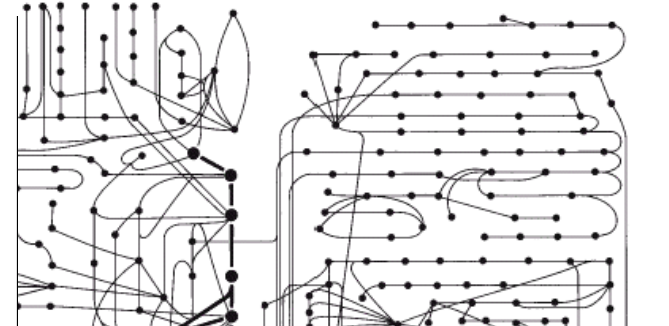


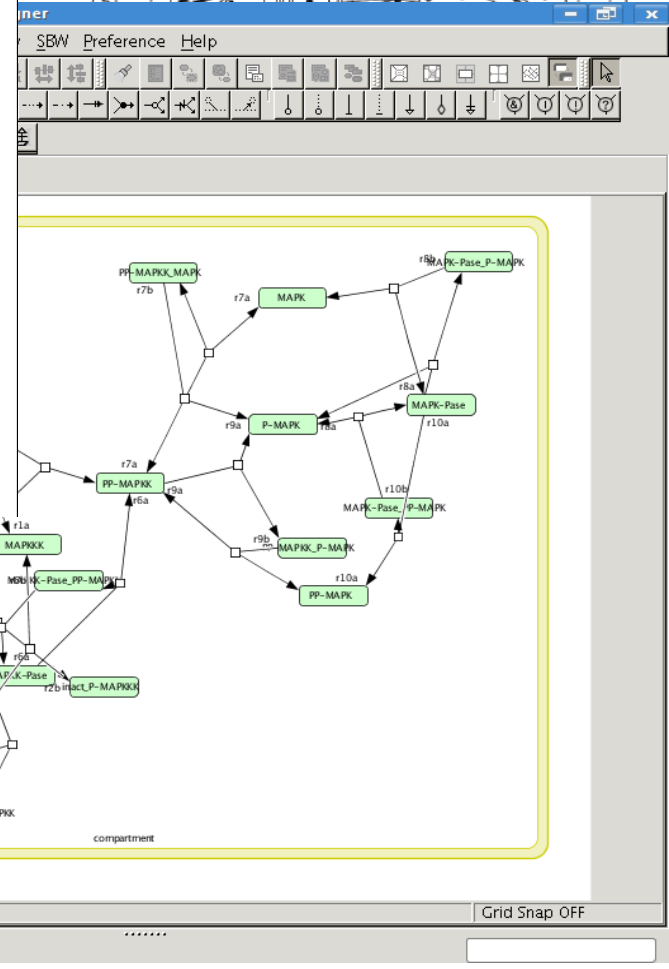
Figure 4 Contribution of copy number-induced gene expression differences to the SCLC and NSCLC phenotypes. Principal component analysis was performed using all 243 Affymetrix probe sets demonstrating expression differences as a result of copy number alterations. The SCLC samples are indicated by solid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

amplification in the NSCLC samples as well suggests that this gene may play an essential role in the development of lung cancers (Garnis et al. 2006).  
It is noteworthy that a subset of the genomic similarities between the SCLC and NSCLC cell lines could be the result of adaptation to culturing conditions. Owing to this, the greatest insight into the biology of the clinical disease will be attainable through analysis of differences (rather than their similarities) in genomic alterations and gene regulation.

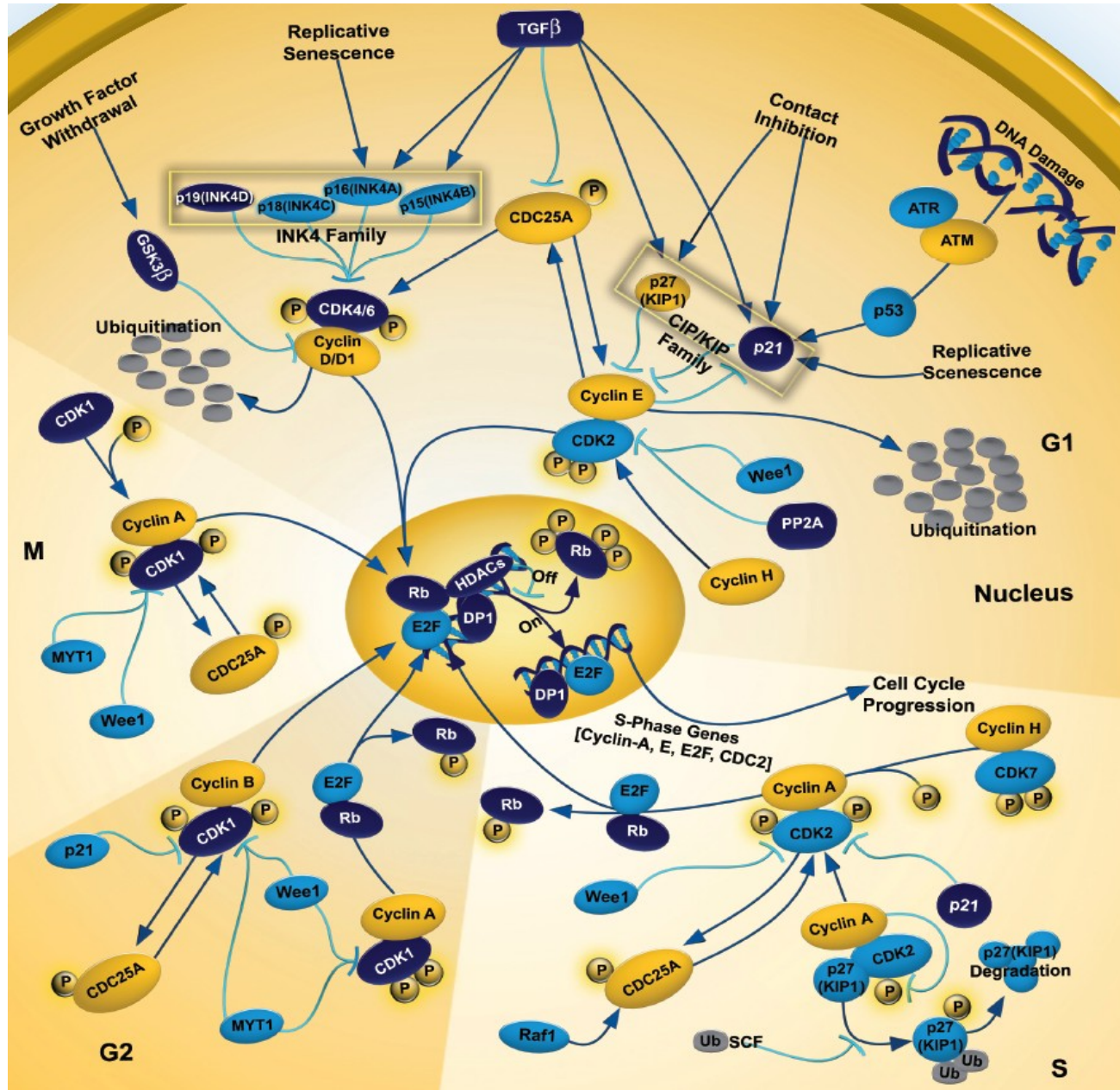
Regions of difference  
Through our analysis, numerous regions throughout the genome were determined to be differentially altered between the SCLC and NSCLC samples. This difference-based approach compensates for random cell culturing artifacts and should identify the regions most strongly linked to clinical disease. These regions ranged in size from whole chromosomes (chromosome 21) to discrete peaks, kilobases in size (Ag1.1). Using our stringent, multiple criteria (Fisher's exact test followed by additional thresholding), we detected several regions that differed strongly in their alteration status between the cell types we refer to these as phenotype-specific copy number alterations (PSCNAs). These included 1p36.33-1p34.2.



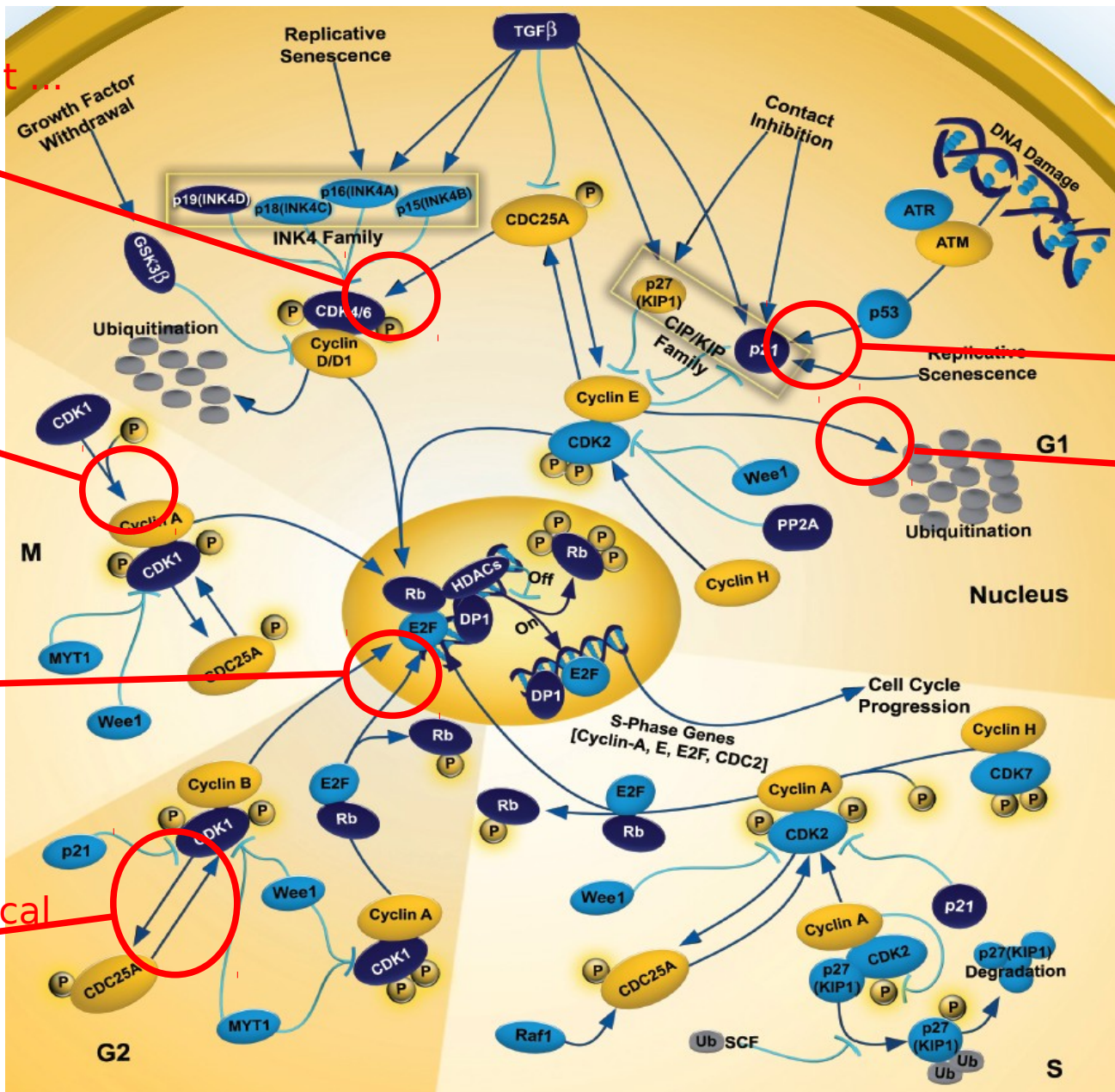
SCLC-specific oncogenic target  
NSCLC-specific oncogenic target  
NSCLC-specific tumour suppressor target  
Critical node validated by RT-PCR



# 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?

# Ambiguity of usual representations



is transformed into

translocates (X "=" Y)

is degraded into

associates into

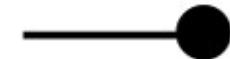
dissociates into

stimulates the activity of

stimulates the expression of

catalyses the formation of

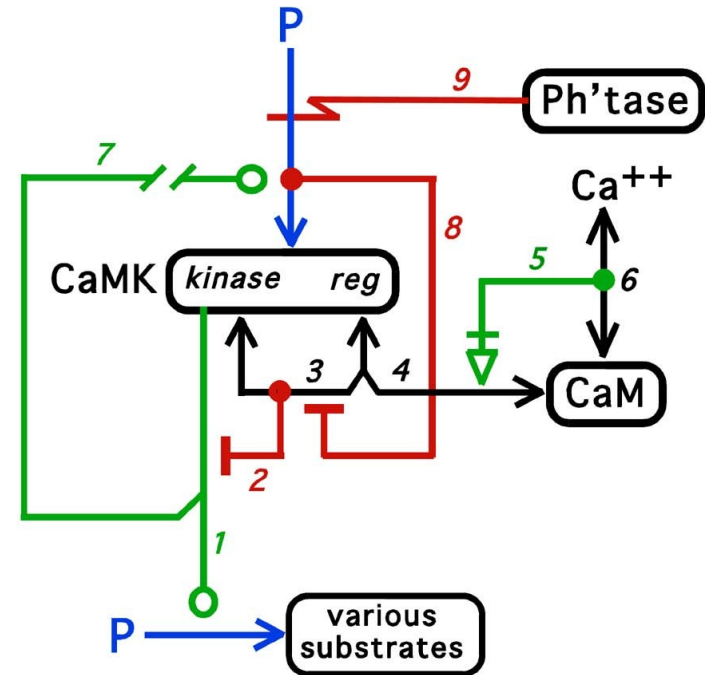
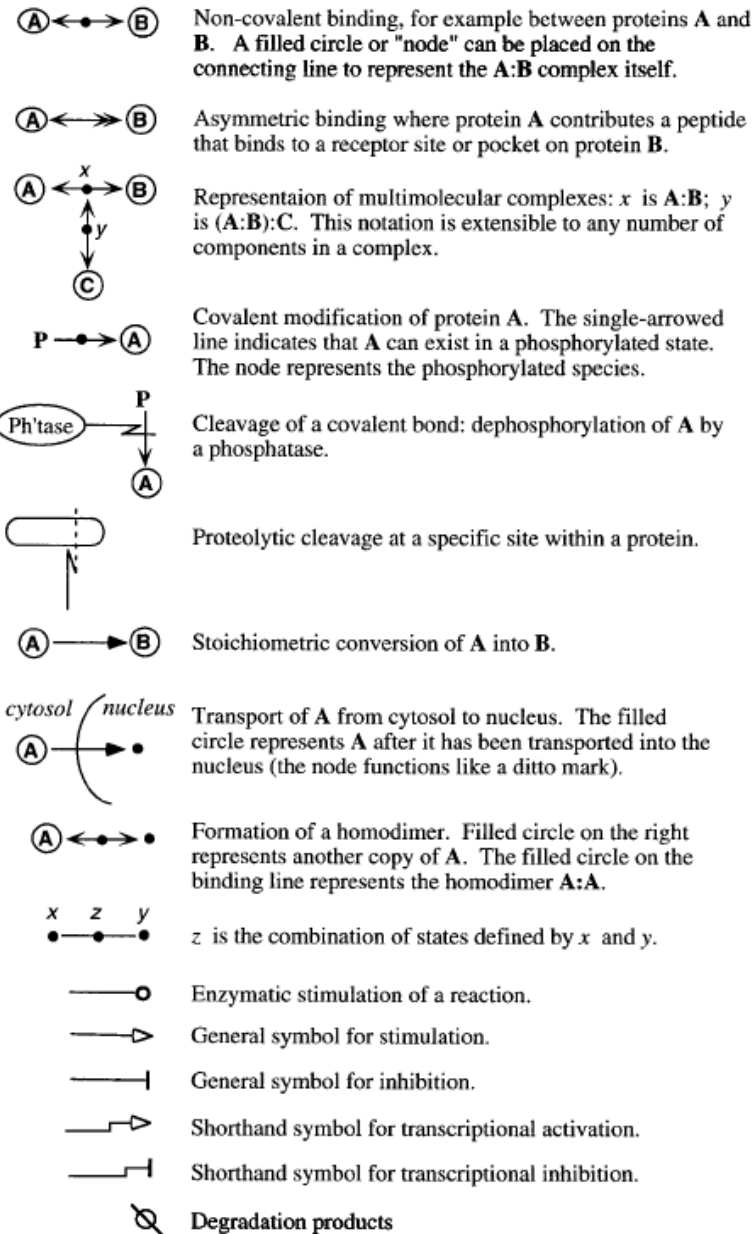
X inhibits Y



inhibition



# Kohn's Molecular Interaction Maps



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

# Kitano's Notation

## 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

### Category-I reduced notation

Degradation	
Transcription	
Translation	
Module	

### Category-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	<ul style="list-style-type: none"> <li> Connectivity (binding, etc.)</li> </ul>
Promotor and coding structure for gene	
Exon structure for RNA	

