Systems Biology Graphical Notation: Entity Relationship language Level 1 Version 2

Anatoly Sorokin[1*], Nicolas Le Novère[2], Augustin Luna[3], Tobias Czauderna[4], Emek Demir[3], Robin Haw[5], Huaiyu Mi[6], Stuart Moodie[7], Falk Schreiber[8] and Alice Villéger[9]

[1]Institute of Cell Biophysics RAS, RU

[2]Babraham Institute, UK

[3]Memorial Sloan-Kettering Institute, USA

[4]Monash University, Australia

[5]Ontario Institute for Cancer Research, Canada

[6]University of Southern California, USA

[7]Eight Pillars Ltd, UK

[8]Monash University, Australia and MLU Halle, Germany

[9]Freelance IT Consultant, UK

Summary

The Systems Biological Graphical Notation (SBGN) is an international community effort for standardized graphical representations of biological pathways and networks. The goal of SBGN is to provide unambiguous pathway and network maps for readers with different scientific backgrounds as well as to support efficient and accurate exchange of biological knowledge between different research communities, industry, and other players in systems biology. Three SBGN languages, Process Description (PD), Entity Relationship (ER) and Activity Flow (AF), allow for the representation of different aspects of biological and biochemical systems at different levels of detail.

The SBGN Entity Relationship language (ER) represents biological entities and their interactions and relationships within a network. SBGN ER focuses on all potential relationships between entities without considering temporal aspects. The nodes (elements) describe biological entities, such as proteins and complexes. The edges (connections) provide descriptions of interactions and relationships (or influences), e.g., complex formation, stimulation and inhibition. Among all three languages of SBGN, ER is the closest to protein interaction networks in biological literature and textbooks, but its well-defined semantics offer a superior precision in expressing biological knowledge.

*To whom correspondence should be addressed. Email: sbgn-editors@lists.sourceforge.net

Systems Biology Graphical Notation: Entity Relationship language Level 1

Version 2.0

Date: August 8, 2015

Editors:

Anatoly Sorokin	*Institute of Cell Biophysics RAS, RU*
Nicolas Le Novère	*Babraham Institute, UK*
Augustin Luna	*Memorial Sloan-Kettering Institute, USA*
Tobias Czauderna	*Monash University, Australia*
Emek Demir	*Memorial Sloan-Kettering Institute, USA*
Robin Haw	*Ontario Institute for Cancer Research, Canada*
Huaiyu Mi	*University of Southern California, USA*
Stuart Moodie	*Eight Pillars Ltd, UK*
Falk Schreiber	*Monash University, Australia & MLU Halle, Germany*
Alice Villéger	*Freelance IT Consultant, UK*

To discuss any aspect of SBGN, please send your messages to the mailing list sbgn-discuss@caltech.edu. *To get subscribed to the mailing list or to contact us directly, please write to* sbgn-editors@lists.sourceforge.net. *Bug reports and specific comments about the specification should be entered in the issue tracker* http://sourceforge.net/p/sbgn/sbgn-er-l1/.

doi:10.2390/biecoll-jib-2015-264

Contents

doi:10.2390/biecoll-jib-2015-264

Chapter 1

Introduction

The goal of the **S**ystems **B**iology **G**raphical **N**otation (SBGN) is to standardize the graphical/visual representation of essential biochemical and cellular processes. SBGN defines comprehensive sets of symbols with precise semantics, together with detailed syntactic rules defining their use. It also describes the manner in which such graphical information should be interpreted. For a general description of SBGN goals, one can read:

> Nicolas Le Novère, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, Sarala M Wimalaratne, Frank T Bergman, Ralph Gauges, Peter Ghazal, Hideya Kawaji, Lu Li, Yukiko Matsuoka, Alice Villéger, Sarah E Boyd, Laurence Calzone, Melanie Courtot, Ugur Dogrusoz, Tom C Freeman, Akira Funahashi, Samik Ghosh, Akiya Jouraku, Sohyoung Kim, Fedor Kolpakov, Augustin Luna, Sven Sahle, Esther Schmidt, Steven Watterson, Guanming Wu, Igor Goryanin, Douglas B Kell, Chris Sander, Herbert Sauro, Jacky L Snoep, Kurt Kohn, Hiroaki Kitano. The Systems Biology Graphical Notation. Nature Biotechnology 27, 735 - 741 (2009). http://dx.doi.org/10.1038/nbt.1558

This document defines the *Entity Relationship* visual language of SBGN. Entity Relationships are one of three views of a biological process offered by SBGN, the others being Process Descriptions and Activity Flows. SBGN Entity Relationship language allows to see all the relationships in which a given entity participates, regardless of the temporal aspects. Entities are defined here in a broad sense as something that can exists. Relationships can be seen as rules describing the influences of entities on other relationships. An overview of Entity Relationships is given in Section 1.1.

1.1 Overview of the Entity Relationship language

To set the stage for what follows in this chapter, we first give a brief overview of some of the concepts in the Entity Relationship language with the help of an example shown in Figure 1.1 on the following page.

doi:10.2390/biecoll-jib-2015-264

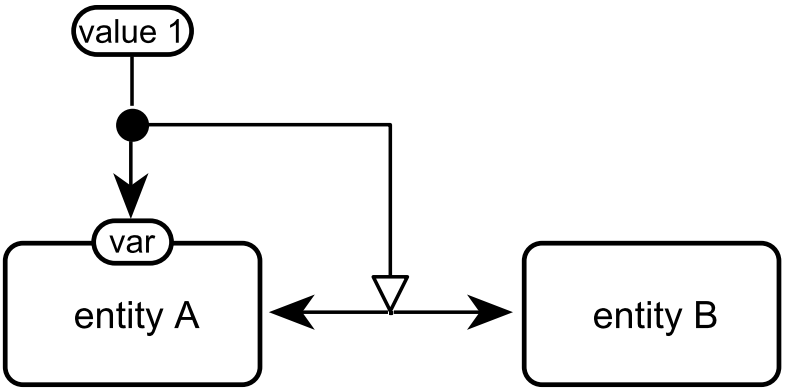

Figure 1.1: *This example of an Entity Relationship map represents an entity A interacting with another entity B. The assignment of value 1 to the state variable var of the entity A stimulates the interaction between entity A and entity B.*

The essence of the Entity Relationships is to depict the influences of entities upon the behaviour of others. The entities are classes of things that can exist, either on their own, or when statements (interactions and assignments) become true, so existence is the key concept in Entity Relationships. For entity to participate in relationship at least one of it's instance should exist. Instances of different entities can interact (such as "entity A" and "entity B" in Figure 1.1), or a value can be assigned to a property of some instance of an entity (such as "value 1" to property "var" of "entity A" in Figure 1.1). The entities can modify the interactions between other entities, for example, stimulation of the interaction between "entity A" and "entity B" in Figure 1.1 takes place if "value 1" is really assigned to property "var" of some instance of the "entity A". The existence of the assignment (set of instances where value is really assigned to the property) is represented by the black dot (called *outcome*). The influences can therefore be understood as logical consequences of existence of participating instances. On the contrary to the Process Description language, where the different processes affect each other in a way that the behaviour of the depicted system can only be understood taking into account the whole system, the relationships are essentially independent. One can imagine that each of the relationships represents a specific conclusion of a scientific experiment reported in an article. Their addition on a map represents the knowledge we have of the effects of the entities represented upon each other. In Figure 1.1, we have three statements:

- "entity A" can interact with "entity B".
- property "var" of "entity A" can take the value "value 1".
- if "var" is set to "value 1", the interaction between "entity A" and "entity B" is stimulated.

The independence of relationships in Entity Relationships is the key to avoid the combinatorial explosions inherent to Process Descriptions.

A more complex and realistic example of Entity Relationship map is shown in Figure 1.2 on the next page. This example will be re-used throughout the description of the graphical symbols (glyphs) used by SBGN Entity Relationship Level 1 (with a few additions when the concepts are missing in the example)

doi:10.2390/biecoll-jib-2015-264

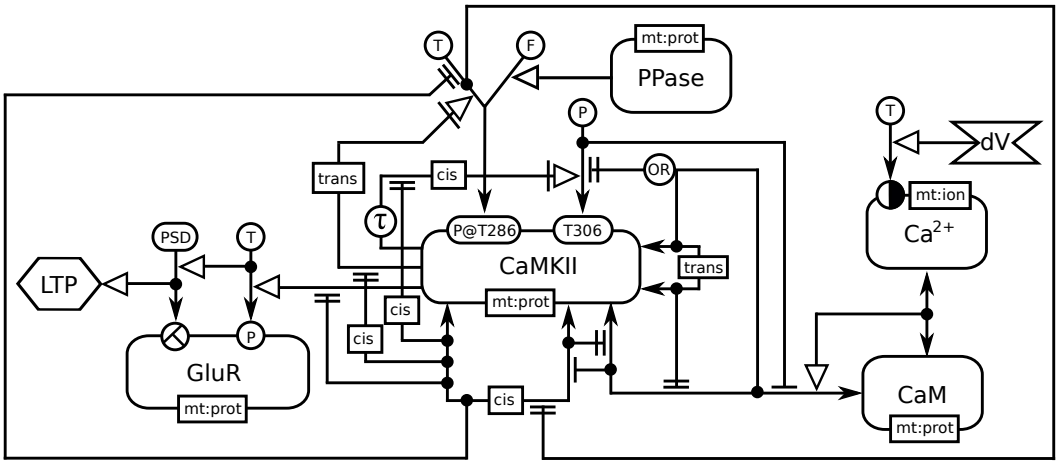

Figure 1.2: *This example of an Entity Relationship map depicts the effect of a depolarisation (dV) on the intracellular calcium, that binds to calmodulin, that itself binds to the calcium/calmoduline kinase II (CaMKII). The binding of calmodulin inhibits the folding of CaMKII monomer on itself, thus relieving the inhibition on the kinase activity. The phosphorylation of the glutamate receptors finally leads to the Long Term Potentiation (LTP) of the synapses. In addition, the map shows the effect of trans-phosphorylation on threonine 286, that makes the enzyme constitutively active, and on threonine 306, that renders the kinase insensitive to calmodulin, as well as the dimerisation of the kinase.*

Table 1.1 summarizes the different SBGN abstractions described in this chapter.

Component	Role	Examples
Entity node	Something that exists, whether a physical object or sets of objects.	An entity, the result of an interaction
Statement arc	Something that can be true or false, and affects or relates entities.	An interaction between entities, the assignment of a value to a variable
Influence	The effect of something true on the realisation of a statement or another influence.	A stimulation, an absolute inhibition

Table 1.1: *Summary of Entity Relationship components and their roles.*

1.2 SBGN levels and versions

It was clear at the outset of SBGN development that it would be impossible to design a perfect and complete notation right from the beginning. Apart from the prescience this would require (which, sadly, none of the authors possess), it also would likely require a vast language that most newcomers would shun as being too complex. Thus, the SBGN community followed an idea used in the development of other standards, i.e. stratify language development into levels.

A *level* of one of the SBGN languages represents a set of features deemed to fit together cohesively, constituting a usable set of functionality that the user community agrees is sufficient for a reasonable set of tasks and goals. Within *levels*, *versions* represent small evolution of a language, that may involve new glyphs, refined semantics, but no fundamental change of the way maps are to be generated and interpreted. Capabilities and features that cannot be agreed upon and are judged insufficiently critical to require inclusion in a given level, are postponed to a higher level or version. In this way, the development of SBGN languages is envisioned to proceed in stages, with each higher levels adding richness compared to the levels below it.

doi:10.2390/biecoll-jib-2015-264

1.3 Developments, discussions, and notifications of updates

The SBGN website (http://sbgn.org/) is a portal for all things related to SBGN. It provides a web forum interface to the SBGN discussion list (sbgn-discuss@caltech.edu) and information about how anyone may subscribe to it. The easiest and best way to get involved in SBGN discussions is to join the mailing list and participate.

Face-to-face meetings of the SBGN community are announced on the website as well as the mailing list. Although no set schedule currently exists for workshops and other meetings, we envision holding at least one public workshop per year. As with other similar efforts, the workshops are likely to be held as satellite workshops of larger conferences, enabling attendees to use their international travel time and money more efficiently.

Notifications of updates to the SBGN specification are also broadcast on the mailing list and announced on the SBGN website.

1.4 Note on typographical convention

The concept represented by a glyph is written using a normal font, while a *glyph* means the SBGN visual representation of the concept. For instance "a biological entity is encoded by the SBGN ER *entity*".

doi:10.2390/biecoll-jib-2015-264

Chapter 2

Entity Relationship glyphs

This chapter provides a catalog of the graphical symbols available for representing entities in Entity Relationship maps. There are different classes of glyphs corresponding to different classes of entity nodes, statements and influences.

In Chapter 3 beginning on page 31, we describe the rules for combining these glyphs into legal SBGN Entity Relationships, and in Chapter 4 beginning on page 37, we describe requirements and guidelines for the way that maps are visually organized.

2.1 Entity nodes

Entity nodes (ENs) represent elements of truth, class of things that exist. Entity nodes are the source of *influences* (Section 2.2.3). SBGN Entity Relationship Level 1 provides three different types of *entity nodes*, the *interactors*, the *logical operators* and the *perturbing agent*.

2.1.1 Interactors

Interactors are nodes that are able to participate in an interaction (Section 2.2.1.2). SBGN Entity Relationship Level 1 provides two interactors, the *entity* and the *outcome* of a statement.

2.1.1.1 *Glyph:* Entity

SBGN Entity Relationship Level 1 defines only one glyph for all entities, whether physical entity, such as protein, a nucleic acid, metabolite or functional entity such as a gene. Indeed the exact nature of entities does not impact the rules of interactions within a map. The nature of a particular entity may then be clarified using its label and decorations, as will become clear below.

SBO Term:

SBO:0000245 ! entity

Container:

An *entity* is represented by a rectangular container with rounded corners, as illustrated in Figure 2.1 on the next page.

Label:

An *entity* is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The center of the label box must be located in the container. The label may spill outside of the container.

Auxiliary items:

An *entity* might carry state variables that can add information about its state (Section 2.3.2). A state variable is represented by a "stadium", that is a rectangle capped

doi:10.2390/biecoll-jib-2015-264

with two hemi-circles. The center of the bounding box of *state variables* is located on the border and the long axis of this stadium is aligned with that part of the border of the *entity*'s container, as illustrated in Figure 2.1. The label of the state variable (which can precise the type of characteristic represented by the state variable, residue type, residue number etc.) is written within the state variable's container. Particular *state variables* are the existence (Section 2.3.2) and the location (Section 2.3.2).

An *entity* can carry one or several *units of information* (Section 2.3.1). Particular *units of information* are available for describing the material type (Section 2.4.1) and the conceptual type (Section 2.4.2) of a macromolecule. The center of the bounding box of a *unit of information* is located on the border and the long axis of this stadium is aligned with that part of the border of the *entity*'s container.