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

# Enters The Systems Biology Graphical Notation



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



# What is SBGN?

- 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

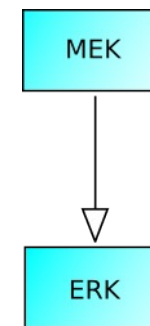
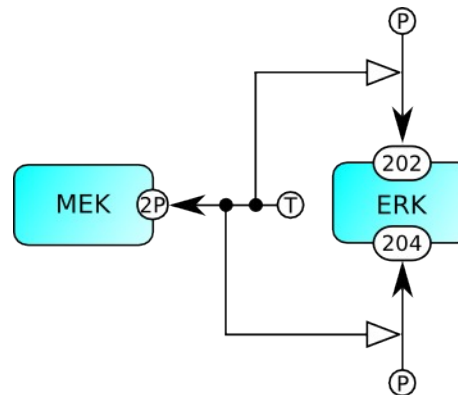
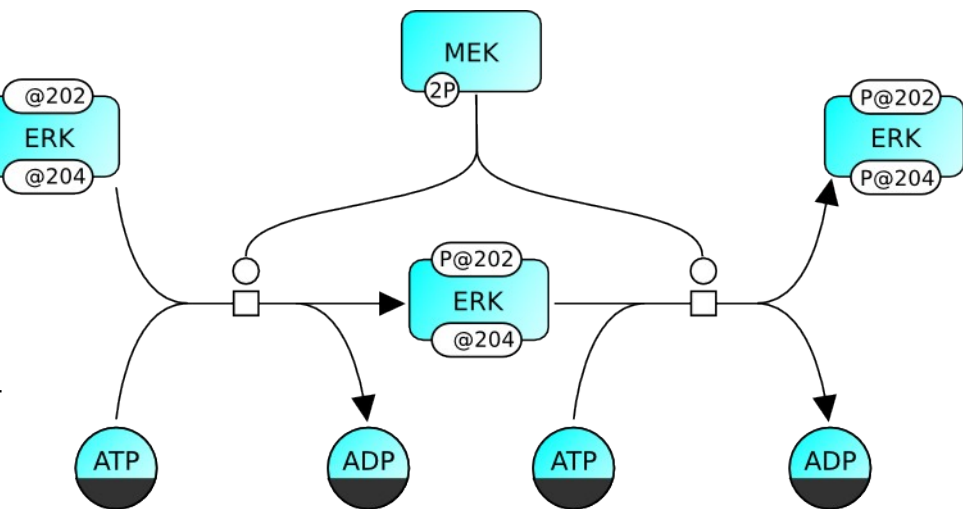
# Graph trinity: three languages in one notation

## Process Descriptions

## Entity Relationships

## Activity Flows

Nature Precedings : doi:10.1038/npre.2010.4963.1 : Posted 8 Oct 2010

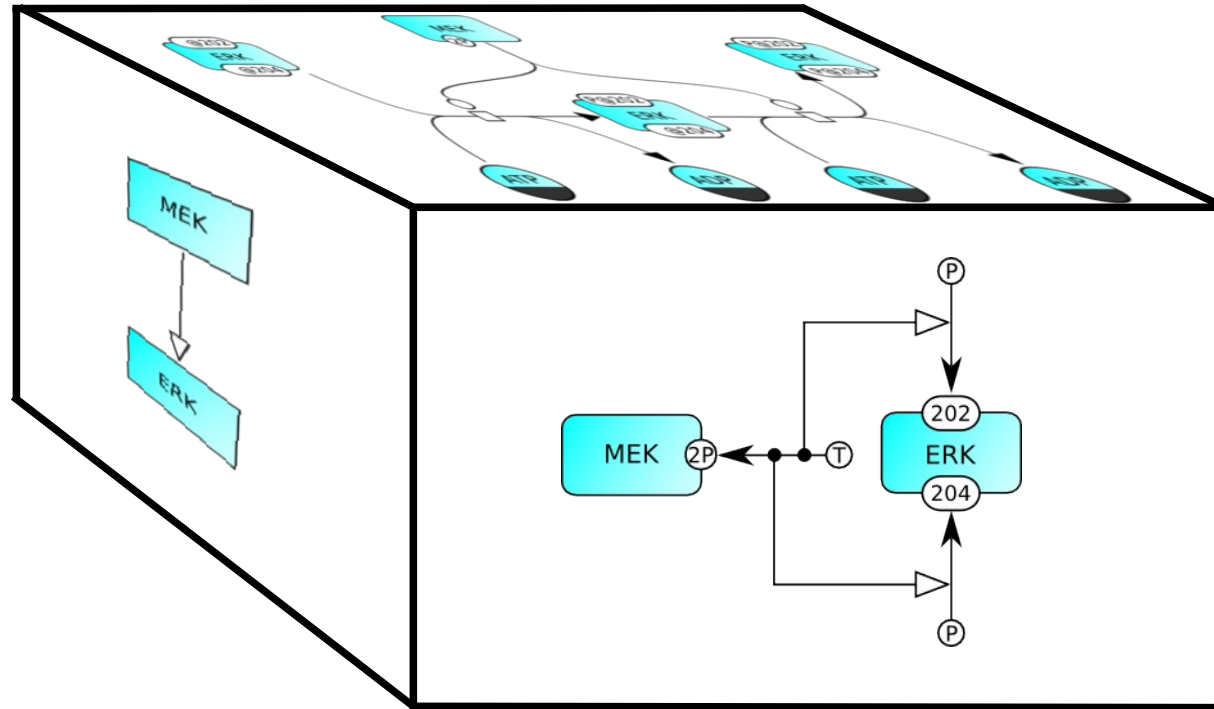


- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

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

- Ambiguous
- Conceptual
- Sequential

# Three orthogonal projections of biology



# The Systems Biology Graphical Notation

Nicolas Le Novère<sup>1</sup>, Michael Hucka<sup>2</sup>, Huaiyu Mi<sup>3</sup>, Stuart Moodie<sup>4</sup>, Falk Schreiber<sup>5,6</sup>, Anatoly Sorokin<sup>7</sup>, Emek Demir<sup>8</sup>, Katja Wegner<sup>9</sup>, Mirit I Aladjem<sup>10</sup>, Sarala M Wimalaratne<sup>11</sup>, Frank T Bergman<sup>12</sup>, Ralph Gauges<sup>13</sup>, Peter Ghazal<sup>4,14</sup>, Hideya Kawaji<sup>15</sup>, Lu Li<sup>1</sup>, Yukiko Matsuoka<sup>16</sup>, Alice Villéger<sup>17,18</sup>, Sarah E Boyd<sup>19</sup>, Laurence Calzone<sup>20</sup>, Melanie Courtot<sup>21</sup>, Ugur Dogrusoz<sup>22</sup>, Tom C Freeman<sup>14,23</sup>, Akira Funahashi<sup>24</sup>, Samik Ghosh<sup>16</sup>, Akiya Jouraku<sup>24</sup>, Sohyoung Kim<sup>10</sup>, Fedor Kolpakov<sup>25,26</sup>, Augustin Luna<sup>10</sup>, Sven Sahle<sup>13</sup>, Esther Schmidt<sup>1</sup>, Steven Watterson<sup>4,22</sup>, Guanming Wu<sup>27</sup>, Igor Goryanin<sup>4</sup>, Douglas B Kell<sup>18,28</sup>, Chris Sander<sup>8</sup>, Herbert Sauro<sup>12</sup>, Jacky L Snoep<sup>29</sup>, Kurt Kohn<sup>10</sup> & Hiroaki Kitano<sup>16,30,31</sup>

---

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

## Systems Biology Graphical Notation: Process Description language Level 1

Version 1.2

Date: 3 October, 2010

### Editors:

Stuart Moodie *School of Informatics, University of Edinburgh, UK*  
Nicolas Le Novère *EMBL European Bioinformatics Institute, UK*  
Anatoly Sorokin *School of Informatics, University of Edinburgh, UK*  
Huaiyu Mi *SRI International, USA*  
Falk Schreiber *IPK Gatersleben & MLU Halle, Germany*

To discuss any aspect of SBGN, please send your messages to the mailing list [sbgn-discuss@sbgn.org](mailto:sbgn-discuss@sbgn.org). To get subscribed to the mailing list or to contact us directly, please write to [sbgn-editors@lists.sourceforge.net](mailto:sbgn-editors@lists.sourceforge.net). Bug reports and specific comments about the specification should be entered in the issue tracker [http://p.sf.net/sbgn/pd\\_tracker](http://p.sf.net/sbgn/pd_tracker).



## Systems Biology Graphical Notation: Entity Relationship language Level 1

Version 1.1

Date: October 6, 2010

### Editors:

Nicolas Le Novère *EMBL European Bioinformatics Institute, UK*  
Stuart Moodie *CSBE, University of Edinburgh, UK*  
Emek Demir *Sloan-Kettering Institute, USA*  
Falk Schreiber *IPK Gatersleben & University of Halle, DE*  
Huaiyu Mi *SRI International, USA*

To discuss any aspect of SBGN, please send your messages to the mailing list [sbgn-discuss@caltech.edu](mailto:sbgn-discuss@caltech.edu). To get subscribed to the mailing list or to contact us directly, please write to [sbgn-editors@lists.sourceforge.net](mailto:sbgn-editors@lists.sourceforge.net). Bug reports and specific comments about the specification should be entered in the issue tracker [https://sourceforge.net/tracker/admin/?atid=117862&group\\_id=178553](https://sourceforge.net/tracker/admin/?atid=117862&group_id=178553).



## Systems Biology Graphical Notation: Activity Flow language Level 1

Release 1.0

Date: September 4, 2009

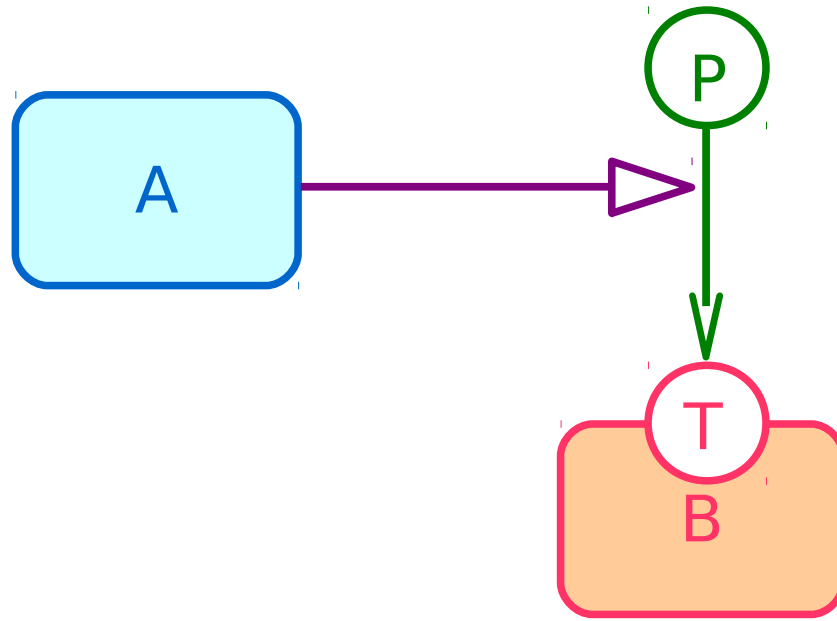
### Editors:

Huaiyu Mi *SRI International, USA*  
Falk Schreiber *IPK Gatersleben & MLU Halle, Germany*  
Nicolas Le Novère *EMBL European Bioinformatics Institute, UK*  
Stuart Moodie *CSBE, University of Edinburgh, UK*  
Anatoly Sorokin *University of Edinburgh, UK*

To discuss any aspect of SBGN, please send your messages to the mailing list [sbgn-discuss@sbgn.org](mailto:sbgn-discuss@sbgn.org). To get subscribed to the mailing list or to contact us directly, please write to [sbgn-editors@lists.sourceforge.net](mailto:sbgn-editors@lists.sourceforge.net). Bug reports and specific comments about the specification should be entered in the issue tracker [http://sourceforge.net/tracker/?func=adddgroup\\_id=178553&atid=1175366](http://sourceforge.net/tracker/?func=adddgroup_id=178553&atid=1175366).