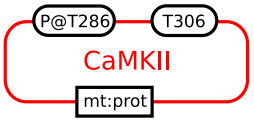

Figure 2.1: *Example of an* entity *named CaMKII, that carries two* state variables *representing the phosphorylated residue threonine 286 and the residue threonine 306, and a* unit of information *precising its material status (protein).*

The granularity of the representation, that is what a given *entity* actually represents, is let to the person generating the map. For instance (Figure 2.2), one could choose to represent phosphorylated and non-phosphorylated forms of a protein using a single *entity*, carrying a *state variable* representing the phosphorylation. The phosphorylated form would be represented by the *outcome* located on the *assignment*. Alternatively one could create separate entity for phosphorylated state, carrying the *existence state variable* and represent phosphorylation by assignment of the true value to the *existence state variable*.

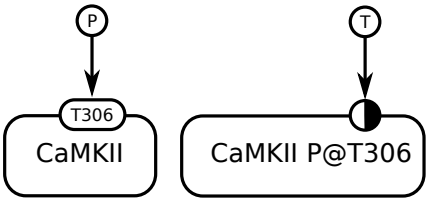

Figure 2.2: *The two ways to represent phosphorylation of the protein: by assigning "P" to the state variable (left), and by setting existence to phosphorylated* entity *(right).*

2.1.1.2 Glyph: Outcome

In Entity Relationships, an *outcome* represents the actualisation of a *statement* (Section 2.2.1). It is not an entity on its own, but a subset of all possible entity instances where *statement* is true. For instance, if an *interaction* represents a non-covalent binding, the *outcome* represents all possible complexes that has two interactors bound to each other. If an *interaction* represents a genetic interaction, for instance derived from genetic screenings, the *outcome* represents the result of the presence of the two polymorphisms. If an *assignment* represents the phosphorylation of a protein, the *outcome* represents all possible complexes and free molecules where this protein is phosphorylated.

An *outcome* represent a particular instance of a realisation, and therefore, from one outcome must depart only one influence. An outcome being an *entity node*, it cannot receive influences. It exists. It cannot more or less exist.

doi:10.2390/biecoll-jib-2015-264

SBO Term:
SBO:0000409 ! interaction outcome

Container:
An *outcome* is represented by a black dot located on the arc of a *statement* (Section 2.2.1). The diameter of the dot has to be larger than the thickness of the arc.

Label:
An *outcome* has no identity on its own and does not carry any label.

Auxiliary items:
An *outcome* does not carry any auxiliary items.

Figure 2.3: *The Entity Relationship glyph for* outcome.

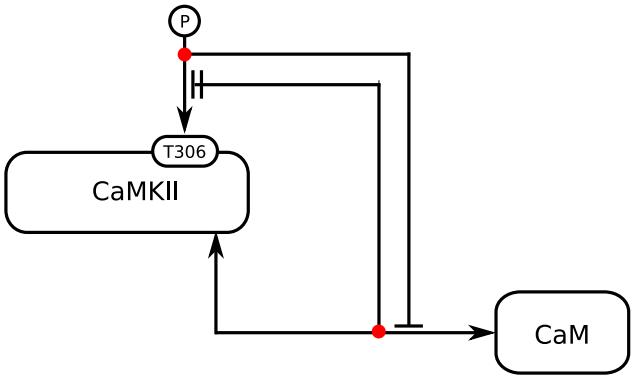

Figure 2.4: *Examples of* outcomes. *The rightmost represents the fact that calmodulin effectively interacts (Section 2.2.1.2) with calcium/calmodulin kinase II. The leftmost represents the fact that the value phosphorylated is assigned (Section 2.2.1.1) to the variable representing threonin 306 of calcium/calmodulin kinase II.*

2.1.2 Logical operators

A *logical operator* allows to combine elements of truth into another element of truth (if A exists and B exits, then A AND B exists) in order to apply influences. SBGN Entity Relationship Level 1 provides three *logical operators*, *and*, *or*, and *not*.

2.1.2.1 Glyph: And

The glyph *and* is used to denote that all the *interactors* linked as input are necessary to produce the output influence.

SBO Term:
SBO:0000173 ! and.

Container:
And is represented by a circle, with two connectors located at the opposite side for inputs and output.

doi:10.2390/biecoll-jib-2015-264

Label:

And is identified by the label "AND" placed in an unbordered box attached to the center of the container.

Auxiliary items:

And does not carry any auxiliary items.

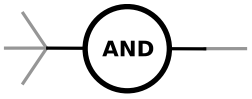

Figure 2.5: *The Entity Relationship glyph for* and. *Three inputs are represented, but two or more than three would be allowed.*

2.1.2.2 Glyph: Or

The glyph *or* is used to denote that any of the *interactors* linked as input is sufficient to produce the output influence.

SBO Term:

SBO:0000174 ! or.

Container:

Or is represented by a circle, with two connectors located at the opposite side for inputs and output.

Label:

Or is identified by the label "OR" placed in an unbordered box attached to the center of the container.

Auxiliary items:

Or does not carry any auxiliary items.

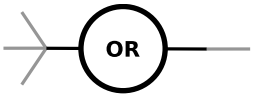

Figure 2.6: *The Entity Relationship glyph for* or. *Three inputs are represented, but two or more than three would be allowed.*

doi:10.2390/biecoll-jib-2015-264

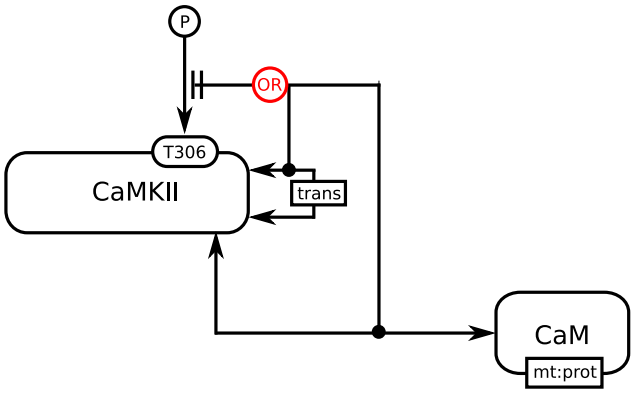

Figure 2.7: *Example of the* or *logical operator, showing that either the dimerisation of CaMKII or its binding to Calmodulin preclude the phosphorylation of threonin 306.*

2.1.2.3 *Glyph:* Not

The glyph *not* is used to denote that the output influence only happens in the absence of the input *interactor*.

SBO Term:
SBO:0000238 ! not.

Container:
Not is represented by a circle, with two connectors located at the opposite side for input and output.

Label:
Not is identified by the label "NOT" placed in an unbordered box attached to the center of the container.

Auxiliary items:
Not does not carry any auxiliary items.

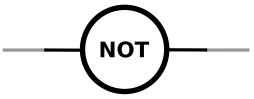

Figure 2.8: *The Entity Relationship glyph for* not.

2.1.2.4 *Glyph:* delay

The glyph *delay* is used to denote that the *entity nodes* linked as input does not produce the influence immediately, but a delay after the decision of influencing has been taken.

SBO Term:
SBO:0000225 ! delay.

Container:
Delay is represented by a circle, with two connectors located at the opposite side for input and output.

Label:
Delay is identified by the greek letter "τ" ("TAU") placed in an unbordered box attached to the center of the container.

doi:10.2390/biecoll-jib-2015-264

Auxiliary items:

Delay does not carry any auxiliary items.

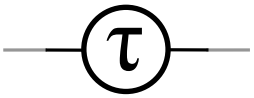

Figure 2.9: *The Entity Relationship glyph for* delay.

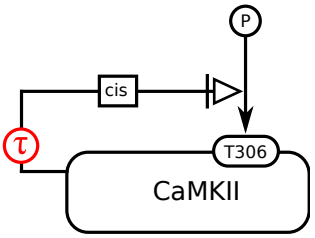

Figure 2.10: *Example of the* delay *logical operator, showing that the stimulation of the phosphorylation of CaMKII on threonin 306 takes place a measurable amount of time after the decision of stimulation is triggered.*

2.1.3 Glyph: *Perturbing agent*

Biochemical networks can be affected by external influences. Those influences can be well-defined physical perturbations, such as a the effect of a light pulse or of a change in temperature; they can also be more complex and not well-defined phenomena, for instance a biological process, an experimental setup, or a mutation. For these situations, SBGN provides the *perturbing agent* glyph. We do not use the word *perturbation* to avoid the misunderstanding with the influence that the *perturbing agent* has on the map.

SBO Term:

SBO:0000405 ! perturbing agent

Container:

A *perturbing agent* is represented by a modified hexagon having two opposite concave faces, as illustrated in Figure 2.11 on the following page.

Label:

A *perturbing agent* is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the *perturbing agent* container. The label may spill outside of the container.

Auxiliary items:

A *perturbing agent* does not carry any auxiliary unit. In particular, its existence being not subjected to any modulation by any other *interactor*, it does not require the state variable existence. *Perturbing agent* do not have location either. pH of lysosome and mitochondria are different perturbing agents.

doi:10.2390/biecoll-jib-2015-264

Figure 2.11: *The Entity Relationship glyph for* perturbing agent.

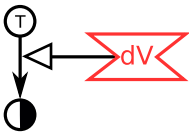

Figure 2.12: *Example of a* perturbing agent *representing the depolarisation of a membrane, that stimulates (Section 2.2.3.2) the existence (see 2.3.2) of an interactor.*

2.2 Relationships

Relationships are rules that decide of the existence of entity nodes, based on the existence of others. SBGN Entity Relationship Level 1 provides two types of relationships, the statements and the influences.

2.2.1 Statements

Statements can be true or false. *Statements* are targets of *Influences*. They are not true themselves, but can carry *Outcomes* (see Section 2.1.1.2). SBGN Entity Relationship Level 1 provides three types of statements, *Assignment*, *Interaction* and *Phenotype*.

2.2.1.1 *Glyph:* Assignment

Assignment is used to describe the setting of a state variable to a certain value. The assignment, represented by an harpoon arrow, goes from one or more *variable values* to a variable identification, represented by a *state variable* attached to the entity affected by the assignment. If an *assignment* takes several *variable values* as input, there is an implicit XOR between them, located at the point of junction between the arcs coming from the alternative values. Since only one value can be assigned at a time, there is therefore no edge overlap in the assignment itself. The result of an assignment is represented by *outcomes*, that is by filled dots on the arrow. The result of an *assignment* can be represented by any number of *outcomes*. In the case of more than one *variable values*, the *outcomes* must be placed on the relevant incoming branch.

SBO Term:

SBO:0000464 ! state variable assignment

Origin:

One or more variable value (section Section 2.3.3) on their own, each containing a variable value.

Target:

A state-variable (section Section 2.3.2) carried by an entity (section Section 2.1.1.1), containing a variable identification.

Symbol:

The target extremity of an *assignment* carries an harpoon arrowhead.

doi:10.2390/biecoll-jib-2015-264

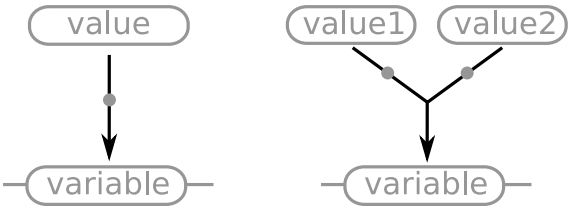

Figure 2.13: *The Entity Relationship glyph for* assignment.

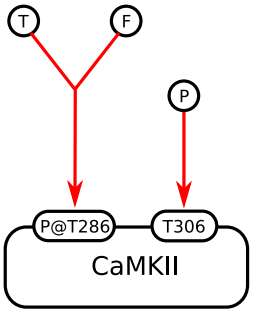

Figure 2.14: *Two examples of* assignment *representing phosphorylation, by one value (phosphorylation) of a variable representing a residue, or two values (true or false) of a variable representing the phosphorylated residue.*

2.2.1.2 *Glyph:* Interaction

Interaction represents an interaction between two or more *entities* or *outcomes*, whether a non-covalent physical interaction, or a functional interaction, e.g. genetic interaction. The interaction is represented by a circle which connects to arrows pointing to the interactors involved in the interaction. In the case of a binary interaction, the circle may be ommitted. The realisation of the interaction is represented by *outcomes* (see section 2.1.1.2), that is by filled dots. These *outcomes* are located on the circle representing the interaction. In the case of a binary interaction represented without circle, the *outcomes* can be placed anywhere between the arrowheads. The realisations of an interaction can be represented by any number of *outcomes*. The *influences* (2.2.3) targeting an interaction end up on the external side of the circle, between the *outcomes*.

SBO Term:

SBO:0000342 molecular or genetic interaction

Origin:

Any *interactor* (Section 2.1.1).

Target:

Any *interactor* (Section 2.1.1).

Symbol:

Both origin and target extremities of an *interaction* carry an harpoon arrowhead. The arrows pointing to the *interactors* originate from a circle. In the case of a binary interaction, the circle is optional.

Auxiliary items:

A *unit of information* containing a *cardinality* (Section 2.4.4) indicates the number of instances of an entity involved in an interaction. The absence of a *cardinality* is synonymous of a cardinality of 1. A *unit of information* on a binary interaction involving only

doi:10.2390/biecoll-jib-2015-264

one entity, or outcomes of interactions involving one entity, carrying the mention *cis* or *trans* precises if the interaction is intra-molecular or between different instances of the same entity or complex.

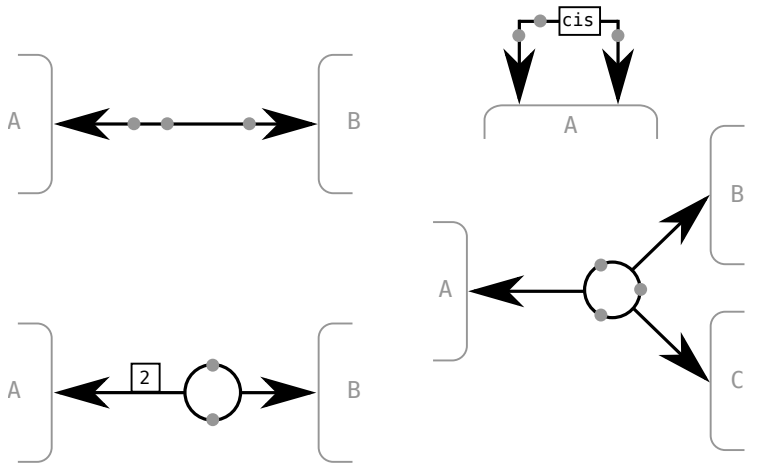

Figure 2.15: *The Entity Relationship glyph for* interaction. *Top left, binary interaction between two entities. The circle can be ommitted, and the* outcomes *located anywhere on the* interaction. *Bottom left, because the cardinality of the entity A is 2, the interaction is not a binary one, but a n-ary one. The circle cannot be ommitted. Bottom right, n-ary interaction with three different entities. Top right, intra-molecular interaction.*

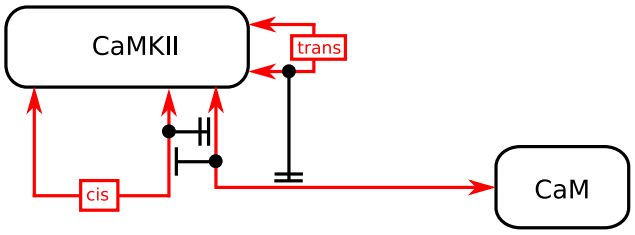

Figure 2.16: *Examples of* interactions, *showing the effect of the binding of calmodulin to CaMKII (binary interaction) on the folding of the kinase (intra-molecular interaction), and the effect of the folding or the dimerisation of CaMKII (inter-molecular interaction between different instances of CaMKII) on the binding of calmodulin.*

doi:10.2390/biecoll-jib-2015-264

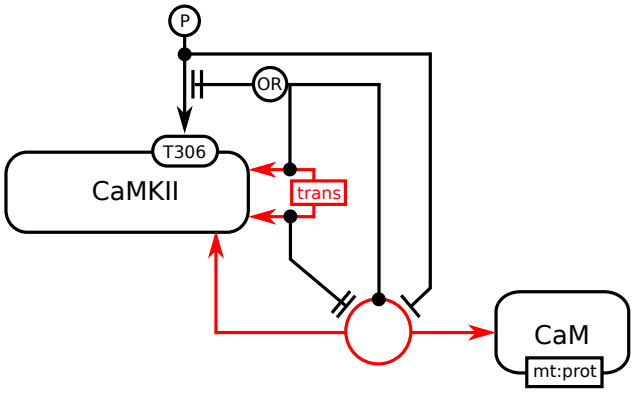

Figure 2.17: *Examples of a binary* interaction *between CaMKII and calmodulin where the interaction is represented by a circle. Interaction between adjacent monomers of CaMKII (trans-interaction) preclude the binding of calmodulin, as represented by an* absolute inhibition *ending on the circle. The phosphorylation of threonine 306 also inhibits the interaction. The realisation of the interaction, represented by an* outcome *located on the circle, itself inhibits the phosphorylation of threonine 306.*

2.2.2 Glyph: *Phenotype*

A biochemical network can generate phenotypes or affect biological processes. Such processes can take place at different levels and are independent of the biochemical network itself. To represent these processes in a map, SBGN Entity Relationship Level 1 defines the *phenotype* glyph.

SBO Term:

SBO:0000358 ! phenotype

Origin:

Non-applicable

Target:

Non-applicable

Symbol:

A *phenotype* is represented by an elongated hexagon, as illustrated in Figure 2.18. It is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the *phenotype* container. The label may spill outside of the container.

Figure 2.18: *The Entity Relationship glyph for* phenotype.

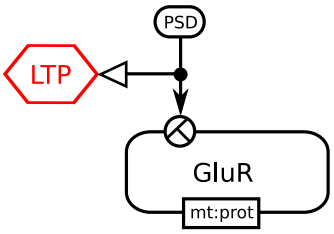

Figure 2.19: *Example of a* phenotype *"Long Term Potentiation (LTP)" enhanced when the entity "GluR" is present in the post-synaptic density.*

2.2.3 Influences

Influence arcs represent the effect of an entity on another relationship. The symbols attached to their extremities precise their semantics. SBGN Entity Relationships' influences can be viewed as logical rules linking *ENs* and other rules. SBGN Entity Relationship Level 1 provides seven influences, *Modulation*, *Stimulation*, *Inhibition*, *Necessary Stimulation*, *Absolute Inhibition*, *Absolute Stimulation*, *Logic Arc*.

2.2.3.1 *Glyph:* Modulation

A *modulation* affects the propensity, or the chances to happen, of the target relationship. Such a modulation can affect the relationship **positively or negatively**, or even both ways depending on the conditions. A *modulation* can also be used when one does not know the precise direction of the effect, for instance if there are conflicting evidences.

SBO Term:
SBO:0000168 ! control.

Origin:
Any *entity node* (Section 2.1).

Target:
Any *relationship* (Section 2.2).

Symbol:
The target extremity of a *modulation* carries an empty diamond.

Auxiliary items:
A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *modulation* origins and either:

- the *entity node* from which the influence targeted by the *modulation* origins
- all the relevant *interactors* of the *interaction* targeted by the *modulation*
- the *entity* subjected to the *assignment* targeted by the *modulation*

Figure 2.20: *The Entity Relationship glyph for* modulation.

doi:10.2390/biecoll-jib-2015-264

2.2.3.2 *Glyph:* Stimulation

A *stimulation* affects **positively** the strength, or the probability, of the target relationship. This stimulation can be for instance a catalysis or a positive allosteric regulation.

SBO Term:
SBO:0000170 ! stimulation.

Origin:
Any *entity node* (Section 2.1).

Target:
Any *relationship* (Section 2.2).

Symbol:
The target extremity of a *stimulation* carries an empty arrowhead.

Auxiliary items:
A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *stimulation* origins and either:

- the *entity node* from which the influence targeted by the *stimulation* origins
- all the relevant *interactors* of the *interaction* targeted by the *stimulation*
- the *entity* subjected to the *assignment* targeted by the *stimulation*

Figure 2.21: *The Entity Relationship glyph for* stimulation.

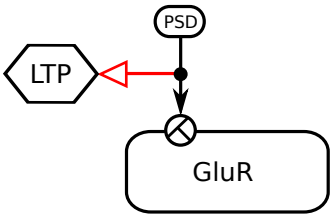

Figure 2.22: *Example of a* stimulation *a* phenotype *"Long Term Potentiation (LTP)" enhanced when the entity "GluR" is present in the post-synaptic density.*

2.2.3.3 *Glyph:* Inhibition

An *inhibition* affects **negatively** the strength, or the probability, of the target relationship. This inhibition can be for instance a steric hindrance or a negative allosteric regulation.

SBO Term:
SBO:0000169 ! inhibition.

Origin:
Any *entity node* (Section 2.1).

doi:10.2390/biecoll-jib-2015-264

Target:
Any *relationship* (Section 2.2).

Symbol:
The target extremity of a *inhibition* carries a bar perpendicular to the arc.

Auxiliary items:
A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *inhibition* origins and either:

- the *entity node* from which the influence targeted by the *inhibition* origins
- all the relevant *interactors* of the *interaction* targeted by the *inhibition*
- the *entity* subjected to the *assignment* targeted by the *inhibition*

Figure 2.23: *The Entity Relationship glyph for* inhibition.

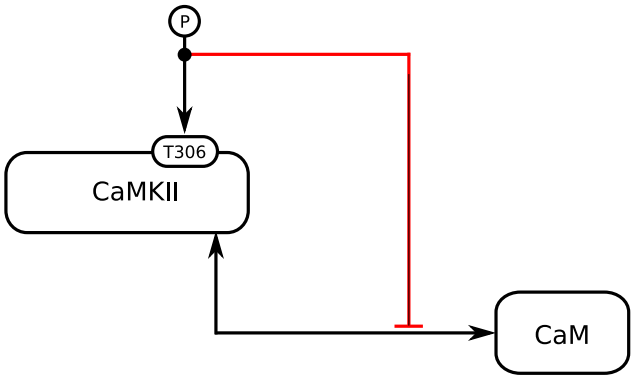

Figure 2.24: *In this example, the phosphorylation of the threonine 306 of the regulatory domain of CaMKII inhibits the interaction between Calmodulin and the kinase.*

2.2.3.4 *Glyph:* Necessary stimulation

A *necessary stimulation* is an influence that has to be present for a relationship to take place (to become true). A relationship modulated by a necessary stimulation can only exist when this stimulation is true, whatever are the other influences this relationship is subjected to.

SBO Term:
SBO:0000171 ! necessary stimulation.

Origin:
Any *entity node* (Section 2.1).

Target:
Any *relationship* (Section 2.2).

doi:10.2390/biecoll-jib-2015-264

Symbol:

The target extremity of a *necessary stimulation* carries an open arrow (to remind that it is a *stimulation*) coming after a larger vertical bar.

Auxiliary items:

A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *necessary stimulation* origins and either:

- the *entity node* from which the influence targeted by the *necessary stimulation* origins
- all the relevant *interactors* of the *interaction* targeted by the *necessary stimulation*
- the *entity* subjected to the *assignment* targeted by the *necessary stimulation*

Figure 2.25: *The Entity Relationship glyph for* necessary stimulation.

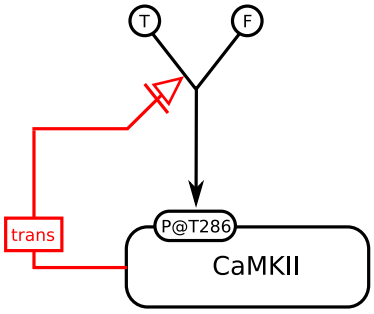

Figure 2.26: *This example shows how threonine 286 of CaMKII is only phosphorylated by the kinase itself, but in a trans-fashion, meaning a molecule of CaMKII does not phosphorylate itself, but another molecule of CaMKII.*

2.2.3.5 Glyph: Absolute inhibition

An *absolute inhibition* precludes the existence of another relationship. A relationship modulated by an absolute inhibition can only exist when an absolute inhibition is false, whatever are the other influences this relationship is subjected to.

SBO Term:

SBO:0000407 ! absolute inhibition.

Origin:

Any *entity node* (Section 2.1).

Target:

Any *relationship* (Section 2.2).

Symbol:

The target extremity of a *absolute inhibition* carries a double bar perpendicular to the arc (to remind that it is an *inhibition*).

doi:10.2390/biecoll-jib-2015-264

Auxiliary items:

A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *absolute inhibition* origins and either:

- the *entity node* from which the influence targeted by the *absolute inhibition* origins
- all the relevant *interactors* of the *interaction* targeted by the *absolute inhibition*
- the *entity* subjected to the *assignment* targeted by the *absolute inhibition*

Figure 2.27: *The Entity Relationship glyph for* absolute inhibition.

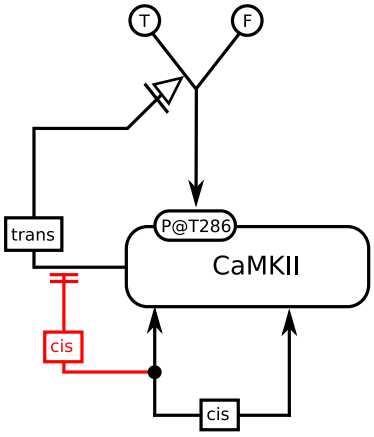

Figure 2.28: *This example shows how an intra-molecular interaction of CaMKII precludes totally its catalytic activity upon another molecule of CaMKII.*

2.2.3.6 *Glyph:* Absolute stimulation

An absolute stimulation always triggers the existence of a target relationship.

SBO Term:

SBO:0000411 ! absolute stimulation

Origin:

Any *entity node* (Section 2.1).

Target:

Any *relationship* (Section 2.2).

Symbol:

The target extremity of a *absolute stimulation* carries a double empty arrowhead (to remind that it is a *stimulation*).

Auxiliary items:

A *unit of information* carrying the mention *cis* or *trans* precises the relationship between the *entity node* from which the *absolute stimulation* origins and either:

- the *entity node* from which the influence targeted by the *absolute stimulation* origins
- all the relevant *interactors* of the *interaction* targeted by the *absolute stimulation*
- the *entity* subjected to the *assignment* targeted by the *absolute stimulation*

Figure 2.29: *The Entity Relationship glyph for* absolute stimulation.

2.2.4 Glyph: *Logic arc*

Logic arc is used to represent the fact that an interactor influences the outcome of a logic operator.

SBO Term:
SBO:0000398 ! logical relationship.

Origin:
Any *interactor* (Section 2.1.1) or *logical operator* (Section 2.1.2).

Target:
Any *logical operator* (Section 2.1.2).

Symbol:
No particular symbol is used to represent a logic arc.

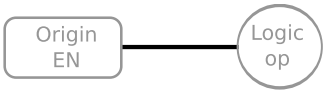

Figure 2.30: *The Entity Relationship glyph for* logic arc.

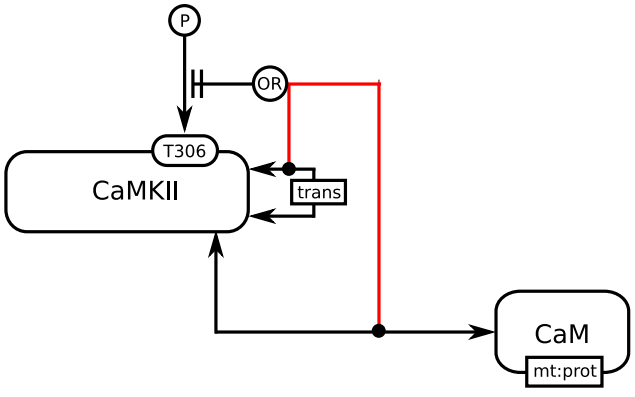

Figure 2.31: *In this example, two logic arcs reflect the fact that the phosphorylation of threonine 306 on CaMKII is precluded either by a dimerisation or the binding of calmodulin.*

doi:10.2390/biecoll-jib-2015-264

2.3 Auxiliary units

Auxiliary units are decorations used on *entities* (Section 2.1.1.1), *interactions* (Section 2.2.1.2) and *influences* (Section 2.2.3) to further refine their semantics. SBGN Entity Relationship Level 1 provides two *auxiliary units*, the *unit of information* and the *state variable*.

2.3.1 Glyph: *Unit of information*

When representing biological entities, it is often necessary to convey some abstract information about the entity's function or structure. The SBGN *unit of information* is a decoration that can be used in this situation to add information to a glyph. Some example uses of a *unit of information* include (but are not limited to) specifying if an interaction is intra or intermolecular, information about the physical environment, or the specific type of biological entity it is decorating.

SBO Term:

Not applicable.

Container:

A unit of information is represented by a rectangle. The long side of the rectangle should be oriented parallel to the border of the *entity*, or the edge, being annotated by the *unit of information*. The center of the bounding box of a *unit of information* should be located on the mid-line of the border of the carrying *entity* or the carrying edge.

Label:

A *unit of information* is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the container. The label may spill outside of the container.

The label defines the information carried by the *unit of information*. SBGN Entity Relationship Level 1 defines several reserved labels, such as "cis" and "trans", or specific prefixes that must be included in the label to indicate the type of information (Section 2.4).