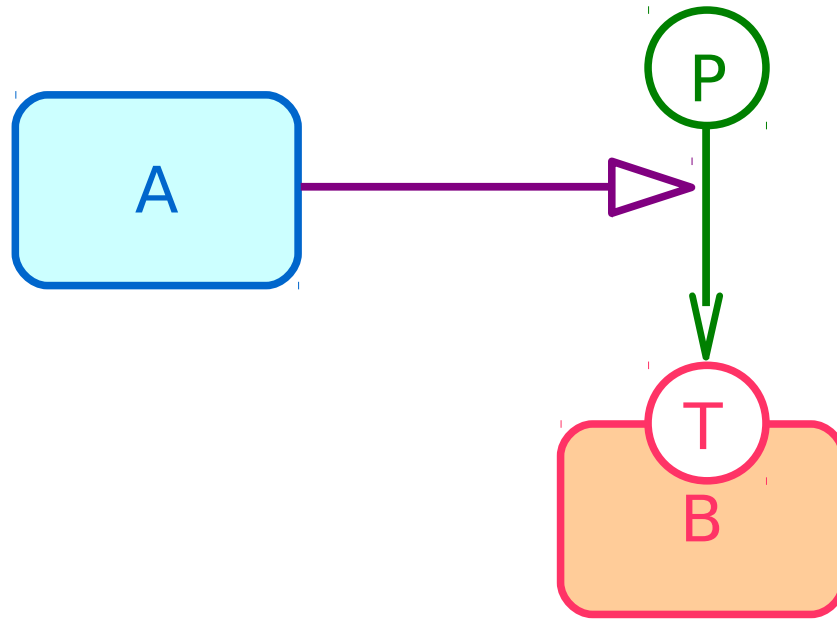


# Entity Relationships can be viewed as rules



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

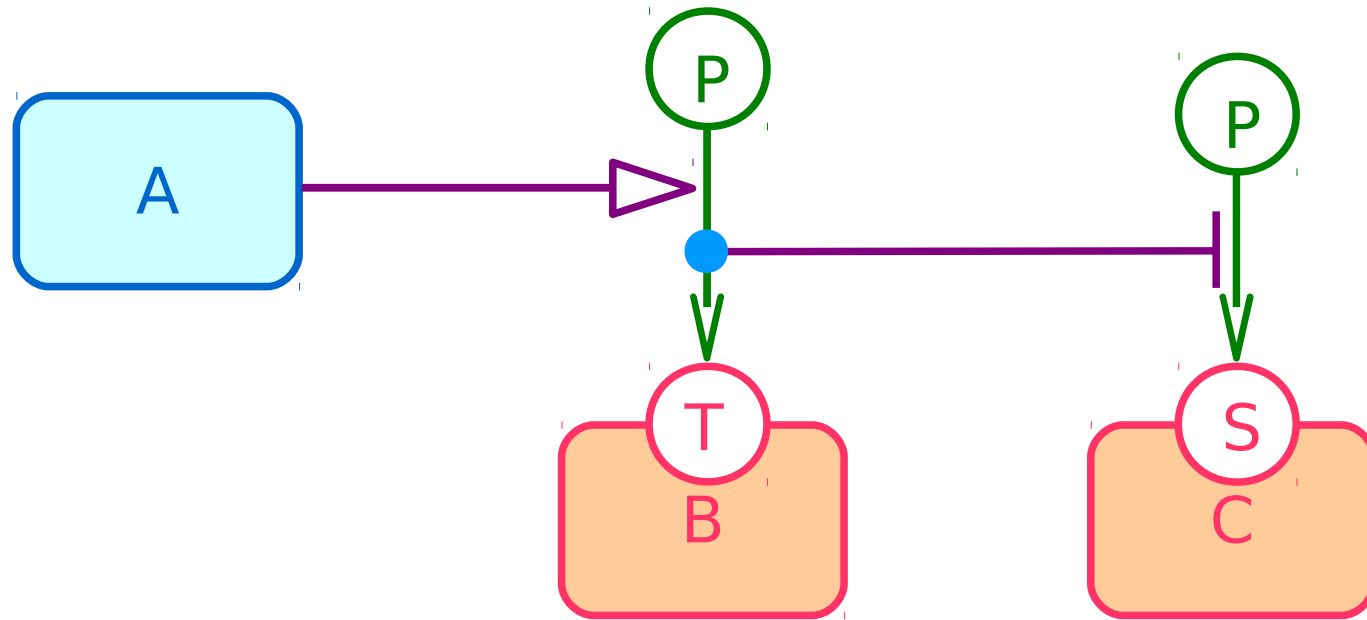
# Entity Relationships can be viewed as rules



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

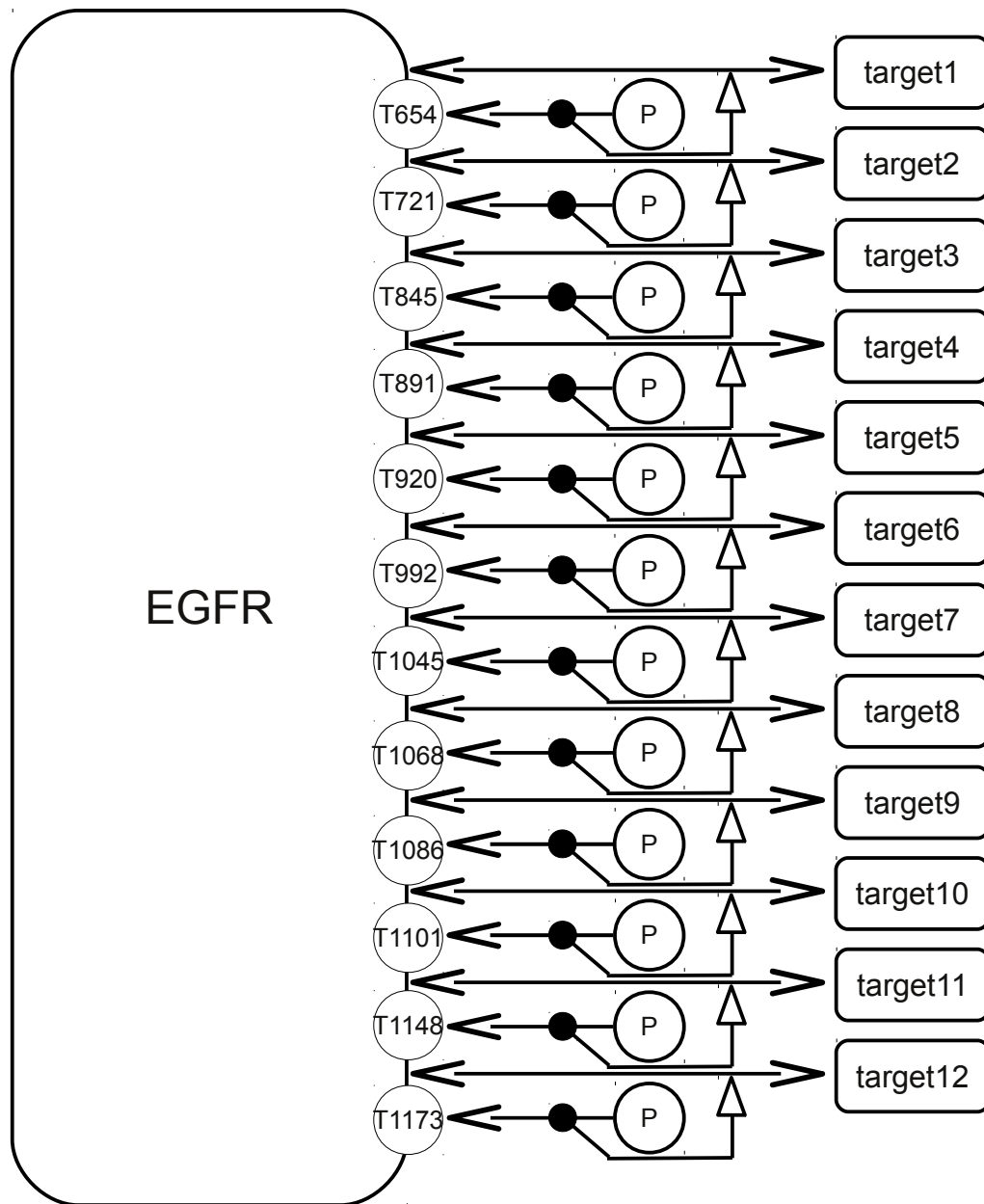


If **A** exists, the assignment of the value **P** to the state variable **T** of **B** is increased

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



# Multi-state and combinatorial explosion



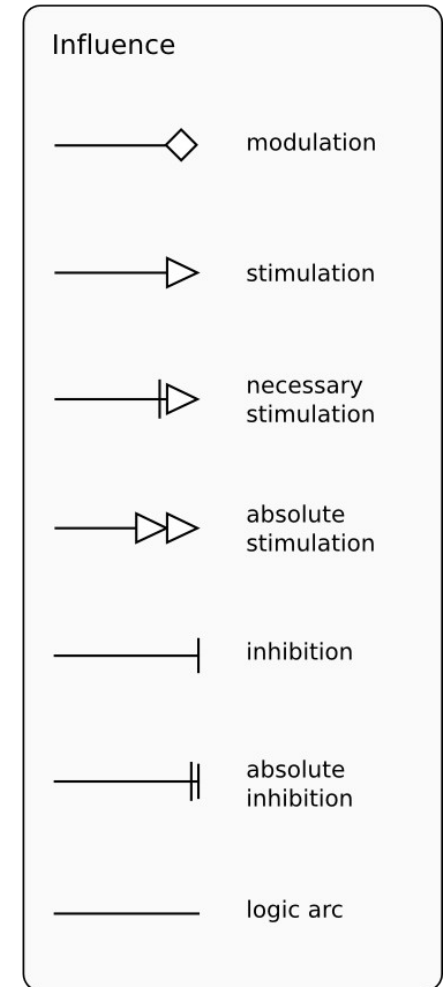
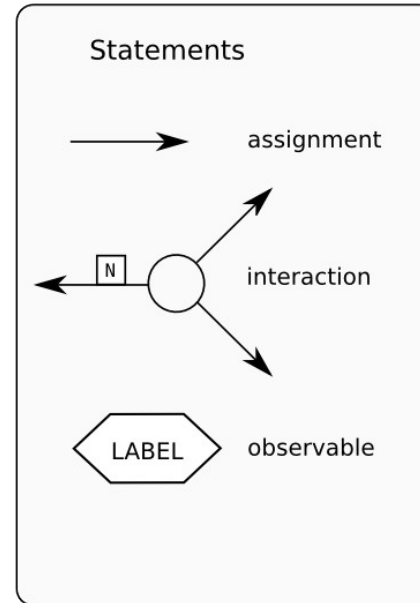
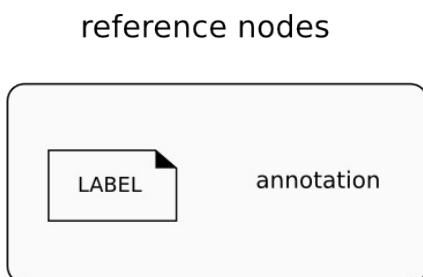
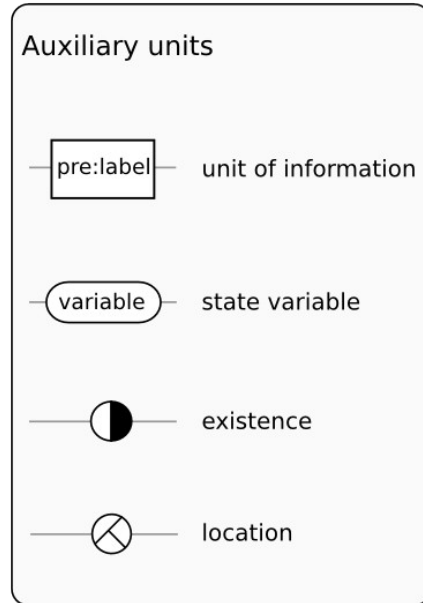
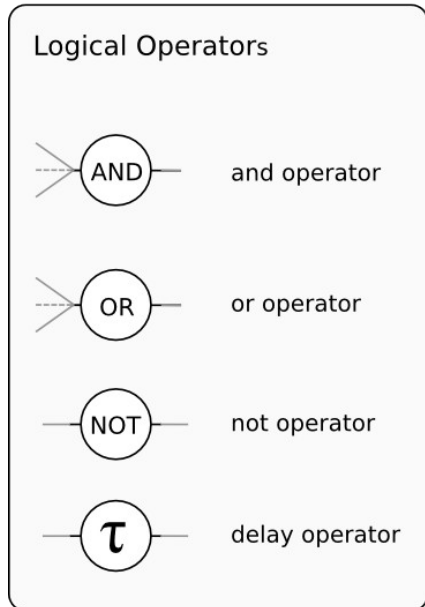
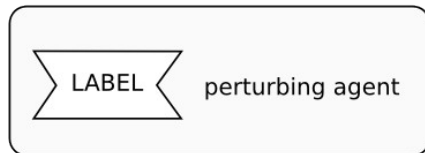
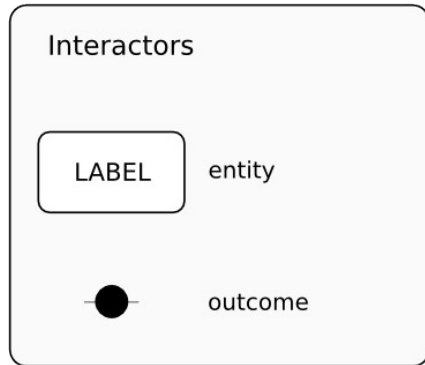
Process Descriptions:  
“once a state variable value,  
always a state variable value”