Auxiliary items:

A *unit of information* does not carry any auxiliary items.

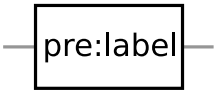

Figure 2.32: *The Entity Relationship glyph for* unit of information.

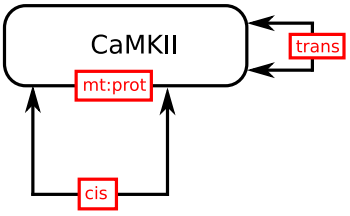

Figure 2.33: *Using* units of information *to represent the fact that the entity "CaMKII" is a protein, and to display intra- (cis) and inter- (trans) molecular interactions.*

doi:10.2390/biecoll-jib-2015-264

2.3.2 Glyph: *State variable*

Many biological entities such as molecules can exist in different *states*, meaning different physical or informational configurations. These states can arise for a variety of reasons. For example, macromolecules can be subject to post-synthesis modifications, wherein residues of the macromolecules (amino acids, nucleosides, or glucid residues) are modified through covalent linkage to other chemicals. Other examples of states are alternative conformations as in the closed/open/desensitized conformations of a transmembrane channel, and the active/inactive forms of an enzyme.

SBGN provides a means of associating one or more *state variables* with an entity; each such variable can be used to represent a dimension along which the state of the overall entity can vary. When an entity can exist in different states, the state of the instance of the entity can be described by the current values of all its *state variables*, and the values of the *state variables* of all its possible components (nested entities Section 2.6), recursively.

In SBGN Entity Relationship Level 1, *state variables* are also used to describe the localisation in compartments (a transport is therefore could be described as assignment of different values to the location state variable, see Section 2.2.1.1).

SBO Term:

Not applicable.

Container:

A *state variable* is represented by a "stadium" container, that is two semicircles of same radius joined by parallel segments, as shown in Figure 2.34. The parallel segment axis should be tangent to the border of the glyph of the *entity* being modified by the *state variable*. The center of the bounding box of a *state variable* should be located on the mid-line of the border of the *entity*.

Label:

A *state variable* is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the container. The label may spill outside of the container.

Auxiliary items:

A *state variable* does not carry any auxiliary items.

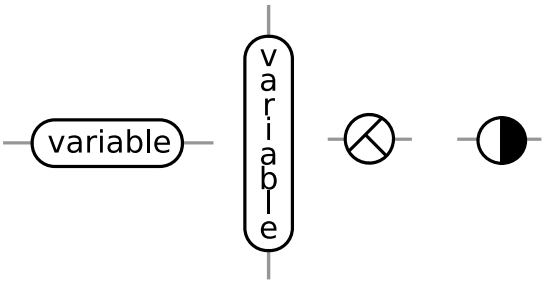

Figure 2.34: *The Entity Relationship glyph for* state variable. *From left to right, horizontal* state variable, *vertical* state variable, location, existence.

Two state variables are predefined. The variable *existence* is used to represent the creation or destruction of instances of an entity, as seen on Figure 2.35 on the next page . *Existence* can take two values, true (T) or false (F). The variable is represented by a circle vertically divided in two. One hemicircle is black, and the other white.

doi:10.2390/biecoll-jib-2015-264

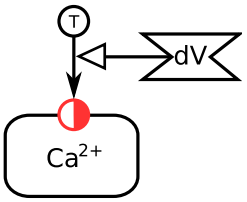

Figure 2.35: *Using the* state variable existence *to represent the appearance of calcium following a depolarisation.*

The variable *location* is used to represent the physical location of an entity, as seen on Figure 2.36 . *Location* can take any value, but there can be only one *location* per instance of an entity. The variable is represented by a circle containing two perpendicular segments, an abstract version of the usual slanted pin.

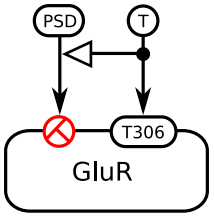

Figure 2.36: *Using the* state variable location *to represent the fact that phosphorylation of glutamate receptors stimulate their incorporation in the post-synaptic density.*

2.3.3 Glyph: *Variable value*

SBO Term:

Not applicable.

Container:

A *variable value* is represented by a "stadium" container, that is two hemicercles of same radius joined by parallel segments, as shown in Figure 2.37.

Label:

A *variable value* is identified by a label placed in an unbordered box containing a string of characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the container. The label may spill outside of the container.

Auxiliary items:

A *variable value* does not carry any auxiliary items.

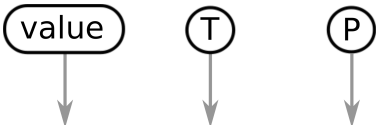

Figure 2.37: *The Entity Relationship glyph for* variable value.

A *variable value* is linked to a *state variable* (see Section 2.3.2) through *assignment* (see Section 2.2.1.1).

doi:10.2390/biecoll-jib-2015-264

A *state variable* does not necessarily have to be Boolean-valued. For example, an ion channel can possess several conductance states; a receptor can be inactive, active and desensitized; and so on. As another example, a *state variable* "ubiquitin" could also carry numerical values corresponding to the number of ubiquitin molecules present in the tail. The set of *variable value* linked to a *state variable* does not necessary have to enumerate all possible values for that state variable, for example 'P' value assigned to post translation modification site does not assume that this site could not be also methylated or glycosylated at that site.

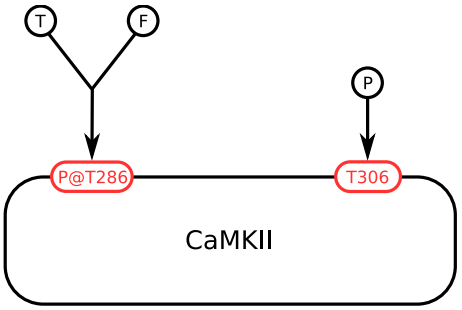

Figure 2.38: *Two examples of* state variables *used to represent phophorylation of a threonine residue. While only the value "phosphorylated" is assigned to T306, the variable T286P can take the values true or false, which allow for representing dephosphorylation as well as phosphorylation.*

2.4 Controlled vocabularies used in SBGN Entity Relationship Level 1

Some glyphs in SBGN Entity Relationships can contain particular kinds of textual annotations conveying information relevant to the purpose of the glyph. These annotations are carried by *units of information* (Section 2.3.1) or *state variables* (Section 2.3.2).

The text that appears as the unit of information decorating an entity must be prefixed with a controlled vocabulary term indicating the type of information being expressed. The prefixes are mandatory. Without the use of controlled vocabulary prefixes, it would be necessary to have different glyphs to indicate different classes of information; this would lead to an explosion in the number of symbols needed. There is no more than one *unit of information* per controlled vocabulary allowed for any *Entity*. So, for example, one *unit of information* of material type and one *unit of information* of conceptual type.

In the rest of this section, we describe the controlled vocabularies (CVs) used in SBGN Entity Relationship Level 1. In each case, some CV terms are predefined by SBGN, but unless otherwise noted, *they are not the only terms permitted*. Authors may use other CV values not listed here, but in such cases, they should explain the terms' meanings in a figure legend or other text accompanying the map.

2.4.1 Entity material types

The material type of an *Entity* indicates its chemical structure. A list of common material types is shown in Table 2.1 on the following page, but others are possible. The values are to be taken from the Systems Biology Ontology (`http://www.ebi.ac.uk/sbo/`), specifically from the branch having identifier `SBO:0000240` (*material entity*). The labels are defined by SBGN Entity Relationship Level 1.

The material types are in contrast to the *conceptual types* (see below). The distinction is that material types are about physical composition, while conceptual types are about functions. For example, a strand of RNA is a physical artifact, but its use as messenger RNA is a function.

doi:10.2390/biecoll-jib-2015-264

Name	Label	SBO term
Non-macromolecular ion	`mt:ion`	SBO:0000327
Non-macromolecular radical	`mt:rad`	SBO:0000328
Ribonucleic acid	`mt:rna`	SBO:0000250
Deoxribonucleic acid	`mt:dna`	SBO:0000251
Protein	`mt:prot`	SBO:0000297
Polysaccharide	`mt:psac`	SBO:0000249

Table 2.1: *A sample of values from the* material types *controlled vocabulary (Section 2.4.1).*

2.4.2 Entity conceptual types

An *entity*'s *conceptual type* indicates its function within the context of a given Entity Relationship map. A list of common conceptual types is shown in Table 2.2, but others are possible. The values are to be taken from the Systems Biology Ontology (`http://www.ebi.ac.uk/sbo/`), specifically from the branch having identifier `SBO:0000241` (*functional entity*). The labels are defined by SBGN Entity Relationship Level 1.

Name	Label	SBO term
Gene	`ct:gene`	SBO:0000243
Transcription start site	`ct:tss`	SBO:0000329
Gene coding region	`ct:coding`	SBO:0000335
Gene regulatory region	`ct:grr`	SBO:0000369
Messenger RNA	`ct:mRNA`	SBO:0000278
Functional domain	`ct:domain`	SBO:0000493
Binding site	`ct:bind`	SBO:0000494
Catalytic site	`ct:cat`	SBO:0000495
Transmembrane domain	`ct:tm`	SBO:0000496

Table 2.2: *A sample of values from the* conceptual types *vocabulary (Section 2.4.2).*

2.4.3 Macromolecule covalent modifications

A common reason for the introduction of state variables on an entity is to allow access to the configuration of possible covalent modification sites on that entity. For instance, a macromolecule may have one or more sites where a phosphate group may be attached; this change in the site's configuration (i.e., being either phosphorylated or not) may factor into whether, and how, the entity can participate in different processes. Being able to describe such modifications in a consistent fashion is the motivation for the existence of SBGN's covalent modifications controlled vocabulary.

Table 2.3 on the following page lists a number of common types of covalent modifications. The most common values are defined by the Systems Biology Ontology in the branch having identifier `SBO:0000210` (*addition* under *events* → *reaction* → *biochemical reaction* → *conversion* → *addition*). The labels shown in Table 2.3 on the next page are defined by SBGN Entity Relationship Level 1; for all other kinds of modifications not listed here, the author of an Entity Relationship map must create a new label (and should also describe the meaning of the label in a legend or text accompanying the map).

Name	Label	SBO term
Acetylation	Ac	SBO:0000215
Glycosylation	G	SBO:0000217
Hydroxylation	OH	SBO:0000233
Methylation	Me	SBO:0000214
Myristoylation	My	SBO:0000219
Palmytoylation	Pa	SBO:0000218
Phosphorylation	P	SBO:0000216
Prenylation	Pr	SBO:0000221
Protonation	H	SBO:0000212
Sulfation	S	SBO:0000220
Ubiquitination	Ub	SBO:0000224

Table 2.3: *A sample of values from the* covalent modifications *vocabulary (Section 2.4.3).*

2.4.4 Miscellaneous terms

SBGN Entity Relationship Level 1 requires several reserved characters. A special unit of information usable on interactions describe the number of identical interactors involved. Note that the value is a unitary number, and not (for example) a range. There is no provision in SBGN Entity Relationship Level 1 for specifying a range in this context because it leads to problems of entity identifiability. Other reserved characters (see Table 2.4) are used in state variable assignments to represent truth or falsehood (T and F). Two reserved words are used in units of information carried by relationships: *cis* and *trans*.

Name	Label	SBO term
cardinality	#	SBO:0000364
true	T	SBO:0000416
false	F	SBO:0000417
cis	cis	SBO:0000414
trans	trans	SBO:0000415

Table 2.4: *Miscellaneous controlled terms (Section 2.4.4). For the cardinality, # stands for a number, for example, "5".*

2.5 Glyph: *Annotation*

SBGN Entity Relationship Level 1 defines a glyph to add additional information to a map, that does not modify the semantic of the graph. This glyph can be used to add free text, or links to external information.

SBO Term:

SBO:0000550 ! annotation.

Container:

An *annotation* is represented by a rectangular container with a folded corner, as illustrated in Figure 2.39 on the next page. This container is linked to the annotated element with a callout. The link ends up on the border of the annotated element.

Label:

An *annotation* contains information placed in an unbordered box containing a string of

doi:10.2390/biecoll-jib-2015-264

characters. The characters can be distributed on several lines to improve readability, although this is not mandatory. The label box must be attached to the center of the container. The label may spill outside of the container.

Auxiliary items:

An *annotation* does not carry any auxiliary unit.

Figure 2.39: *The Entity Relationship glyph for* annotation.

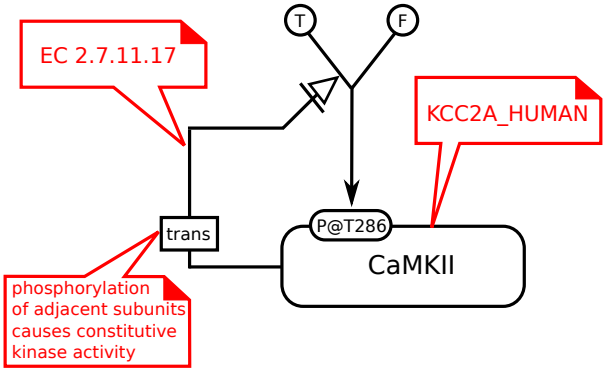

Figure 2.40: *Example of* annotations *adding information to the description of the trans-phosphorylation of CaMKII.*

2.6 *Entity* nesting

SBGN Entity Relationship Level 1 Version 2 allows *entities* to be nested. The enclosed *entity* represents a part of the enclosing *entity*, note that this relation between *entities* of different levels does not imply anything about *entities* of the same level. In the Figure 2.41, A is part of X and B is part of X. However, nothing is said about A and B. A and B could be of different nature (transmembrane domain and transduction domain) or could be overlapping (binding domain and catalytic domain sharing some amino-acids).