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

# Entity Relationships L1 V1 reference card

## Entity Nodes

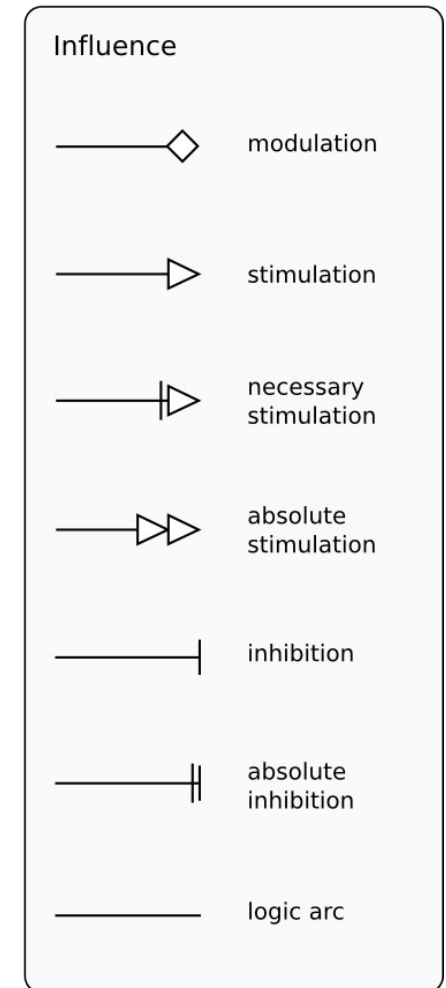
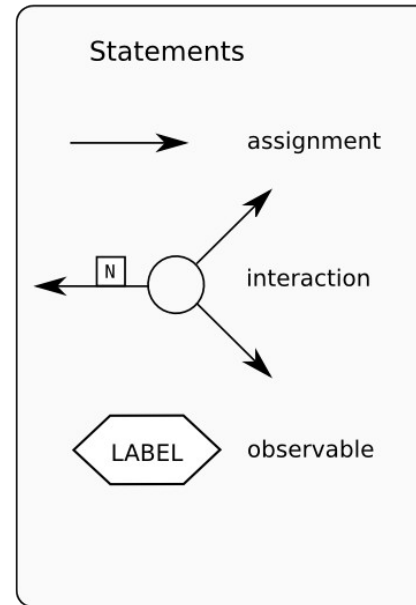
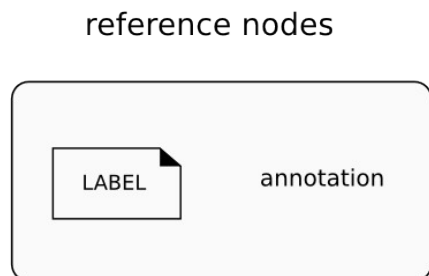
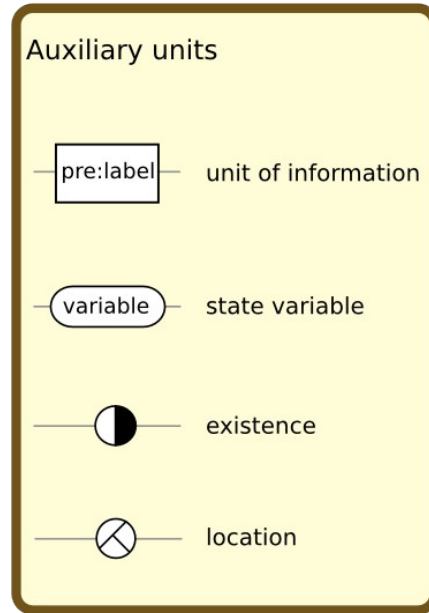
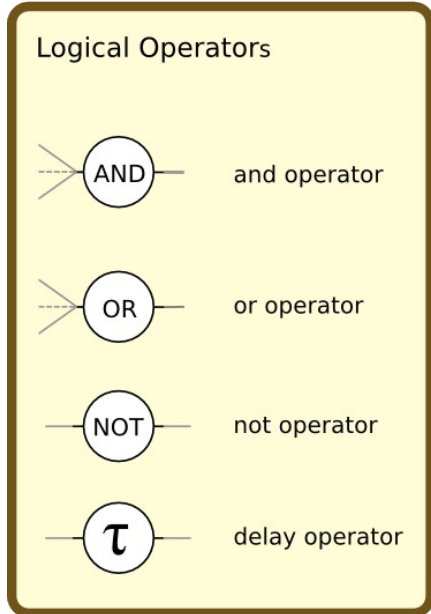
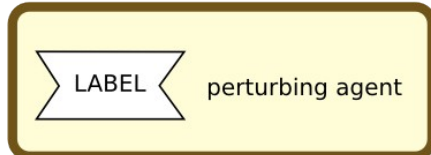
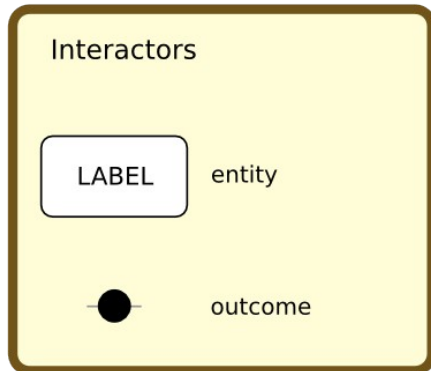
## Relationship Nodes



# Entity Relationships L1 V1 reference card

## Entity Nodes

## Relationship Nodes

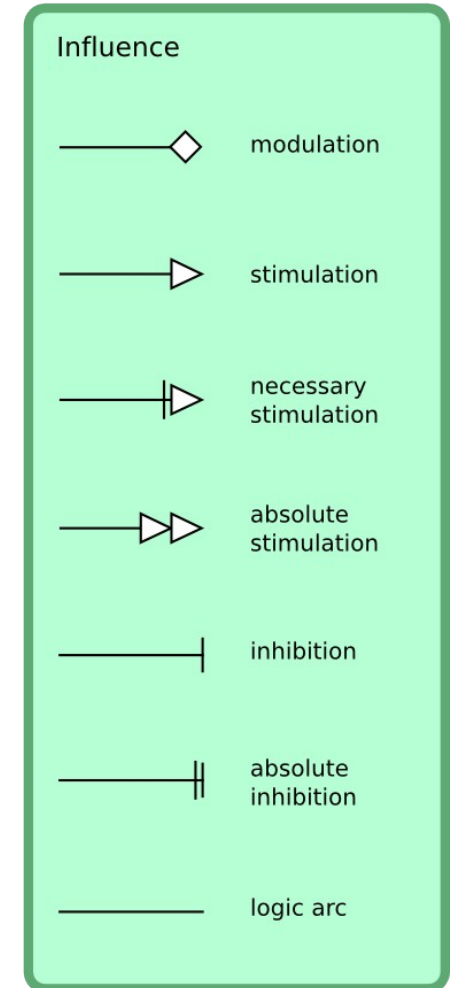
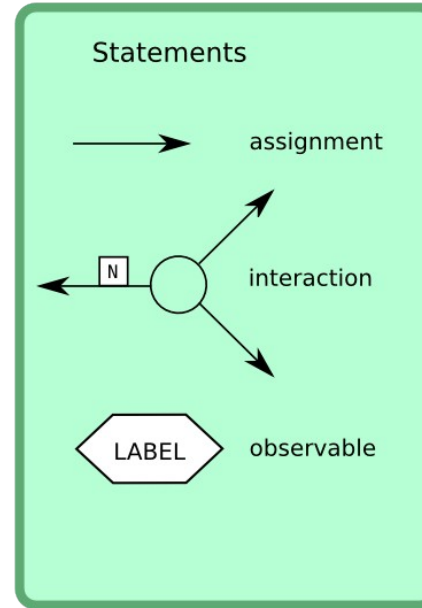
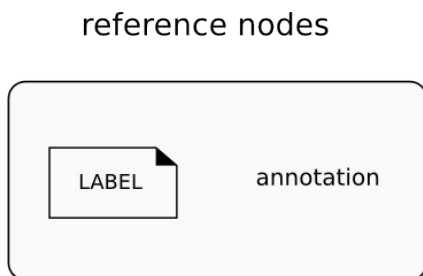
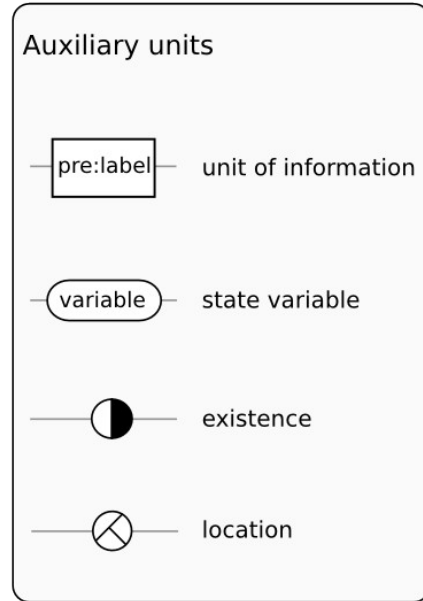
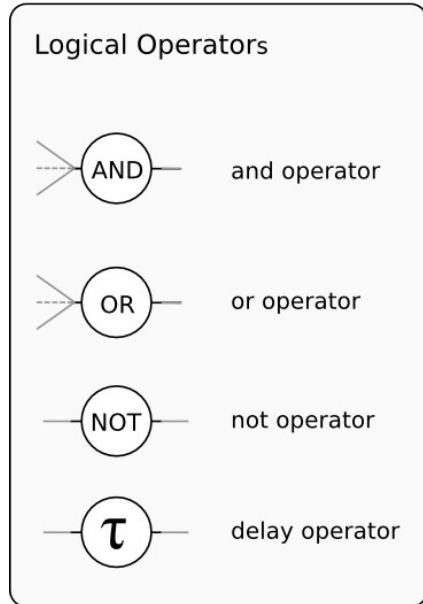
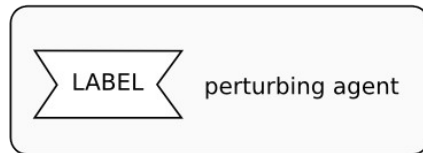
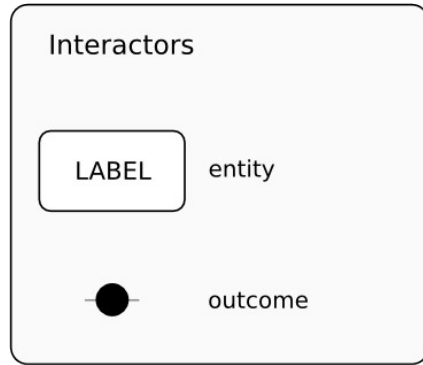


*continuants,*  
 things that exists (or not)

# Entity Relationships L1 V1 reference card

## Entity Nodes

## Relationship Nodes



*occurrences,*  
events that may happen (or not)

# Entity Relationships L1 V1 syntax

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	
<i>stimulation</i>				O	O	O	O	O	
<i>inhibition</i>				O	O	O	O	O	
<i>necessary stimulation</i>				O	O	O	O	O	
<i>absolute stimulation</i>				O	O	O	O	O	
<i>absolute inhibition</i>				O	O	O	O	O	
<i>assignment</i>				O	O	O	O	O	
<i>interaction</i>				O	O	O	O	O	
<i>phenotype</i>				O	O	O	O	O	

# Example of Entity Relationships L1 V1 semantics

## 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”.

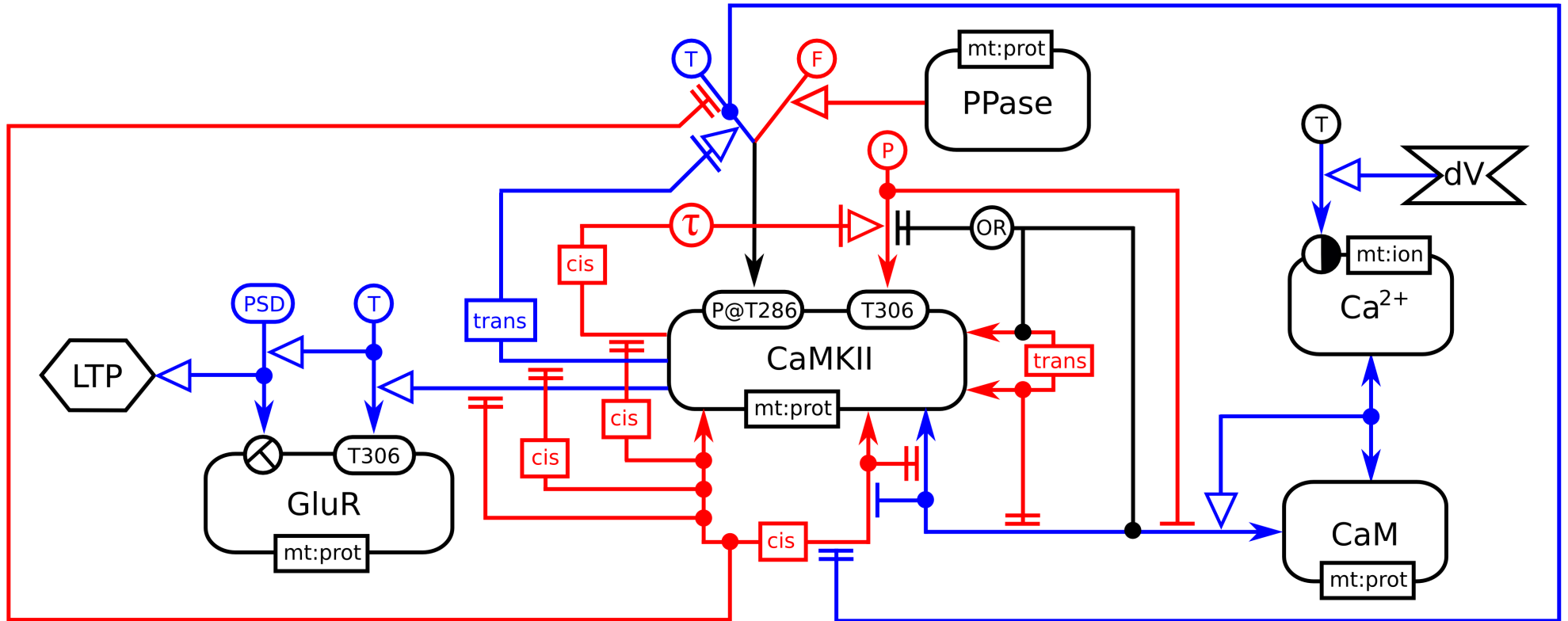
An *absolute stimulation* (Section 2.4.3.6) linking an *entity node*  $E$  and a relationship  $R$  means: “If  $E$  exists then  $R$  always takes place”.