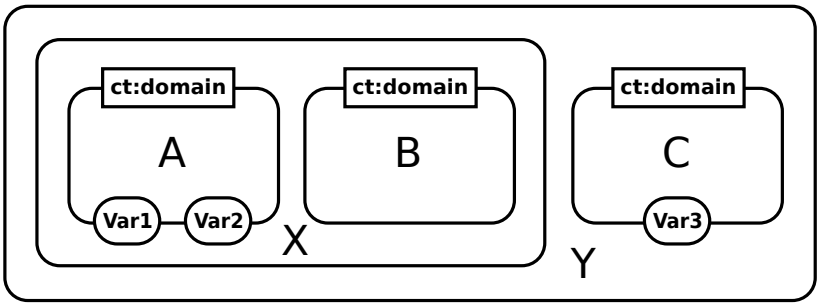

Figure 2.41: *The Entity Relationship glyph for* domain.

doi:10.2390/biecoll-jib-2015-264

As any other *entity*, a nested *entity* can carry *state variables* and *units of information*. The contour of the containing *entity* must surround the totality of all contained *entities* including their *state variables* and *units of information*. This does not include the *assignments* and the *variable values*. A nested *entity* can participate in *interactions*, with "sibling" *entities* (part of the same containing *entity*), or others. They can also generate *influences*. For more information about the use and interpretation of entity nesting, see Section 3.4.6

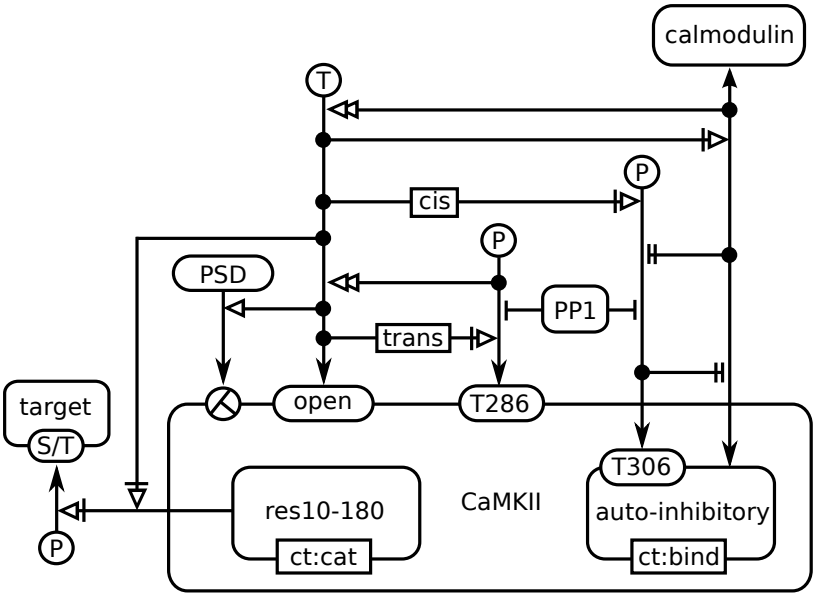

Figure 2.42: *Illustration of the use of* domains. *This map contains five* entities. *Three are top-level, while the* entities *"res10-180" and "auto-inhibitory" are domains of the* entity *"CaMKII". The controlled terms carried by the* units of information *tell us that those* domains *are a catalytic domain and a binding site respectively. As far as the relationships with the* entities *"calmoduline" and "target" are concerned, the interpretation does not differ from the one we would derive from a map where all the* entities *would be entirely separated. However the regulation of localisation of "CaMKII" to the post-synaptic densite (PSD) by phosphorylation of threonine 286 also causes the localisation of "res10-180" and "auto-inhibitory", even if none carry the* state variables *involved.*

2.7 Synchronous events

SBGN Entity Relationship Level 1 Version 2 allows the description of multiple consequences of the event that happens simultaneously. There are cases in which one interaction should influence several others in synchrony, for example the chemical reaction:

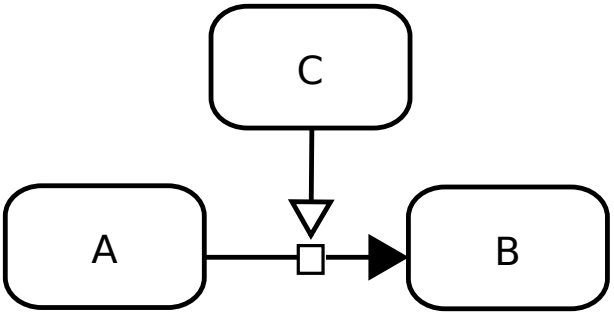

Figure 2.43: *The reaction to represent in Entity Relationship.*

Having two outcomes and two stimulation links (Figure 2.44) does not work in this case because all statement on Entity Relationship map are independent. So it is possible that *entity B* is created, but *entity A* is not deleted.

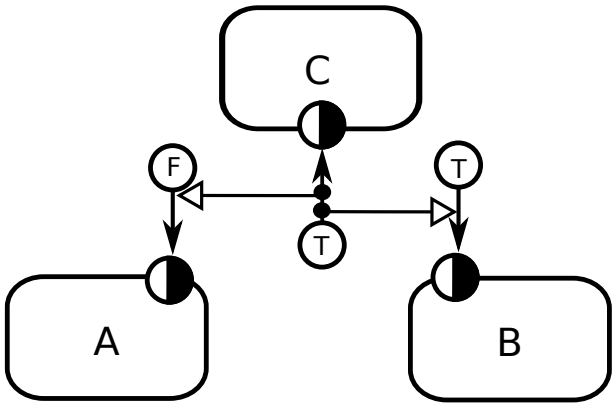

Figure 2.44: *Independent events in Entity Relationship can not describe Figure 2.43 on the previous page.*

To represent multiple consequences of the statement the influence arc is splitted to represent multiple consequences of the event as shown in Figure 2.45.

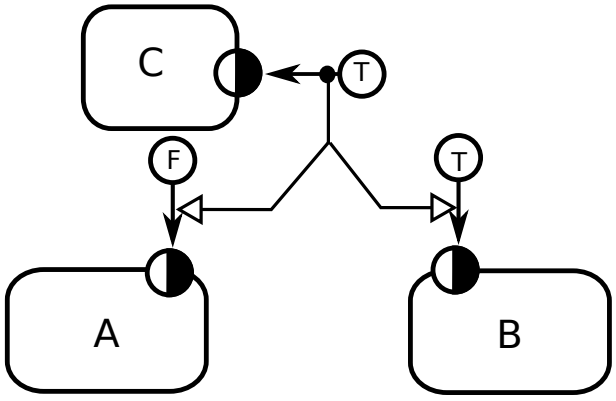

Figure 2.45: *The reaction from Figure 2.43 on the preceding page in Entity Relationship.*

The diagram in the Figure 2.45 should be read as "If (or when) C exists, it stimulates degradation of A and creation of B at the same time".

Synchronous influences do not need to be of the same type and number of branches is not limited. A more general form of the proposed synchronous influence is presented in Figure 2.46 on the following page.

doi:10.2390/biecoll-jib-2015-264

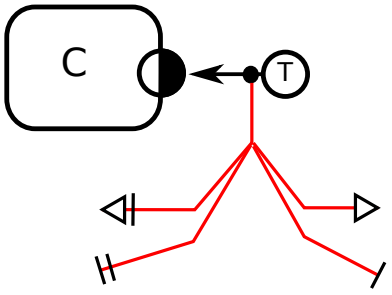

Figure 2.46: *Multiple synchronous consequences of the event.*

Another example (Figure 2.47) is activation of small GTPase by Guanine nucleotide exchange factor (GEF).

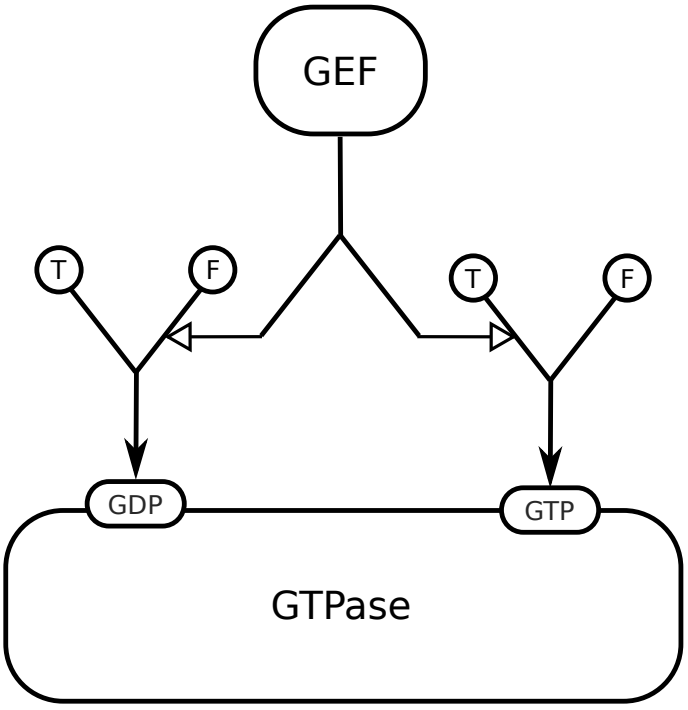

Figure 2.47: *GTPase activation by GEF.*

doi:10.2390/biecoll-jib-2015-264

Chapter 3

Grammar of Entity Relationships

3.1 Overview

In this chapter, we describe how the glyphs of SBGN Entity Relationship Level 1 can be combined to make a valid Entity Relationship map. To do this, we must at the very least define what glyphs can be connected to each other. This is called syntax. Next, we must define additional rules, such as whether duplicate symbols are permitted. In addition, we must define what the notation "means" — how does it represent a body of biological knowledge? This is semantics, and it is essential if a reader is to understand an SBGN map without external help, and a writer is to create a map that reflects his understanding of a biological system.

In this section we start off by describing the concepts of the Entity Relationship language. A detailed description of the syntax is provided next, followed by a description of the syntactic rules of the language.

3.2 Concepts

The SBGN Entity Relationship language is more than a collection of symbols. It is a visual language that uses specific abstractions to describe the biological processes that make up, e.g., a quantitative model, a signalling pathway or a metabolic network. This abstraction is the semantics of SBGN, and to describe it requires more than a definition of the symbols and syntax of the language. We first need to define the abstractions we are using.

SBGN Entity Relationships describe biological interactions involving biological entities. An *entity node* (Section 2.1), such as a molecule, influences the behaviour of other *entities* via a relationship.

It may be convenient to think of a SBGN Entity Relationships as listing independent rules that describe influences between *entities*. A map can then be analysed with "what if?" queries.

3.3 Syntax

The syntax of SBGN Entity Relationships can be defined in the form of an incidence matrix. This incidence matrix has symbols as rows and arcs as columns. Each element of the matrix represents the role of a symbol in connection to an arc. Input (I) means that the arc can begin on that symbol. Output (O) indicates that the arc can end on that symbol. Numbers in parenthesis represent the maximum number of arcs of a particular type to have this specific connection role with the node. No numbers means any number is allowed. Empty cells means the arc is not able to connect to the symbol.

doi:10.2390/biecoll-jib-2015-264

3.3.1 Interactor Nodes connectivity definition

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

3.3.2 Syntactic rules

In addition to the incidence matrix, additional rules refine the syntax of Entity Relationships.

1. Name of the *state variable* should be unique within the *entity* definition.
2. From an *outcome* can only originate one relationship, whether influence or interaction. The relationships being seen as independent rules, separate consequences of an assignment or an interaction have to originate from different outcomes, that is assertion of truth of this assignment or interaction.
3. In the case of a non-binary interaction, the “cis” or “trans” *unit of information* must be carried by the circle representing the n-ary interaction, and not the arc connecting this circle and a given interactor.
4. If an *influence* targeting an *interaction* carries a “cis” or “trans” unit of information, at least one of the *interactors* must be the same *entity* as the origin of the influence.
5. If more than one instance of an *entity* is involved in an *interaction*, a *unit of information cardinality* (Section 2.4.4) must be associated with each entity involved in the statement.
6. A *cis* or *trans* unit of information can only be carried by a relationship involving an instances of a single *entity* directly or by traceable set of *outcomes*.

doi:10.2390/biecoll-jib-2015-264

3.4 Semantic description of Entity Relationships

3.4.1 Entities and states

An *entity* (Section 2.1.1.1) represents a set of individual instances of some kind (molecules, agents etc). The difference from SBGN Process Description Level 1 *entity pool node* is that *entity* instances are independent, while *entity pool node* form one inseparable object acting as a whole. A set of *state variables* (Section 2.3.2) and their *values* (Section 2.3.3) describe a part of the *entity* instances state space. The *entity*, or interaction *outcome* (Section 2.1.1.2) acquire the true state if the set of entity instances compatible with the *entity* state, or actualisation of interaction is non-empty.

3.4.2 Statements

An *interaction* (Section 2.2.1.2) linking the *interactors* A and B means: "any instance of A can interact with instances of B". An *outcome* on an *interaction* represents the cases when the statement is true, that is subset of all instances of A and B between which the interaction effectively exists. If the interaction is a physical interaction between molecules, the *outcome* represents all possible complexes where interactors are really bound to each other, and in that case interaction is always considered as reversible. It is used as follow: "when (or if) A interacts with B then ...".

An *assignment* (Section 2.2.1.1) linking a *variable value* v to a *state variable* V of an *entity* E means: "v is assigned to V of E" or "V of E takes the value v". An *outcome* on an *assignment* represents the cases when the statement is true, that is all instances of E (free or as a complex subunint) on which the variable effectively displays the value. It is used as follows: "when (or if) V of E takes the value v then ..." or more succintly "when (or if) $E\{V => v\}$ then ...".

3.4.3 Influences

A *modulation* (Section 2.2.3.1) linking an *entity node* (Section 2.1) E and a relationship R means: "If E exists then R is either reinforced or weakened".

A *stimulation* (Section 2.2.3.2) linking an *entity node* (Section 2.1) 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 2.2.3.6) linking an *entity node* (Section 2.1) E and a relationship R means: "If E exists then R always takes place with necessity".

A *necessary stimulation* (Section 2.2.3.4) linking an *entity node* (Section 2.1) E and a relationship R means: "R only takes place if E exists.

An *inhibition* (Section 2.2.3.3) linking an *entity node* (Section 2.1) 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 2.2.3.5) linking an *entity node* (Section 2.1) E and a relationship R means: "If E exists then R never takes place".

doi:10.2390/biecoll-jib-2015-264

3.4.4 Logical Operators

An *and* (Section 2.1.2.1) linking several *logic arcs* originating from *entity nodes* (Section 2.1) E_i and an influence F means: "if for each i, E_i exists, then F".

An *or* (Section 2.1.2.2) linking several *logic arcs* originating from *entity nodes* (Section 2.1) E_i and an influence F means: "if for any i, E_i exists, then F".

A *not* (Section 2.1.2.3) linking a *logic arc* originating from an *entity node* (Section 2.1) E and an influence F means: " F takes place only if E does not exist".

3.4.5 Cis and trans relationships

The use of cis and trans units of information on a combination of relationships brings power and versatility to Entity Relationships. However, the resulting semantics may be difficult to grasp. Here are the basic rules that permit to understand the graphs.

- The unit of information "cis" or "trans" carried by an *interaction* refers to the *interactors* targeted by the *interaction*.
- The unit of information "cis" or "trans" carried by an *influence* targeting a state variable *assignment* refers to the origin of the *influence* and to the *entity* carrying the target of the *assignment*.
- The unit of information "cis" or "trans" carried by an *influence* targeting another *influence* refers to the origin of the carrying *influence* and to the origin of the targeted *influence*.
- The unit of information "cis" or "trans" carried by an *influence* targeting an *interaction* refers to the origin of the *influence* and all the relevant *interactors* targeted by the *interaction* (see Section 3.3.2).