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

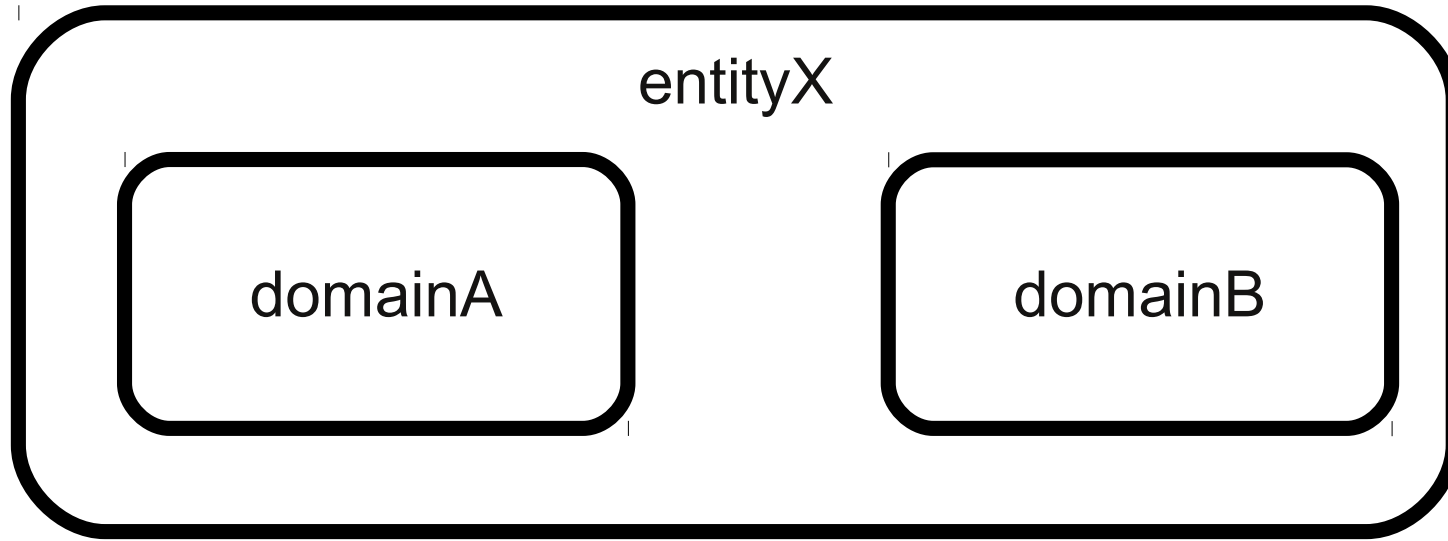
decreases synaptic weight

# Entity Relationships L1 V2

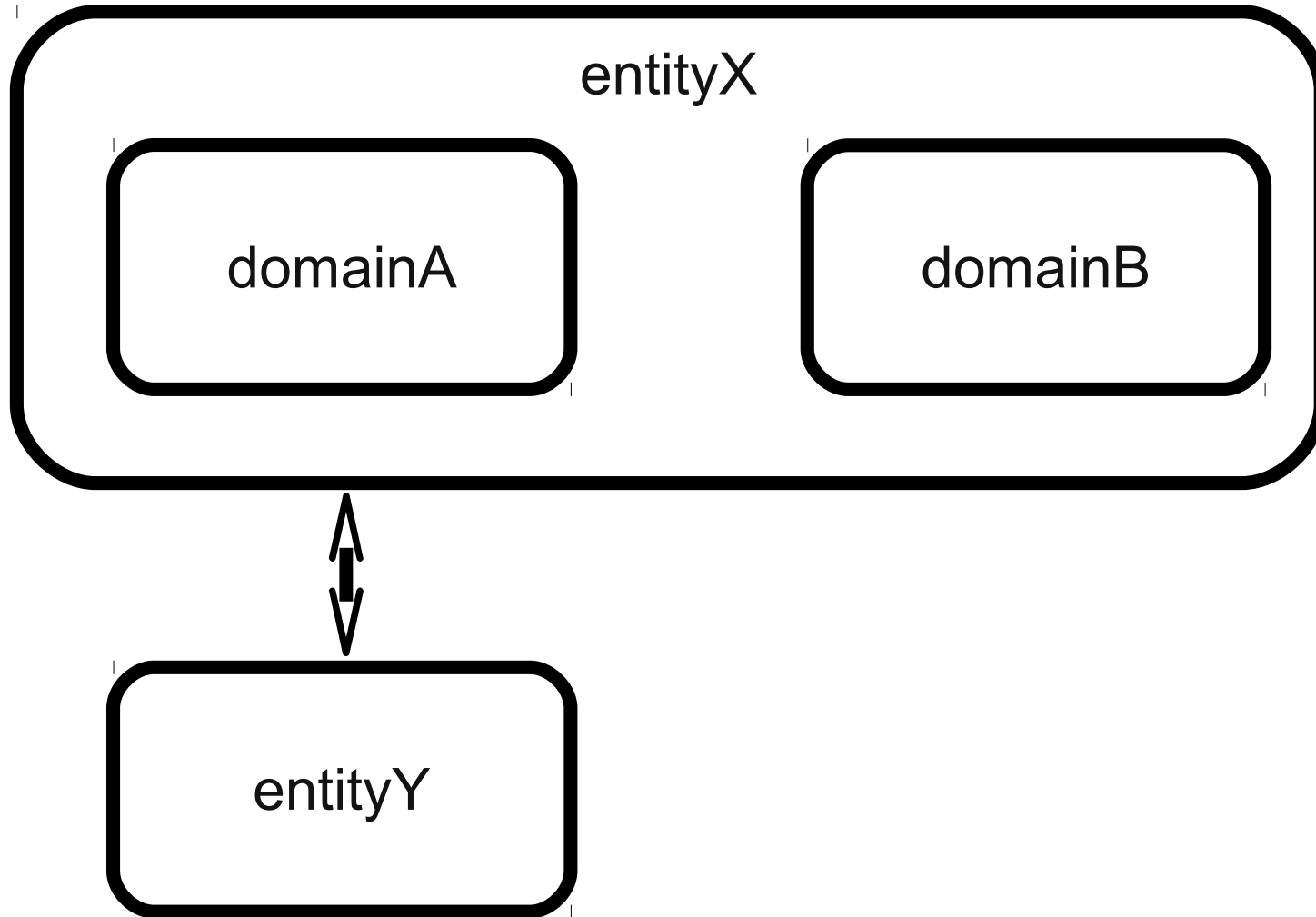
## *nested entities*



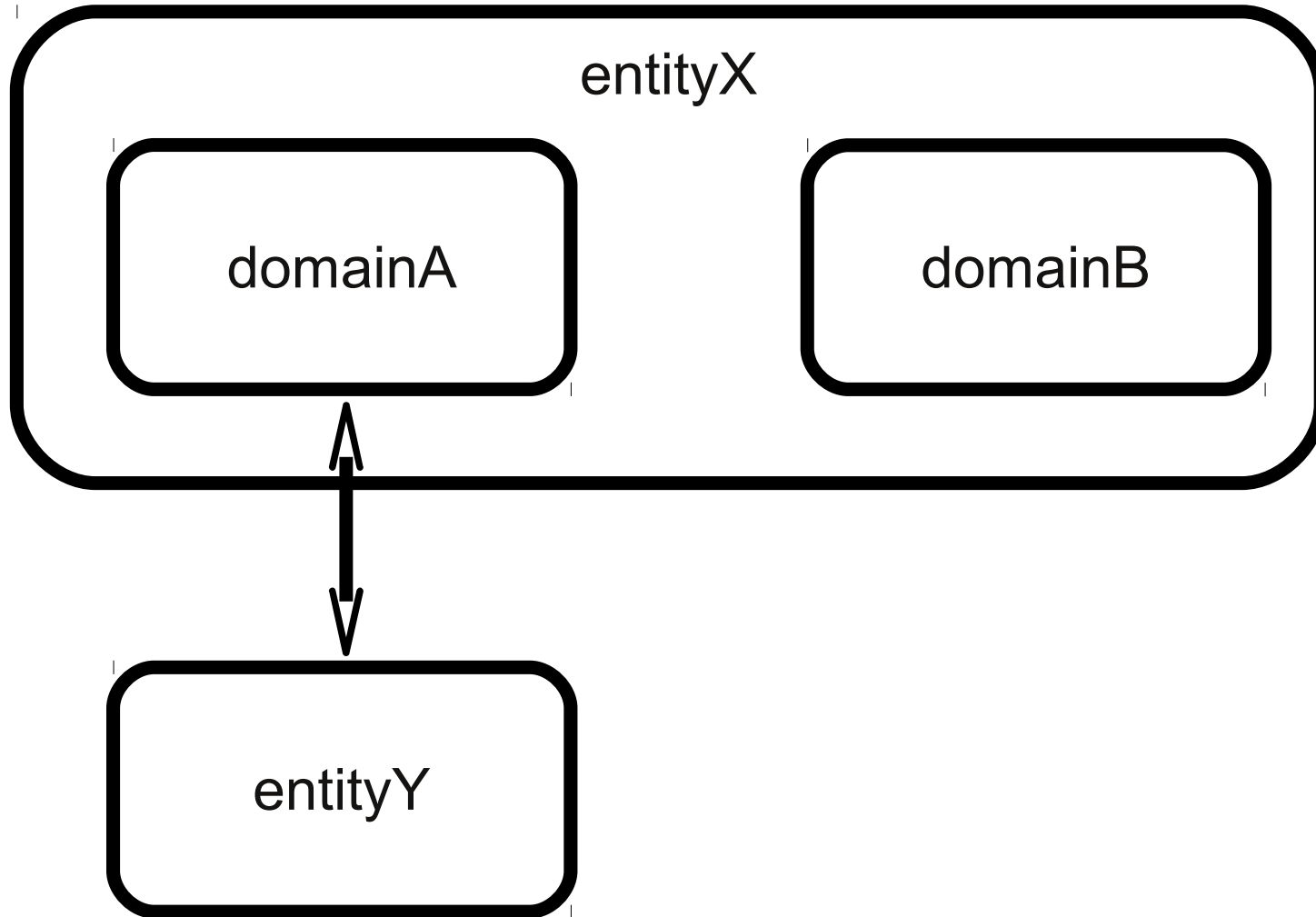
# A and B are part of X



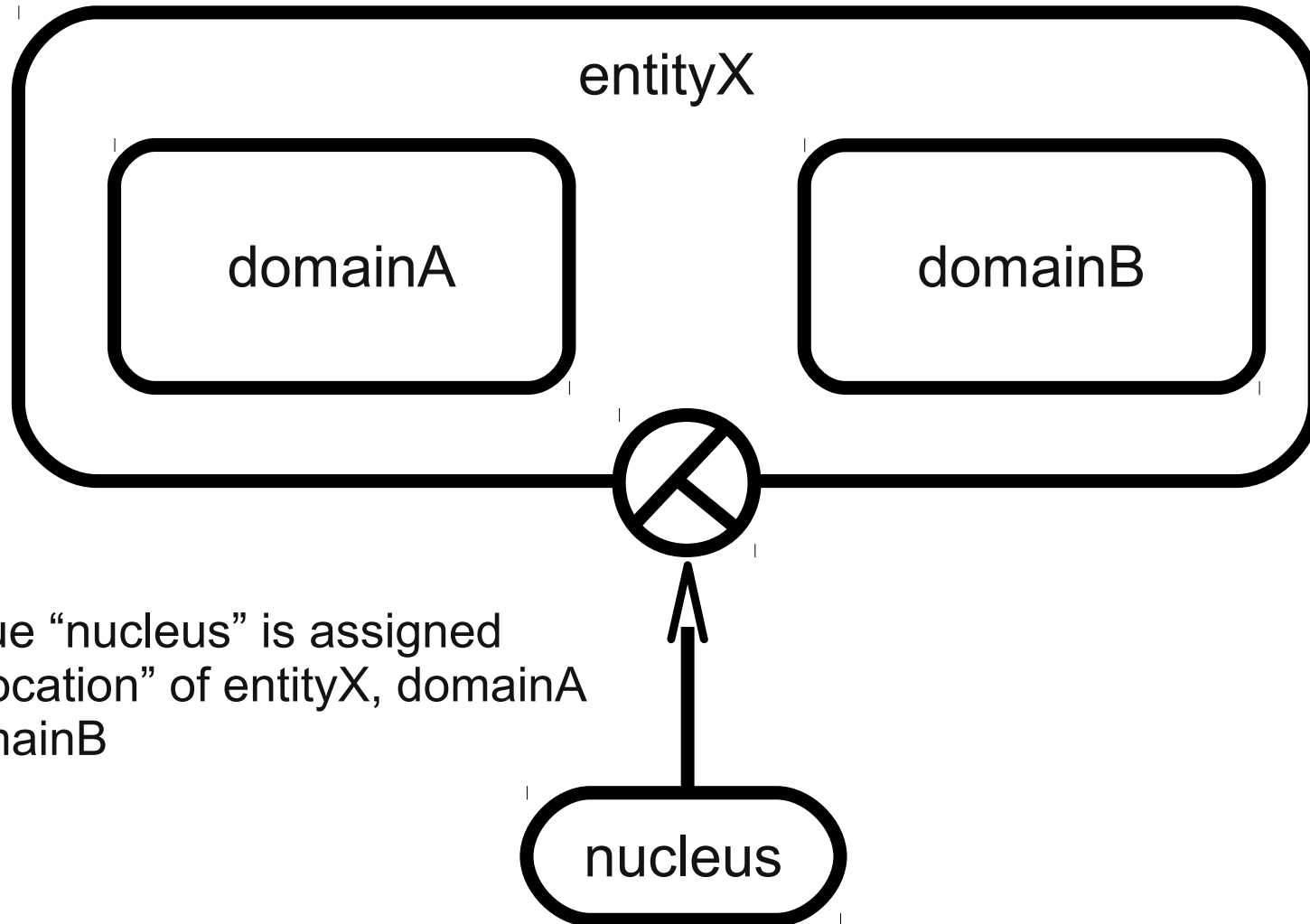
# X interacts with Y



# A of X interacts with Y

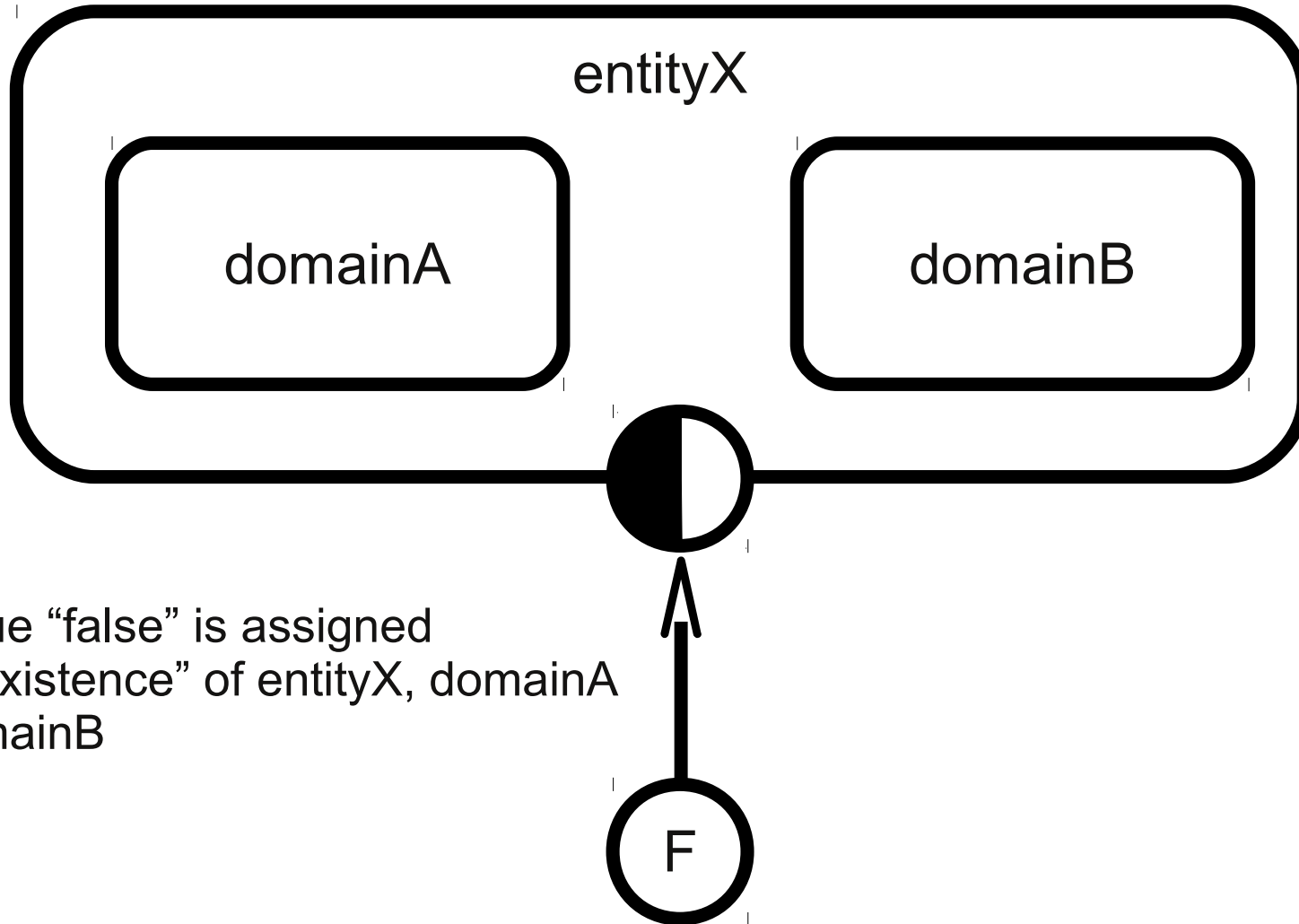


# Translocation of X in the nucleus



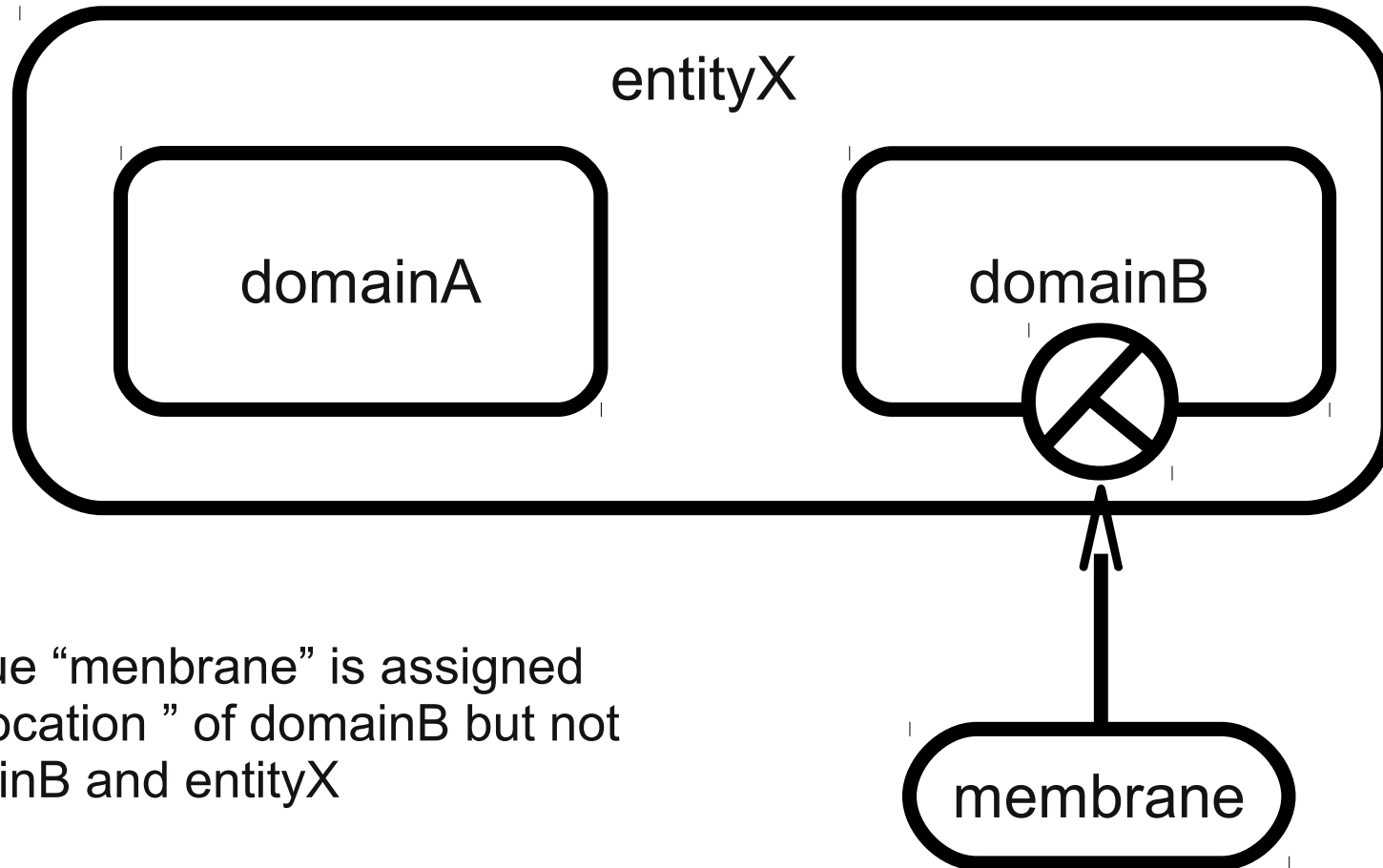