3.4.6 Use of nested entities

The relationship between an *entity* and the *entities* it contains is a partonomy (meronymy). Any instance of the contained *entity* is part of an instance of the containing *entity*.
No functional relationship are implied between *entities* that are part of the same *entity*. Such relationships if any must be explicit. In particular, there is no implied disjunction.
Entity nesting is transitive. If $E1$ contains $E2$ and $E2$ contains $E3$, then any instance of $E3$ is part of an instance of $E2$ and any instance of $E2$ is part of an instance of $E1$, therefore any instance of $E3$ is part of an instance of $E1$.
The assignment of a certain value to the variable *location* of the containing *entity* effectively assigns this value to the variable *location* of all contained *entities*. For instance, a protein that translocates from the cytosol to the nucleus causes the translocation of all its domains.
The assignment of a certain value to the variable *location* of a contained *entity* does not assign this value to the variable *location* of the containing *entity*. For instance a protein domain can translocate from the cytosol into the plasma membrane, but the whole protein still belong to the cytosol.

doi:10.2390/biecoll-jib-2015-264

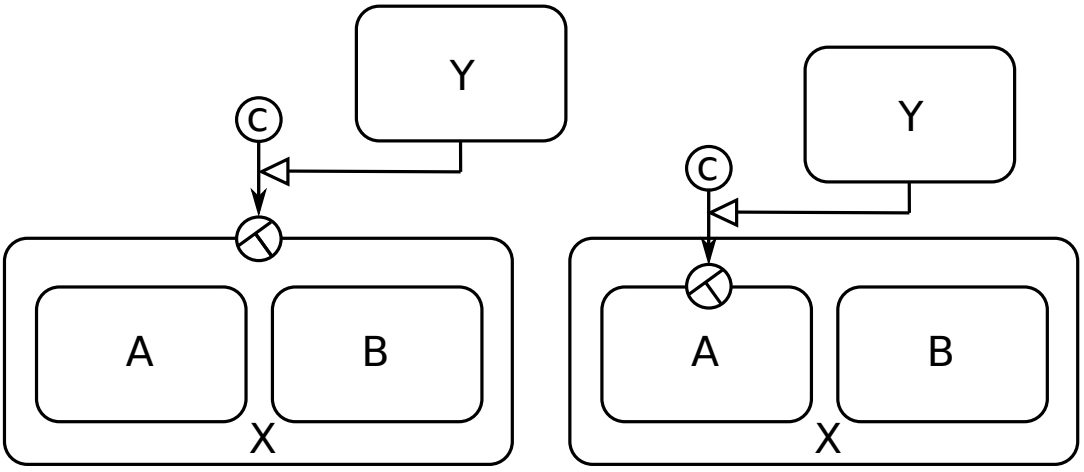

Figure 3.1: *On the left, the entity Y causes the whole entity X to translocate to the location c, including its A and B components. On the right, only the entity A is translocated, without effect on B and X.*

The assignment of a certain value to the variable *existence* of the containing *entity* effectively assigns this value to the variable *existence* of all contained *entities*. For instance, the degradation of a protein implies the degradation of all its domains.
The assignment of a certain value to the variable *existence* of a contained *entity* does not assign this value to the variable *existence* of the containing *entity*. For instance the degradation of a protein domain does not imply the degradation of the entire protein.

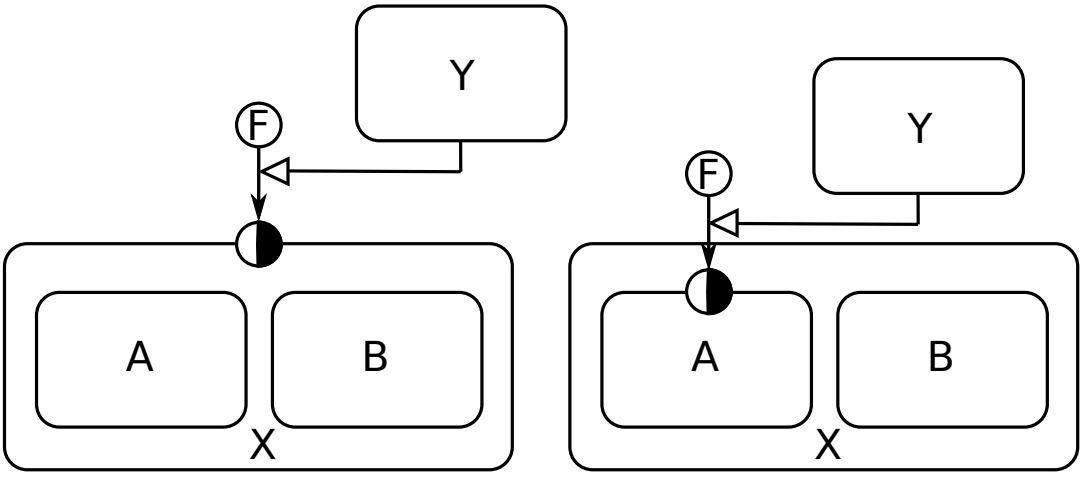

Figure 3.2: *On the left, the entity Y causes the disappearance of the whole entity X, including its A and B components. On the right, only the entity A disappear, without effect on B and X as a whole.*

Outcome of an interaction that involves contained entities is defined along the same line as in Section 2.1.1.2: the outcome represents set of all possible instances of topmost containing entity has domains represented by contained entity involved in interaction.

3.4.7 (In)Validation of ER maps

Based on the definitions above, it should be possible to use the toolkit of formal logic to analyse Entity Relationships. In particular, one can envision to build truth tables describing the consequences of the existences of the various entities. Those tables should point to inconsistencies

doi:10.2390/biecoll-jib-2015-264

leading to contradictory predicates.

doi:10.2390/biecoll-jib-2015-264

Chapter 4

Layout Guidelines for an Entity Relationship map

4.1 Introduction

The previous chapters describe the appearance and meaning of SBGN Entity Relationship Level 1 components which are *entity nodes* as well as *relationships*. The components of an Entity Relationship map have to be placed in a meaningful way – a random distribution with spaghetti-like connections will most likely hide the information encoded in the underlying model, whereas an elegant placement of the objects, giving a congenial appearance of the maps, may reveal new insights. The arrangement of components in a map is called a *layout*.

SBGN Entity Relationship maps should be easily recognizable not only by the glyphs used, but also by the general style of the layout. However, the arrangement of the components is a complex art in itself, and there is no simple rule which can be applied to all cases. Therefore this section provides guidelines for the layout of Entity Relationships, divided into two categories:

1. requirements, i. e. rules which **must** be fulfilled by a layout, and
2. recommendations, i. e. rules which **should** be followed if possible.

In addition, we provide a list of additional suggestions which may help in producing aesthetically more pleasant layouts, possibly easier to understand.

Those layout guidelines are independent of the method used to produce the map, and apply to both manually drawn maps as well as maps produced by an automatic layout algorithm. The guidelines do not deal with interactive aspects (e. g. the effect of zooming). Further information about automatic network layout (graph drawing) can be found, for example, in the books of Di Battista and co-authors [**?**] and Kaufmann and Wagner [**?**].

Please note that the color of objects has no meaning in SBGN. Although one can use colors to emphasize parts of a map or encode additional information, the meaning of the map should not depend on the colors. Furthermore, objects can have different sizes and size is also meaningless in SBGN. For example, a transition node may be larger than a protein node. Also the meaning of a graph should be conserved upon scaling as far as possible.

doi:10.2390/biecoll-jib-2015-264

4.2 Layout guidelines

4.2.1 Requirements

Requirements are rules which **must** be fulfilled by a layout to produce a valid SBGN Entity Relationship Level 1 graph.

4.2.1.1 Node-node overlaps

Nodes are only allowed to overlap in the case that the overlapping nodes define a glyph. Examples are stacking auxiliary items such as an *unit of information* on top of an *entity*, or an *interaction* and nesting *entities* to represent domains (Section 2.6). In other cases, nodes are not allowed to overlap (Figure 4.1). This includes the touching of nodes.

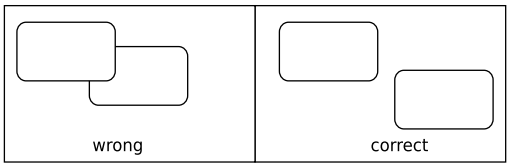

Figure 4.1: *Nodes must not overlap.*

In the case of domain nesting, the internal *entities* must be completely included in the enclosing *entities*.

4.2.1.2 Node-edge crossing

In case of node-edge crossing the edge must be drawn on the top of the node (Figure 4.2). See also recommendation 4.2.2.2 (crossing between edges and nodes should be avoided).

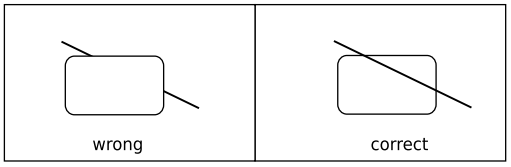

Figure 4.2: *If an edge crosses a node, the edge must be drawn on top of the node.*

4.2.1.3 Node border-edge overlaps

Edges are not allowed to overlap the border lines of nodes (Figure 4.3 on the following page).

4.2.1.4 Edge-edge overlaps

Edges are not allowed to overlap (Figure 4.4 on the next page). This includes touching of edges. Furthermore, an edge is neither allowed to cross itself nor to cross a boundary of node more than twice or other edges more than once.

4.2.1.5 Node orientation

Nodes have to be drawn horizontally or vertically, any other rotation of elements is not allowed (Figure 4.5 on the following page).

4.2.1.6 Interactions

The *interaction arcs* linking more than two *interactor nodes* are attached to a circle. Several outcomes of an interaction are not allowed to overlap (Figure 4.6 on the next page).

doi:10.2390/biecoll-jib-2015-264

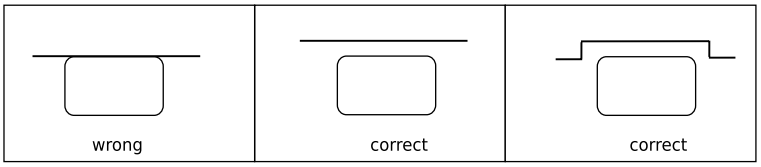

Figure 4.3: *Edges must not overlap node borders.*

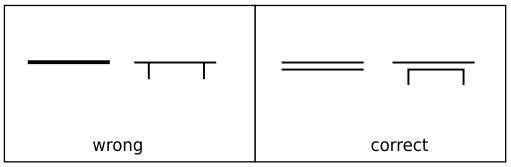

Figure 4.4: *Edges must not overlap.*

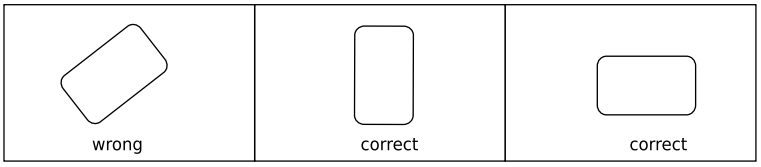

Figure 4.5: *The node orientation must be horizontally or vertically.*

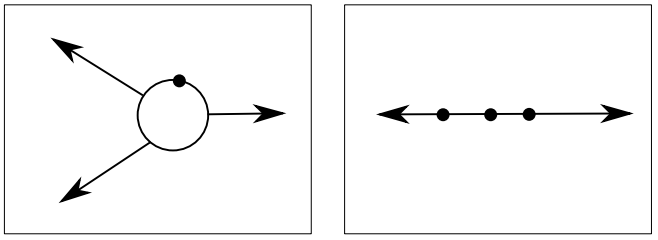

Figure 4.6: *Arcs linking more than two* interactor nodes *are attached to a circle and outcomes of an interaction are not allowed to overlap.*

4.2.1.7 Node labels

At least a part of the label (unbordered box containing a string of characters) has to be placed inside the node it belongs to. Node labels are not allowed to overlap nodes or other labels (this includes touching of other nodes or labels).

4.2.1.8 Edge labels

Edge labels are not allowed to overlap nodes. This includes touching of nodes.

4.2.2 Recommendations

Recommendations are rules which should be followed if possible to produce layouts which are easier to understand.

doi:10.2390/biecoll-jib-2015-264

4.2.2.1 *Multiple* entities *to represent the same concept*

Because rules (the influence of one entity node on a relationship) are independent of each other, a given "entity" (the concept) can be represented by many *entities* (the symbols). If a map is particularly large and an entity highly influenced or influential, it may be a good idea to represent the entity several times, limiting the influences to or from each instance. However, if systematised, such a procedure would lead to disconnected maps difficult to read and interpret. It is recommended to adopt a parsimonious approach, and multiply the symbols representing an entity only when the map become unreadable without doing so.

4.2.2.2 *Node-edge crossing*

Crossings between edges and nodes should be avoided. See also requirement 4.2.1.2 (in case of node-edge crossings the edge must be drawn on the top of the node).

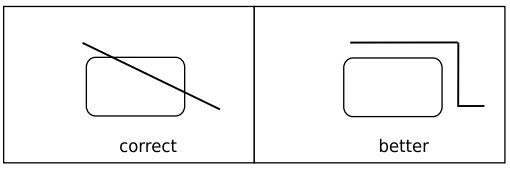

Figure 4.7: *Edges should not cross node.*

4.2.2.3 *Labels*

Labels should be horizontal. Node labels should be placed completely inside the node if possible. Edge labels should be placed close to the edge and avoid overlapping the edge as well as other edge labels.

4.2.2.4 *Avoid edge crossings*

The amount of crossings between edges should be minimized.

4.2.2.5 *Units of information*

Units of information should not hide the structure of the corresponding node and should not overlap other elements.

4.2.3 Additional suggestions

Here is a list of additional layout suggestions which may help in producing aesthetically more pleasing layouts which may be easier to understand.

- Angle of edge crossings: If edge crossings are not avoidable edges should cross with an angle close to 90 degrees.
- Drawing area and width/height ratio: The drawing should be compact and the ratio between the width and the height of the drawing should be close to 1.
- Edge length: Long edges should be avoided.
- Number of edge bends: Edges should be drawn with as few bends as possible.
- Similar and symmetric parts: Similar parts of a map should be drawn in a similar way, and symmetric parts should be drawn symmetrically.
- Proximity information: Related elements (e. g. nodes connected by an arc) should be drawn close together.

doi:10.2390/biecoll-jib-2015-264

Chapter 5

Acknowledgments

SBGN specifications are developed by many people, and with the support of many organisations.

5.1 Main contributors

In addition to the SBGN editors, the specification of SBGN ER benefited enormously from the contribution of many people. In particular, the specification of SBGN Entity Relationship Level 1 benefited much from deep discussions with Mirit Aladjem, Kurt Kohn, Emek Demir, Sohyoung Kim, Yukiko Matsuoka and Hiroaki Kitano.