The value "nucleus" is assigned to the "location" of entityX, domainA and domainB

# degradation of X



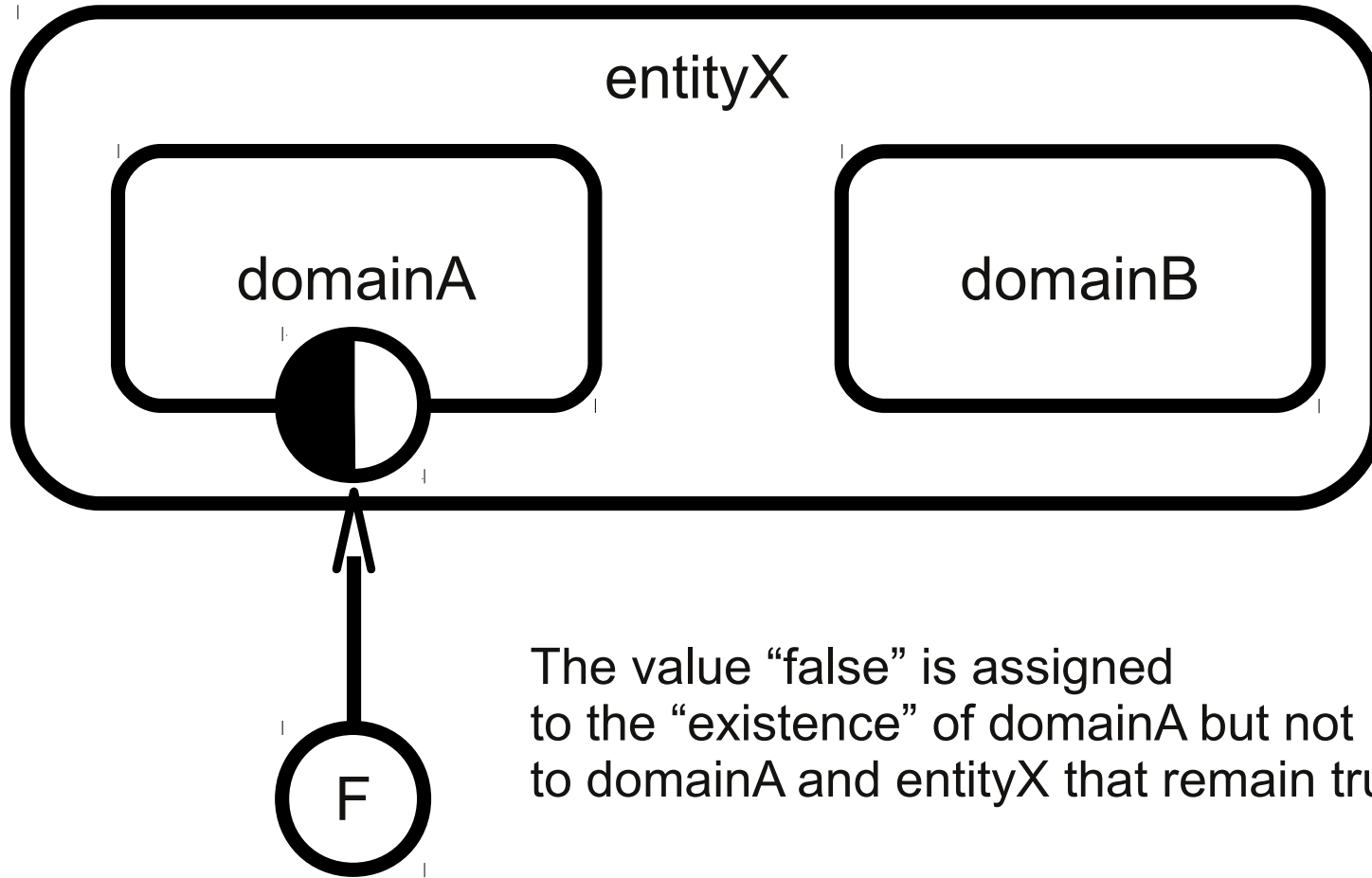
The value "false" is assigned to the "existence" of entityX, domainA and domainB

# insertion of B of X in the membrane



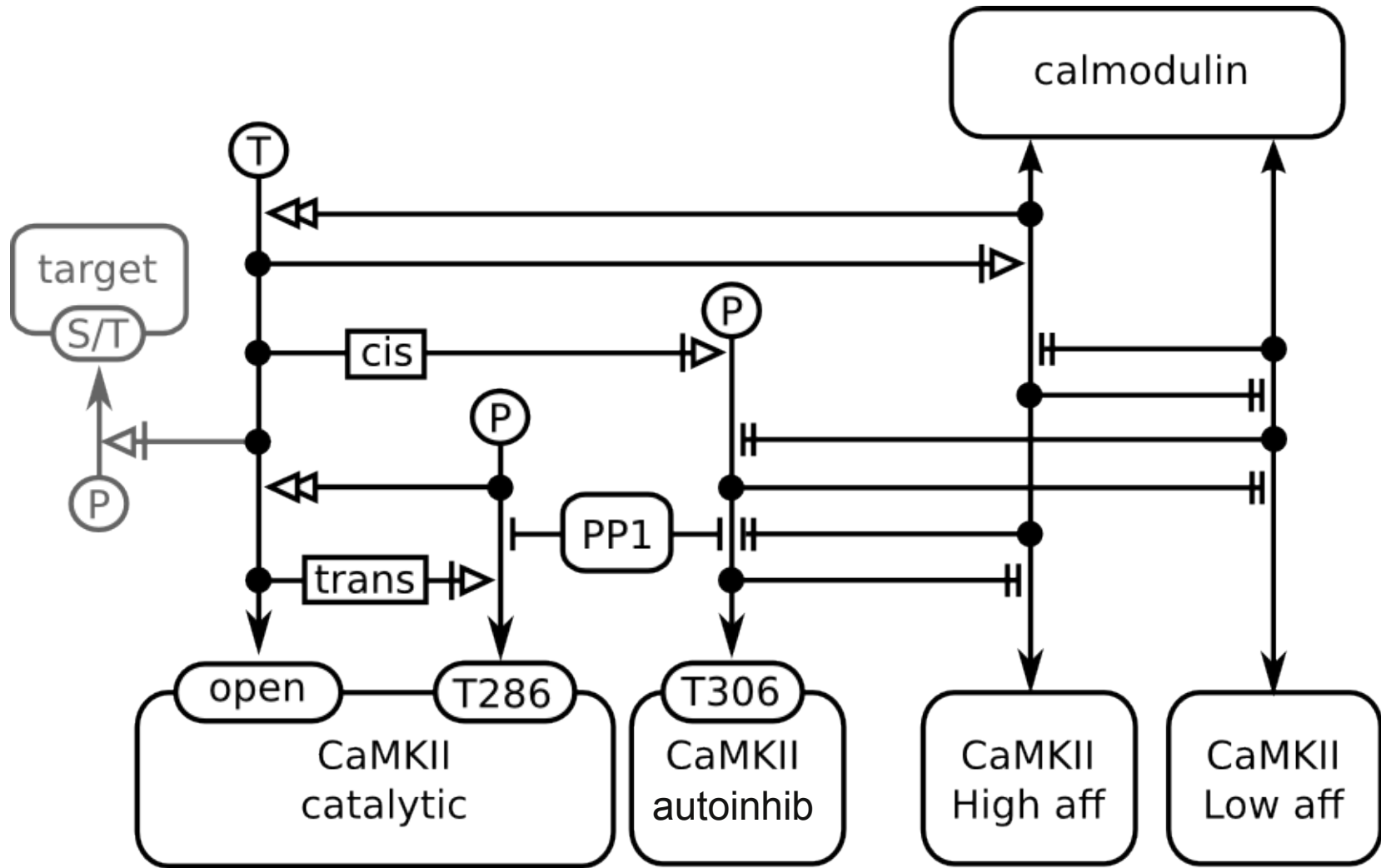
The value "membrane" is assigned to the "location" of domainB but not to domainB and entityX

# degradation of domain A of X



The value "false" is assigned to the "existence" of domainA but not to domainA and entityX that remain true

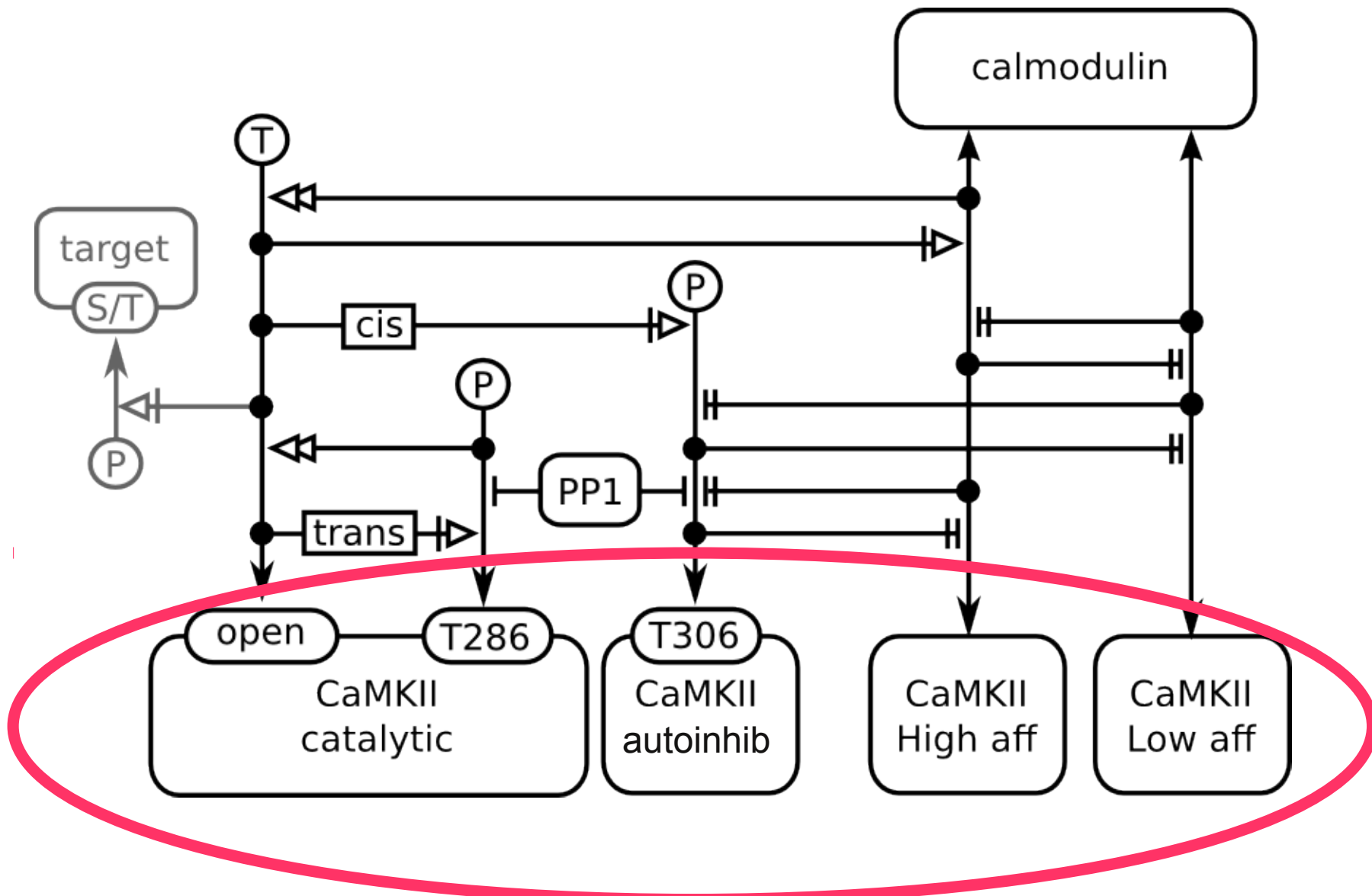
# Level 1 Version 1





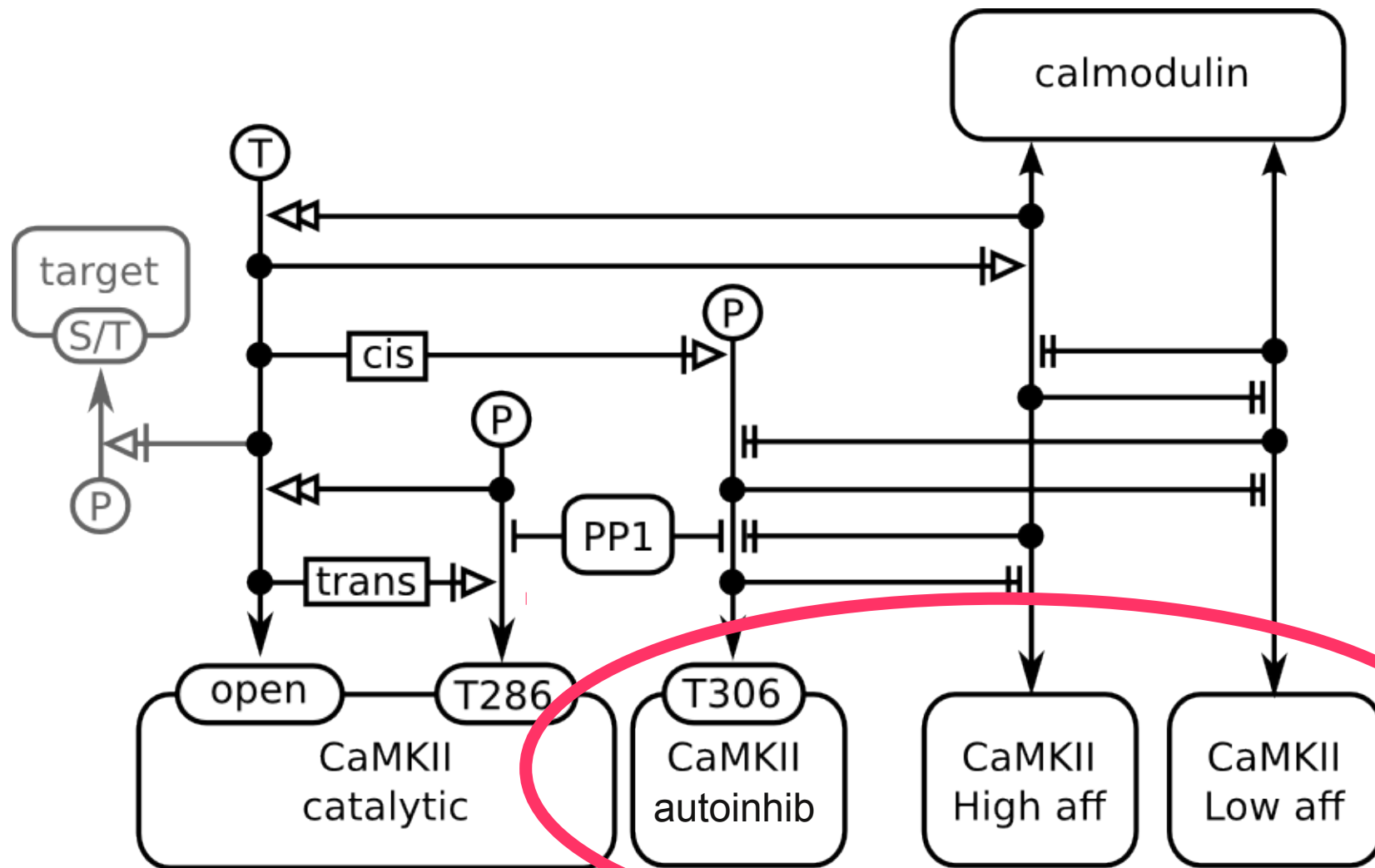
# Level 1 Version 1

Nature Precedings : doi:10.1038/npre.2010.4963.1 : Posted 8 Oct 2010



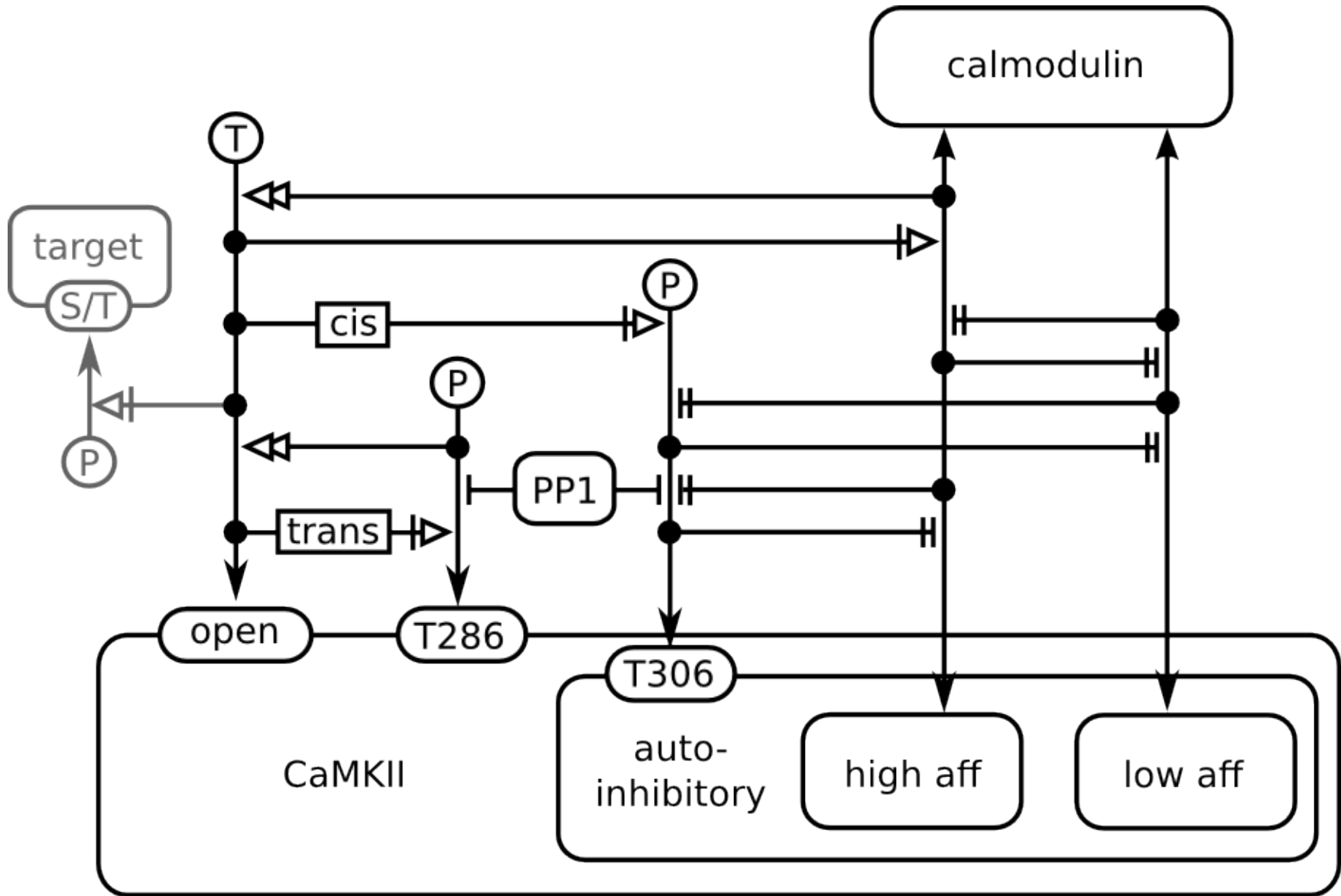
Same molecule=> confusing

# Level 1 Version 1



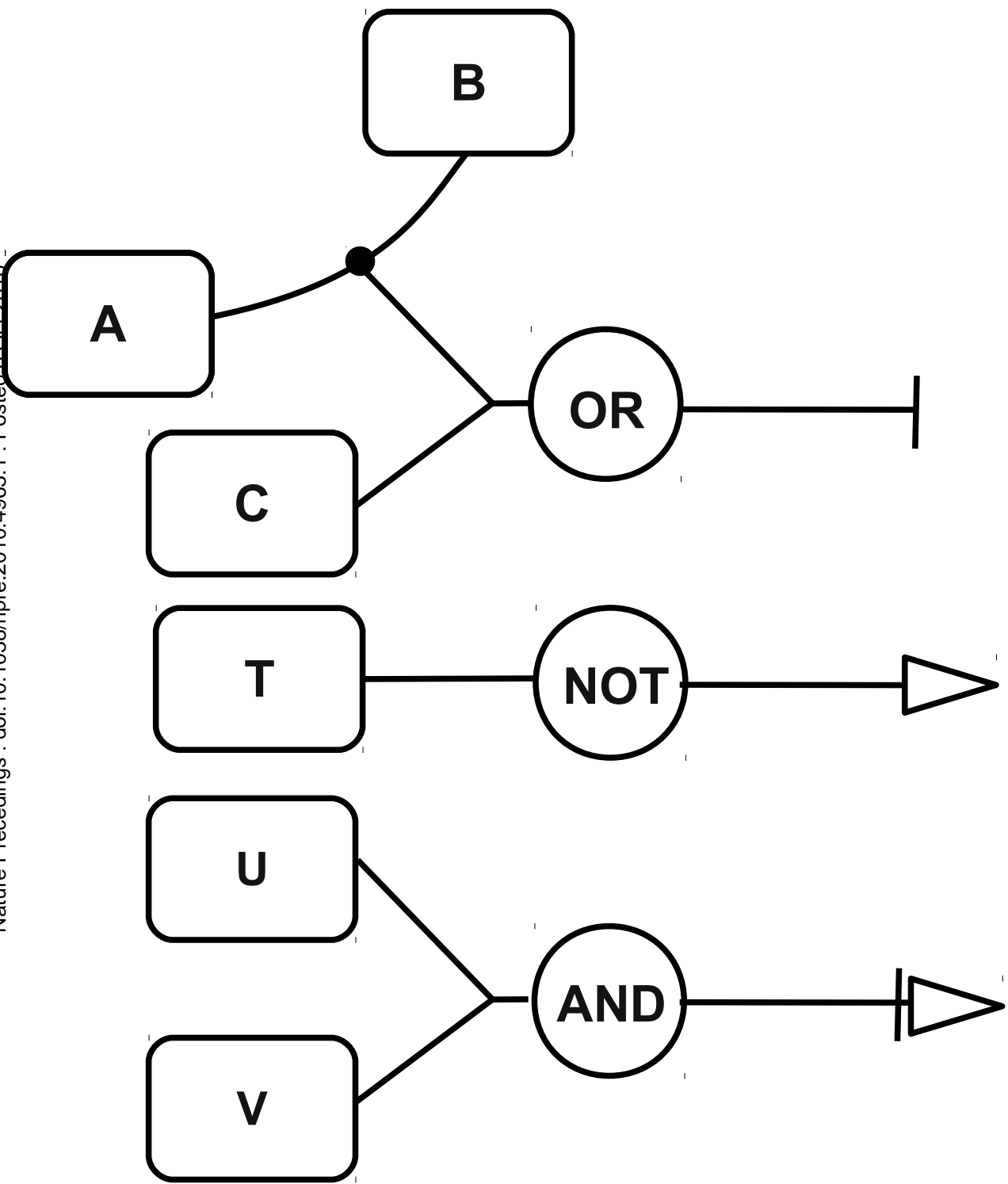
**T306 part of both low and high affinity => confusing**

# Level 1 Version 2



# Pending issues

- Logical operation on statements: In L1 V1, logical operators only output influences. Implicit XOR on variable assignments. Nothing for interactions.
- Topology of domains: Might be necessary if we want adoption by genomics, synthetic biology etc.
- 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).



either C or the complex of A and B inhibits ...

The absence of T stimulates ...

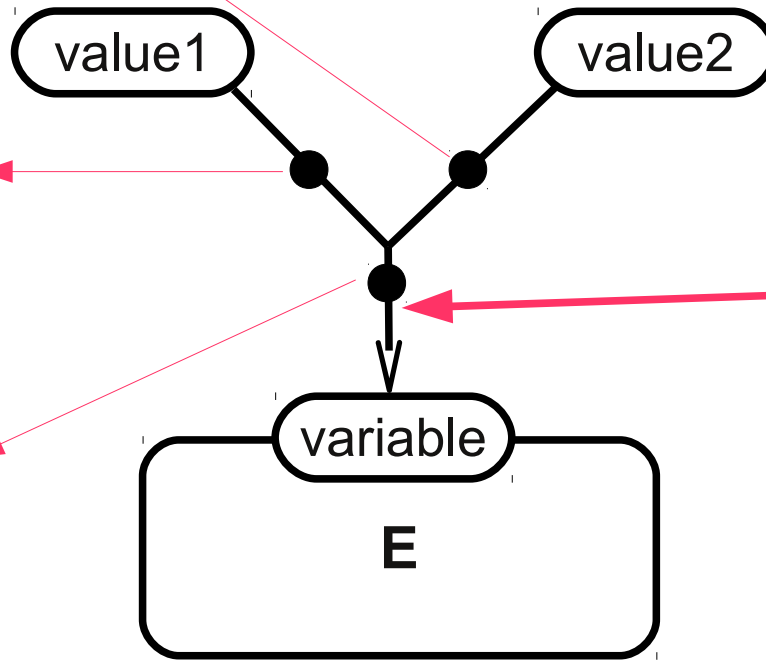
U and V are necessary to the existence of ...