5.2 Comprehensive list of contributors

Here is a more comprehensive list of people who have been actively involved in SBGN development, either by their help designing the languages, their comments on the specification, help with development infrastructure or any other useful input. We aim this list to be rather complete. We are very sorry if we forgot someone, and will be grateful if you notified us of any omission.

Mirit Aladjem, Frank Bergmann, Michael Blinov, Bernard de Bono, Sarah Boyd, Laurence Calzone, Melanie Courtot, David Croft, Tobias Czauderna, Emek Demir, Johannes W. Dietrich, Ugur Dogrusoz, Damien Fleury, Tom Freeman, Akira Funahashi, Ralph Gauges, Peter Ghazal, Samik Ghosh, Igor Goryanin, Anja Hartmann, Robin Haw, Michael Hucka, Matthias Jeschke, Mathias John, Akiya Jouraku, Astrid Junker, Hideya Kawaji, Douglas Kell, Sohyoung Kim, Hiroaki Kitano, Christian Klukas, Kurt Kohn, Fedor Kolpakov, Nicolas Le Novère, Lu Li, Augustin Luna, Yukiko Matsuoka, Carsten Maus, Alexander Mazein, Huaiyu Mi, Stuart Moodie, Ulrike Münzner, Anushya Muruganujan, Michael Pedersen, Jacqueline Quinn, Stefan Rybacki, Sven Sahle, Chris Sander, Herbert Sauro, Esther Schmidt, Falk Schreiber, Jacky Snoep, Anatoly Sorokin, Jessica Stephens, Linda Taddeo, Carolyn Talcott, Lin Uhrmacher, Martijn van Iersel, Alice Villéger, Steven Watterson, Katja Wegner (Wengler), Sarala Wimalaratne (Dissanayake), Guanming Wu, Röbbe Wünschiers.

The authors are also grateful to all the attendees of the SBGN meetings, as well as to the subscribers of the sbgn-discuss@caltech.edu mailing list.

5.3 Financial support

The development of SBGN was mainly supported by a grant from the Japanese New Energy and Industrial Technology Development Organization (NEDO, http://www.nedo.go.jp). The Okinawa Institute of Science and Technology (OIST, http://www.oist.jp), the AIST Computational Biology Research Center (AIST CBRC, http://www.cbrc.jp/index.eng.html), the British Biotechnology and Biological Sciences Research Council (BBSRC, http://www.bbsrc.ac.uk) through a Japan Partnering Award, the European Media Laboratory (EML Research gGmbH, http://www.eml.org/english), the Beckman Institute at the California Institute

doi:10.2390/biecoll-jib-2015-264

of Technology (http://bnmc.caltech.edu), Ontario Institute for Cancer Research (OICR, http://oicr.on.ca), Ontario Genomics Institute (OGI, http://www.ontariogenomics.ca), National Science Foundation (NSF, http://www.nsf.gov), USC Norris Comprehensive Cancer Center (http://uscnorriscancer.usc.edu), Martin Luther University Halle-Wittenberg (http://www.uni-halle.de), Monash University (http://www.monash.edu), IPK Gatersleben (http://www.ipk-gatersleben.de/en), University of Rostock (http://www.uni-rostock.de) and German Federal Ministry of Research and Education (http://www.bmbf.de) provided additional support for SBGN workshops. Some help was provided by the Japan Science and Technology Agency (JST, http://www.jst.go.jp) and the Genome Network Project of the Japanese Ministry of Education, Sports, Culture, Science, and Technology (MEXT, http://www.mext.go.jp) for the development of the gene regulation network aspect of SBGN, and from the Engineering and Physical Sciences Research Council (EPSRC, http://www.epsrc.ac.uk) during the redaction of the specification.

doi:10.2390/biecoll-jib-2015-264

Appendix A

Complete examples of SBGN Entity Relationship Level 1 maps

The following are complete examples of SBGN Entity Relationships maps.

A.1 Activation of RTK

The SBGN Entity Relationship Level 1 language is designed under the "open world" assumption. Under this assumption, all entities, state variables and interactions are assumed to be independent, unless explicitly specified with influence arc. This approach helps to avoid a combinatorial explosion of the states represented in the map and makes drawing diagrams much easier compared to SBGN Process Description Level 1, but it makes reading and analysis of the diagram more difficult.

To illustrate creation and analysis of SBGN Entity Relationship Level 1 diagram we will create the diagram of receptor tyrosine kinase (RTK) activation pathway in a step by step manner. Wikipedia (https://en.wikipedia.org/wiki/Receptor_tyrosine_kinase) describes the process of RTK activation as follows:

> "When a growth factor binds to the extracellular domain of an RTK, its dimerization is triggered with other adjacent RTKs. Dimerization leads to a rapid activation of the protein's cytoplasmic kinase domains, the first substrate for these domains being the receptor itself. The activated receptor as a result then becomes autophosphorylated on multiple specific intracellular tyrosine residues."

Drawing SBGN Entity Relationship Level 1 diagrams is simple, we will follow the textual description of the pathway and add required elements. Let's start with identification of key players of the pathway. From the text, we can determine that the two main entities are "RTK" and "growth factor". We will draw them as "receptor" and "ligand" entities (Section 2.1.1.1), respectively. We also notice that RTK has two domains: the "extracellular domain" ("receptor") and the "kinase domain", they will also be shown in Figure A.1 on the following page. The pathway description also mentions "multiple specific intracellular tyrosine residues", so we have added two state variables (Section 2.3.2) "Y1" and "Y2"' to the receptor entity to be able to show events related to modification of the tyrosine residues.

doi:10.2390/biecoll-jib-2015-264

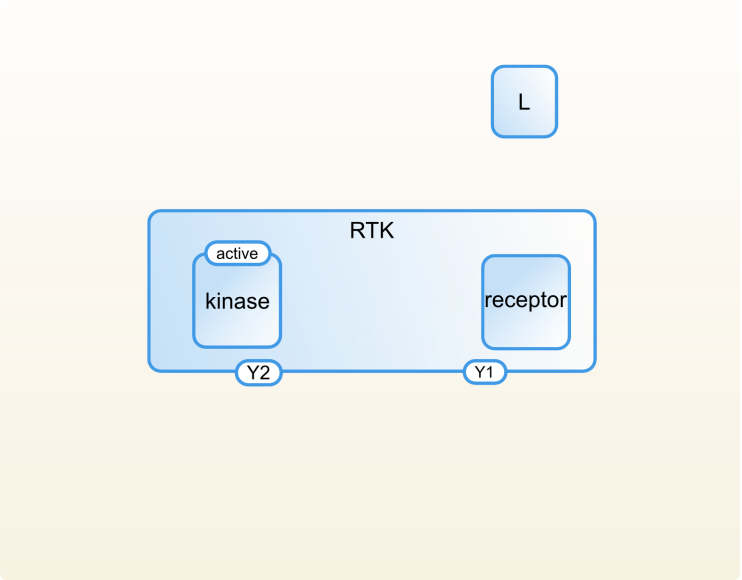

Figure A.1: *Receptor tyrosine kinase activation: definition of entities.*

Through this entity creation process, we have populated our system with instances of two kinds: a "receptor" and a "ligand". We also show the internal structure of the receptor by depicting its two domains as nested entities (Section 2.6). The receptor and its "kinase" domain have "state variables" (Section 2.3.2) that define the basis for the entity state space.

We will now describe the entity state space for each entity that we are interested in. The two assignment arcs (Section 2.2.1.1) pointing to the "Y1" and "Y2" receptor state variables on Figure A.2 on the next page demonstrate that we are going to analyse phosphorylation of tyrosine residues. From the statement: "Dimerization leads to a rapid activation of the protein's cytoplasmic kinase domains", we need to take into account that the "active" state variable of "kinase" domain could become "true" after some perturbations in the system. Therefore, we have added an appropriate variable value (Section 2.3.3) node and an assignment arc.

Given that the SBGN Entity Relationship Level 1 follows the "open world" assumption, diagram authors are not obliged to describe the entire system, and we can focus on the most important parts of the system. For this reason, we include only the "P" state value for the residues. For a more accurate description of the system, we need to include a default value, such as the "unphosphorylated" state (see Section 2.2.1.1 to check how to do it properly), but in our case it is not important so we will rely on the "open world" assumption and omit this detail. SBGN Entity Relationship Level 1 assumes that all interactions are internally reversible, so if we only describe the switch from default to phosphorylated state explicitly, that description implies that there is a mechanism to put "Y1" and "Y2" variables back to their default state. For the same reason, we have not added "false" value to the "active" state of the kinase domain.

doi:10.2390/biecoll-jib-2015-264

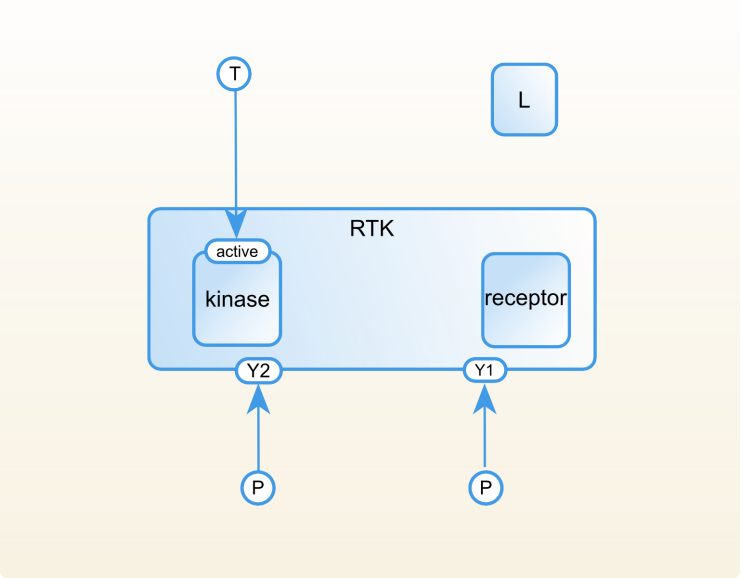

Figure A.2: *Receptor tyrosine kinase activation: state variable assignment.*

We have now prepared all elements that are taking part in the process, and we are ready to draw the pathway itself. The first sentence of the pathway description states "When a growth factor **binds** to the extracellular domain of an RTK", so the first event we are going to draw is a receptor ligand binding (Figure A.3). We will use an interaction arc (Section 2.2.1.2) to show that the ligand can bind the receptor.

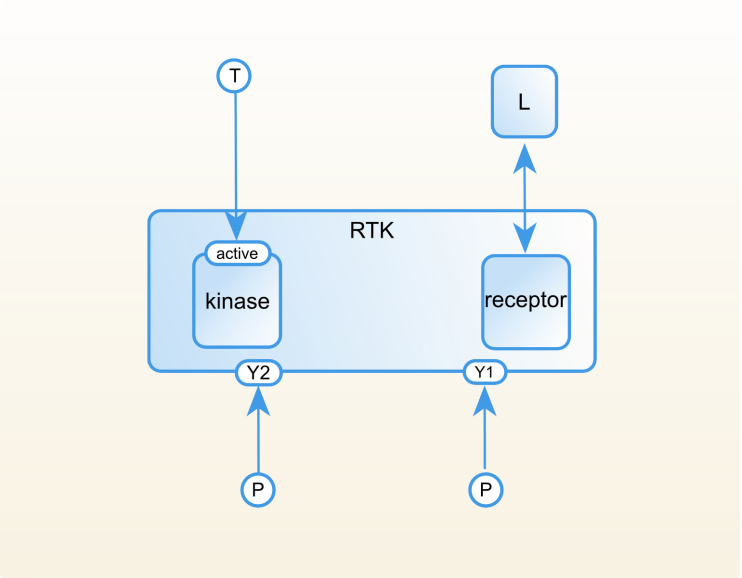

Figure A.3: *Receptor tyrosine kinase activation: receptor-ligand binding.*

The binding happens between the ligand entity and the receptor domain of the RTK entity. The interaction event is reversible in a same way as assignment. That means that even if two instances of ligand and receptor are bound at some time that bond will not last forever, it will dissociate with some probability.

The next statement in the description is "When a growth factor binds to the extracellular domain of an RTK, its **dimerisation** is triggered with other adjacent RTKs". The dimerisation event is added in Figure A.4 on the next page. The interaction arc connects the entity with itself,

doi:10.2390/biecoll-jib-2015-264

so we need to distinguish two possibilities: 1) the receptor molecule binds another molecule of the same kind (what we need according to the pathway description) or 2) that RTK undergoes an intramolecular interaction. In the first case, we add a *unit of information* "trans" to the arc to emphasise that the interaction is intermolecular. The second case would require *unit of information* "cis" (see Section 2.3.1, Section 2.4.4 and Section 3.4.5). Definition of "cis" or "trans" interaction is meaningful only for events involving instances of the same entity occurring directly or via chain of events. If you do not specify cis/trans type of interaction, this information will be undefined, because of the "open world" nature of SBGN Entity Relationship Level 1. This is correct and sometimes unavoidable, but the diagram is more understandable when the appropriate *unit of information* is added to the interaction arc in accordance with biological knowledge.