# Implicit OR

variable has value2

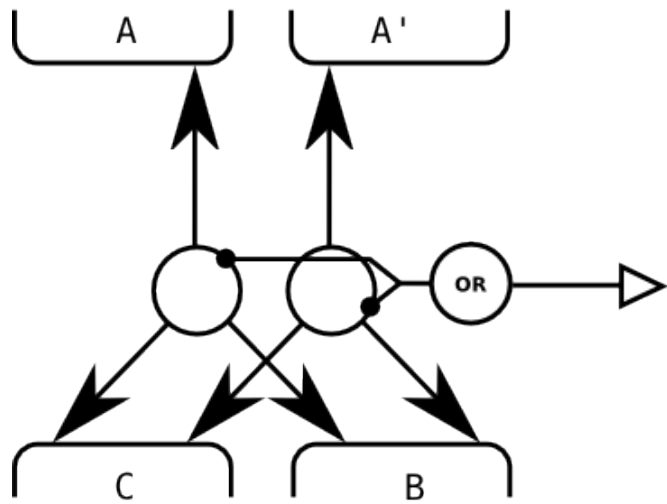
variable has value1

variable has value1  
or value2

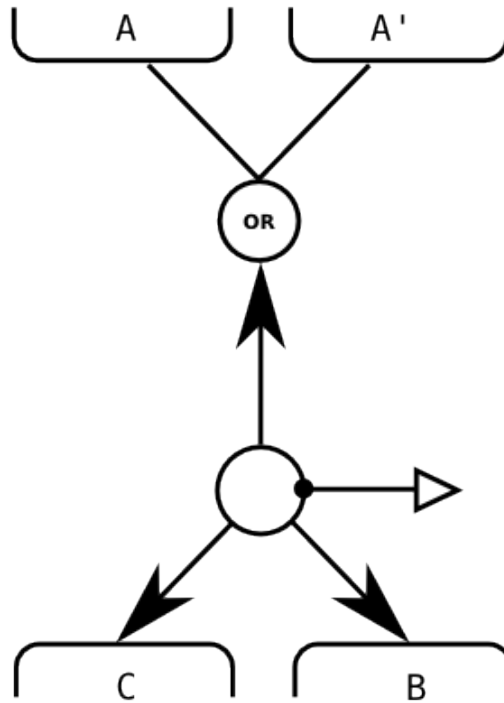


Only ONE arc, not a  
superposition of two.

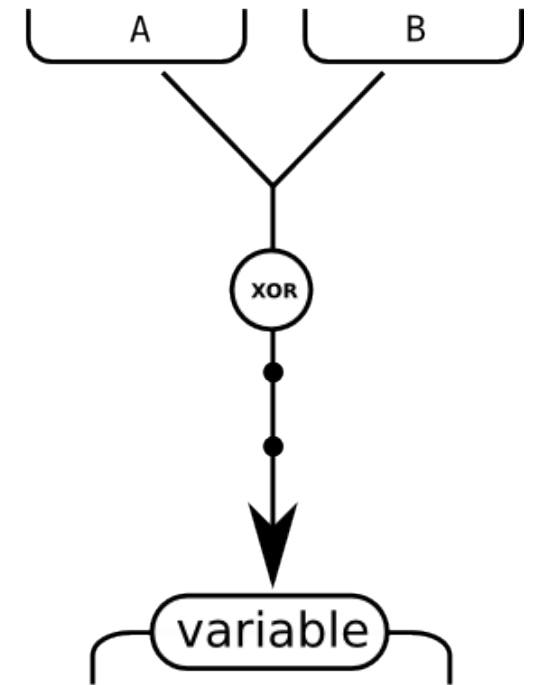
# Logical operator participating to statements



current situation  
for interactions



proposal for  
interactions



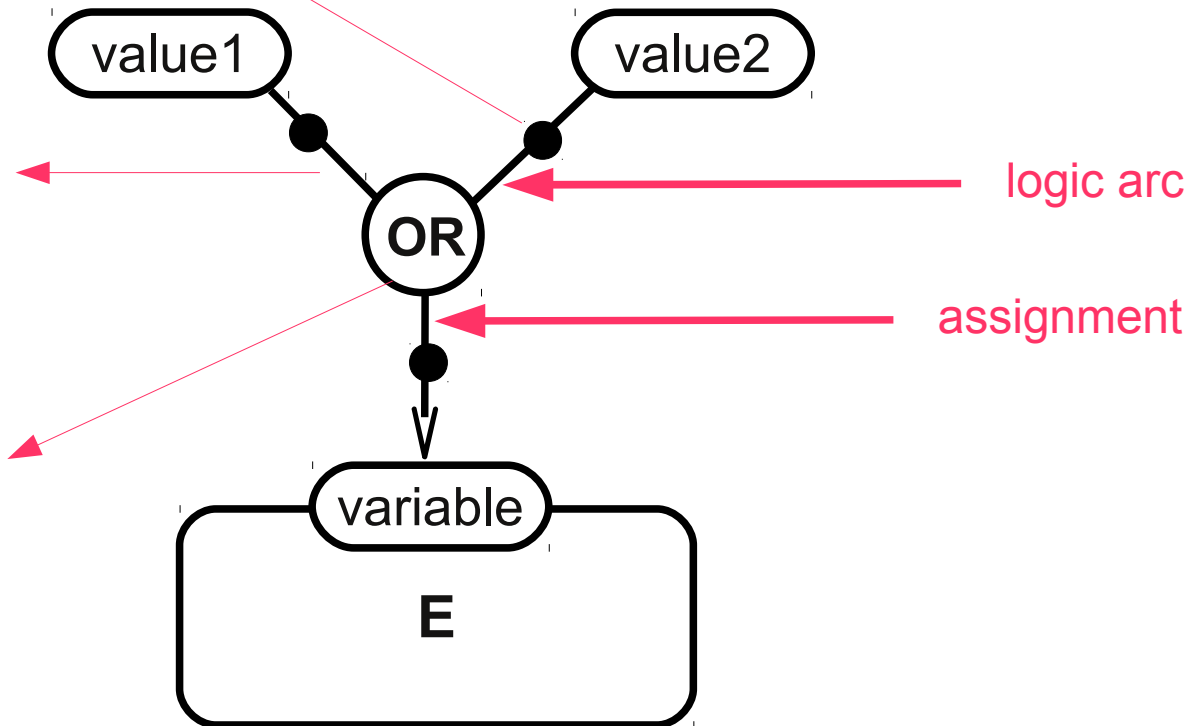
proposal for  
assignment

# Consequence: outcomes on influences

value2 is sent to variable

value1 is sent to variable

variable has value1 or value2



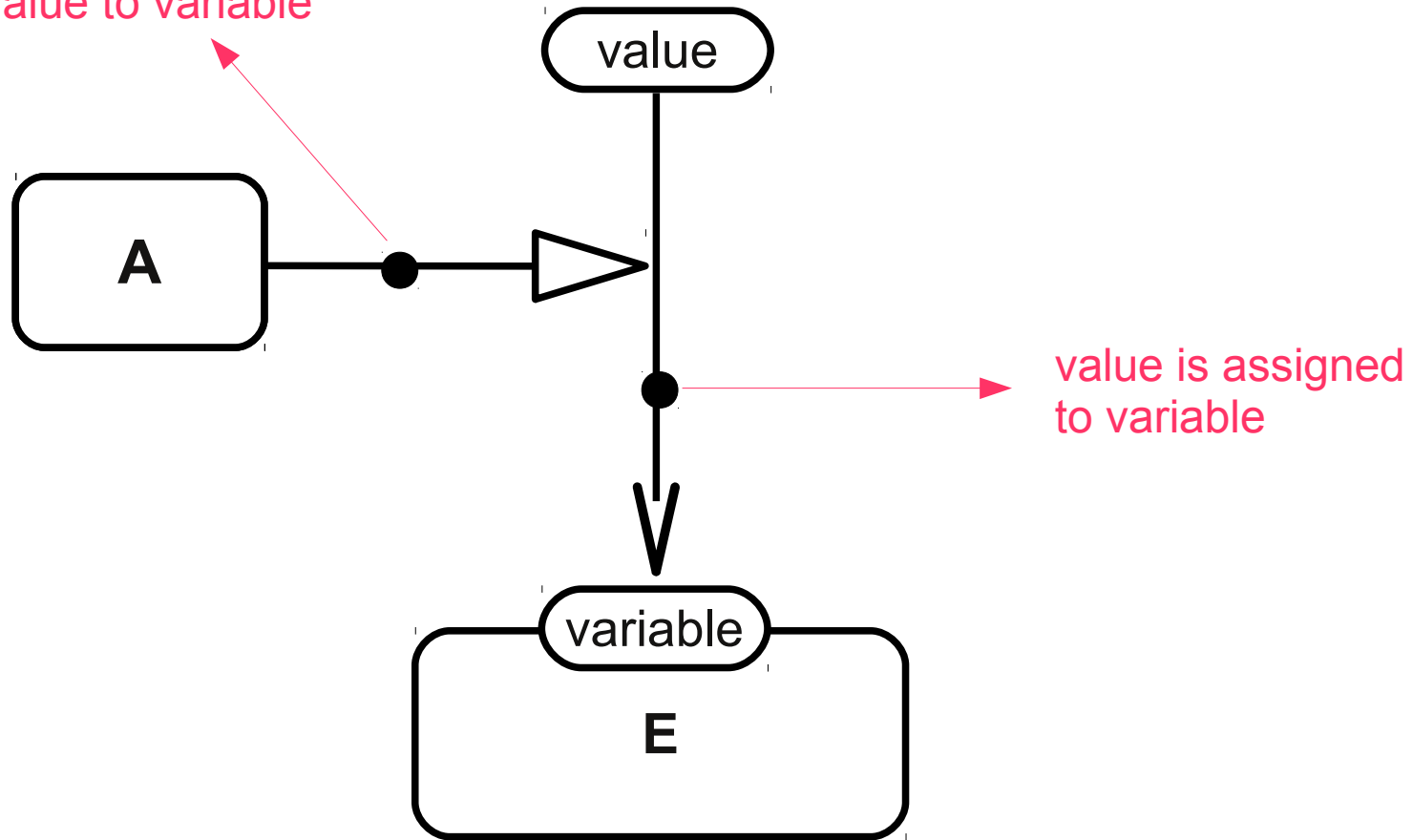
logic arc

assignment



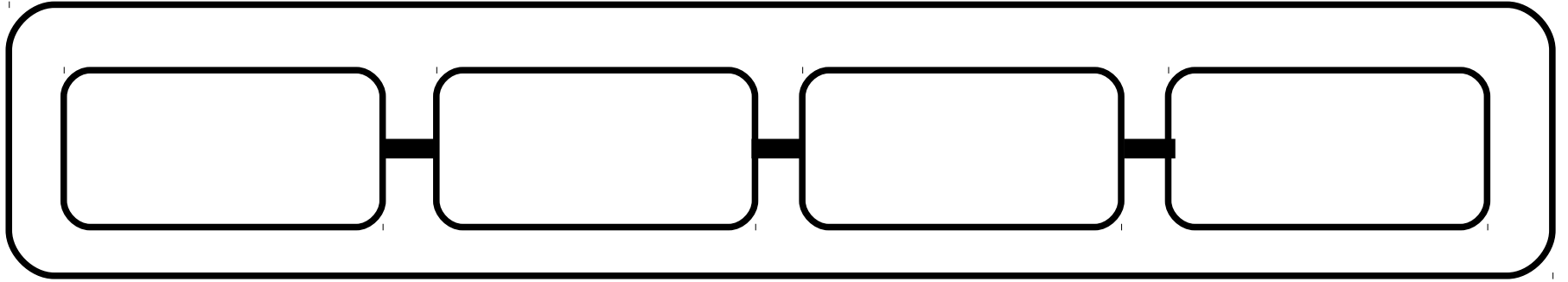
# On all influences ...

A stimulates the assignment  
of value to variable

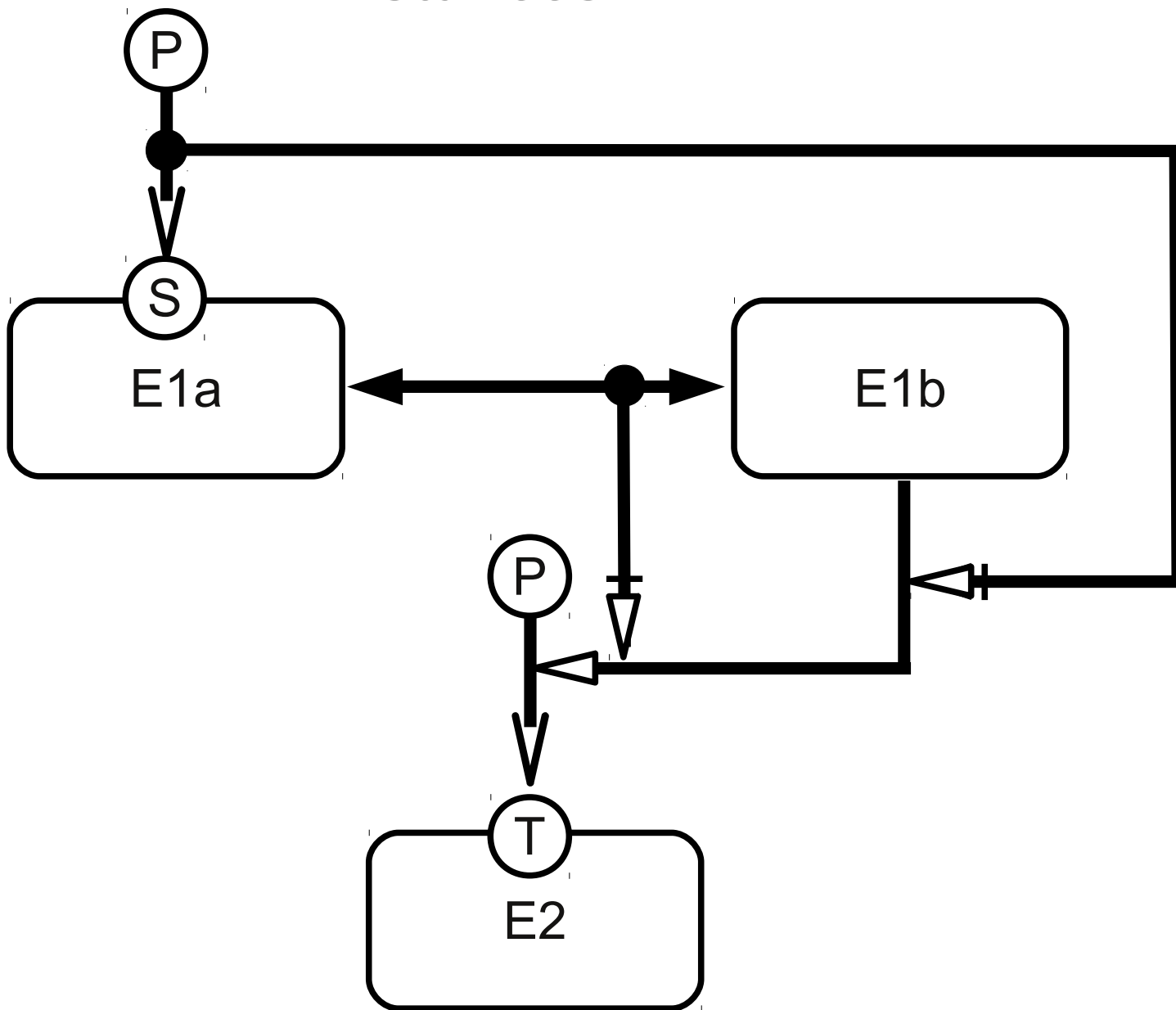


value is assigned  
to variable

# Topology of domains



# Instances in ER



# Generics in ER