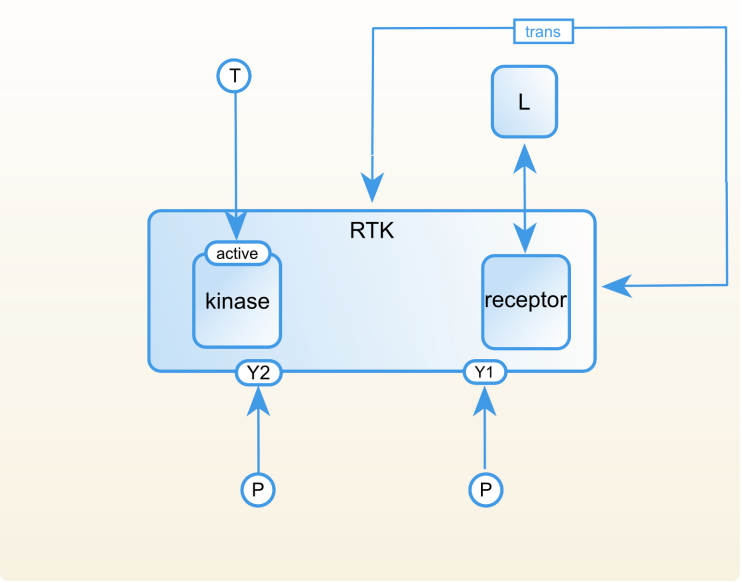

Figure A.4: *Receptor tyrosine kinase activation: receptor dimer formation.*

Now, we have to emphasise that the dimerisation depends upon the ligand-receptor binding. There are two key elements of SBGN Entity Relationship Level 1 we will use in this step (see Figure A.5 on the following page): outcome (Section 2.1.1.2) and influence arc (Section 2.2.3). In our case the result of interaction is a complex containing instances of receptor and ligand, but this is not limited to the binary complex, as we will see soon. To show the results of an interaction on the diagram an outcome node is placed on the on the interaction arc, as a black dot. This outcome dot is the origin of our influence arc (Section 2.2.3). Influence arcs are designed to depict the way one part of the system controls the behaviour of the other. The word "triggered" in a pathway description implies that the interaction between receptor and ligand is necessary for the dimerisation process, so we have used necessary stimulation (Section 2.2.3.4), which indicates that without the stimulator entity the target interaction cannot take place. Furthermore, we have placed "cis" *unit of information* on the influence arc to emphasise that the control happens within the same molecular complex. That means that the receptor-ligand complex is a member of the formed dimer.

doi:10.2390/biecoll-jib-2015-264

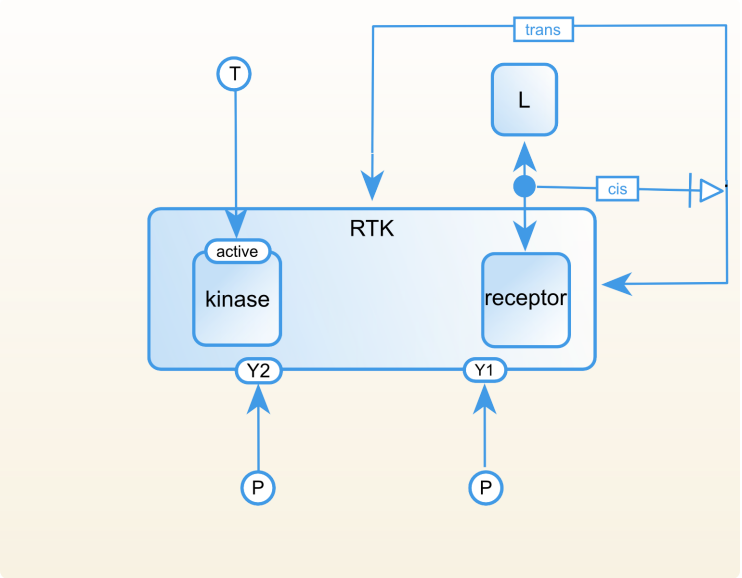

Figure A.5: *Receptor tyrosine kinase activation: control of the dimer formation.*

The next sentence in the description states: "Dimerization leads to a rapid activation of the protein's cytoplasmic kinase domains,". In Figure A.6, we will add another influence arc, stimulation (Section 2.2.3.2) that means the kinase domain may be active to some extent even in monomeric form of a receptor. On the dimerisation interaction, we again have an outcome dot to represent the dimer, as the source of the influence arc, but in this case the outcome represents any kind of complex where two receptor molecules bound together.

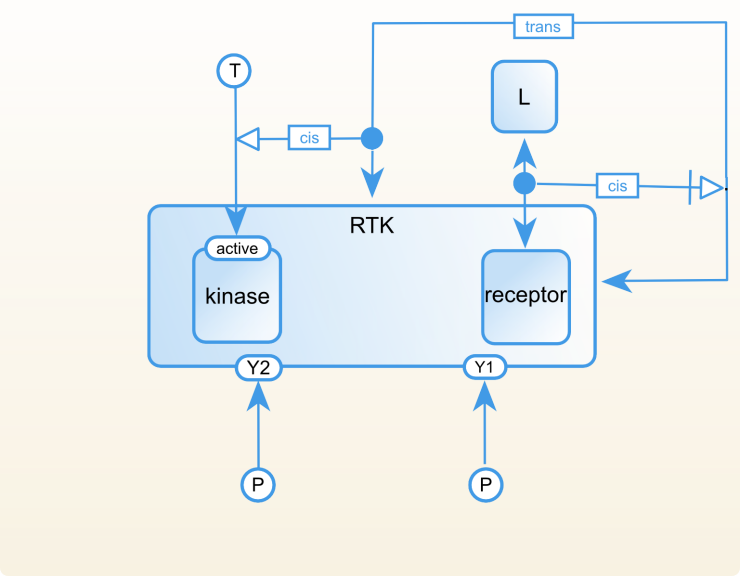

Figure A.6: *Receptor tyrosine kinase activation: activation of the kinase domain of the receptor.*

The last part of our pathway is the "autophosphorylation of multiple specific intracellular tyrosine residues" by activated kinase domain. The complete pathway is shown in Figure A.7 on the following page. All interactions and influences in SBGN Entity Relationship Level 1 are independent, so each stimulation arc representing autophosphorylation of the tyrosine residue has its own outcome, and these outcomes represent independent sets of instances where kinase

doi:10.2390/biecoll-jib-2015-264

is active.

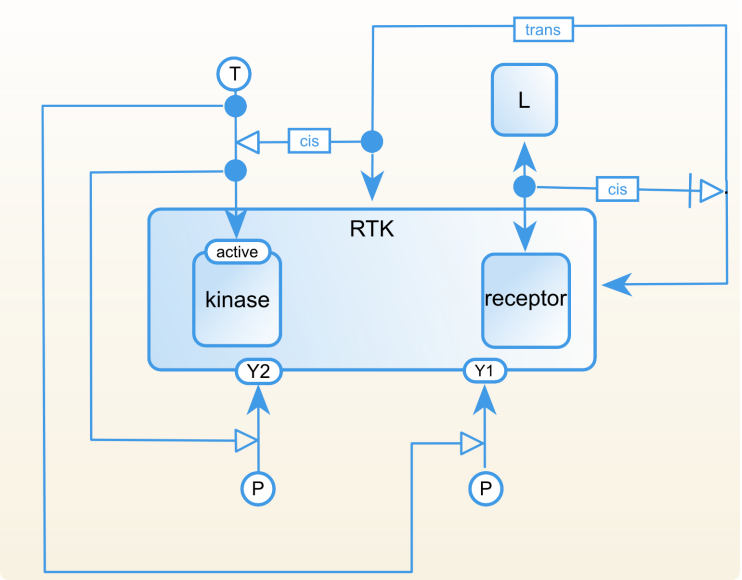

Figure A.7: *The complete diagram of receptor tyrosine kinase activation.*

Now that our diagram is ready let's analyse it to get a better understanding of rules of the language and meaning of the diagram elements. The diagram consists of two types of entities, three types of outcomes, two interaction arcs, three assignment arcs, and four influence arcs. If we take into account independence and reversibility of all interactions we can count the total number of complexes that could be created in the system. It will be 153 complex types in total: 8 monomeric receptors, 8 monomeric receptor-ligand complexes, 36 receptor dimers, 64 receptor dimers with one ligand, 36 receptor dimers with two ligands and free ligand. Those numbers come from simple combinatorics, taking into account the number of state variables and assuming that each of them can have two values: "undefined" or specified by assignment arc. Figure A.8 on the next page shows three of the 144 complexes compatible with the defined outcome.

doi:10.2390/biecoll-jib-2015-264

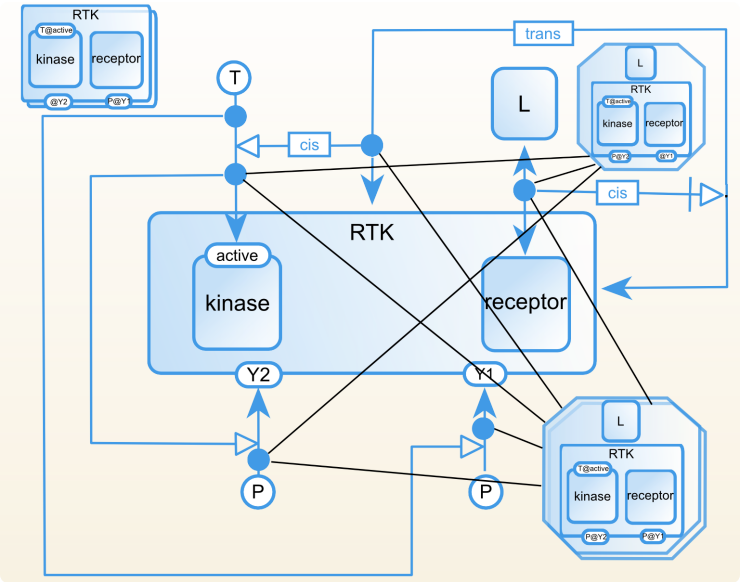

Figure A.8: *Receptor tyrosine kinase activation: examples of active kinase complexes are shown. Black lines connect two complexes with outcomes, which definition is compatible with the complex status.*

The large number possible complexes is defined by the influence arcs, which dictate the number of available states. For example, blocking any single phosphorylation site reduces the number of reachable complexes to 45, and blocking the kinase activation reduces the total set of possible complexes types to just 6. This can easily be seen from the diagram. In the first case, we switch off one receptor state variable, and in the second case, we effectively block all three; blocking activation assignment we render all its outcomes to false state. This means that activation will never happen and that the tyrosine residues will never be phosporylated. In the next section, we will present more complex examples.

A.2 Additional examples

Figure A.9 on the following page presents the different relations between the four entities involved in a Polymerase Chain Reaction (PCR). This example shows the usage of the *entity*, the logical operator *or*, the *state variable* "existence", the *unit of information*, as well as the relationships *interaction*, *assignment*, *necessary stimulation* and *absolute inhibition*.

doi:10.2390/biecoll-jib-2015-264

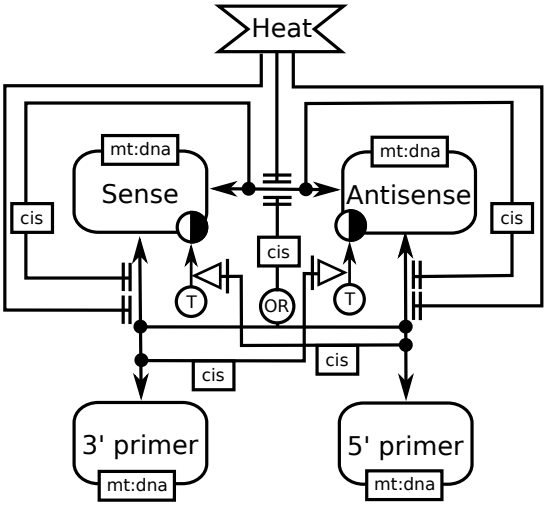

Figure A.9: *Principle of the Polymerase Chain Reaction.*

Figure A.10 depicts the effect of depolarisation (dV) on the intracellular calcium, that binds to calmodulin, that itself binds to the calcium/calmoduline kinase II (CaMKII). In this diagram, the binding of calmodulin inhibits the folding of CaMKII monomer on itself, thus relieving the inhibition on the kinase activity. The phosphorylation of the glutamate receptors finally leads to the Long Term Potentiation (LTP) of the synapses. In addition, the map shows the effect of trans-phosphorylation on threonine 286, that makes the enzyme constitutively active, and on threonine 306, that renders the kinase insensitive to calmodulin, as well as the dimerisation of the kinase.

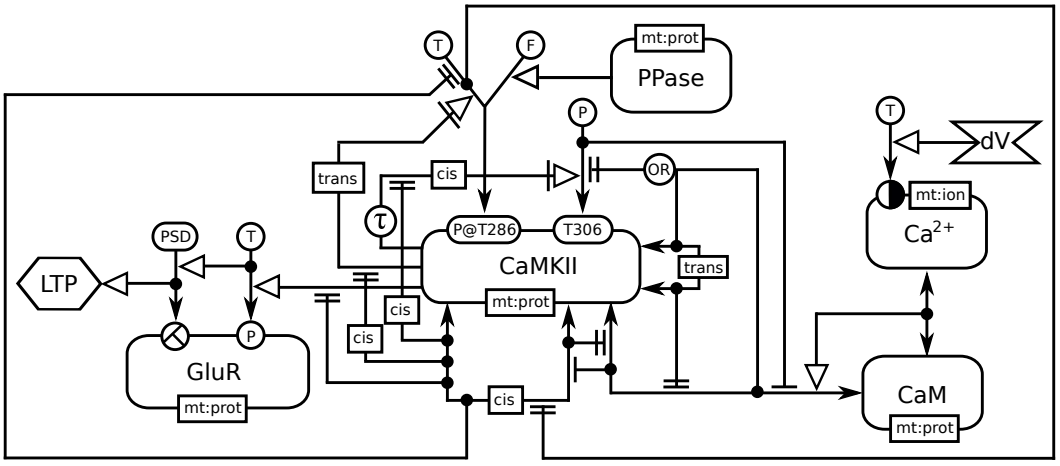

Figure A.10: *Regulation of calcium/calmoduline kinase II effect on synaptic plasticity.*

doi:10.2390/biecoll-jib-2015-264

Appendix B

Reference card

Print this summary of SBGN Entity Relationship symbols for a quick reference.

doi:10.2390/biecoll-jib-2015-264

SYSTEMS BIOLOGY GRAPHICAL NOTATION: ENTITY RELATIONSHIP REFERENCE CARD

doi:10.2390/biecoll-jib-2015-264

Appendix C

Issues postponed to future levels

C.1 Generics and instances

In SBGN Entity Relationship Level 1, an entity is represented only once. One cannot explicitly represent different instances of the "same" entity. Several instances can be inferred from relationships acting in *trans*. However, one cannot generally express the fact that several relationships involving the same entity actually involve the same, or different, instances of this entity. This problem is tied to the problem of generics. Indeed, if one discriminate between classes of instances, how can one represent, in the same map, the generic entity?

C.2 Groups

The SBGN Entity Relationship Level 1 needs a mechanism to "tag" an SBGN glyph as belonging to one or several groups. Those groups would correspond for instance to "pathways" or "metabolic network", glyphs associated with a disease or biological function. A group does not affect the syntax of the map, but is merely a multi-glyph annotation.

Groups would be a way for instance to organise nodes together in a certain subpart of the plan, or highligtht them in some way. Software would have to conserve groups, and it could thus be a way to lightly constrain the layout, without going all the way to specify position and size of the nodes. A group would not be "linked" to any node using edges, but would "contain" EPN/PN in PD, EN in ER, AN in AF (i.e. in PD and AF, a group could span several compartment). Groups could be named and could be annotated with "floating" annotations, similarly to the default compartment.

C.3 Submap

The SBGN Process Description Level 1 has a *Submap* as placeholder for another process and is used when one wishes to hide the detail of this process from the Process Description map, but make it available to the reader as a separate related map. The similar concept could be implemented in SBGN Entity Relationship Level 1.

C.4 Spatial organisation

Introduction of *Nested entities* makes representation of domains possible. The additional feature required by some users is to be able to show relation between domains. For example, promoter, coding region and terminator would be domains of the plasmid entity and it would be important to show that promoter is located before coding region.

doi:10.2390/biecoll-jib-2015-264

Appendix D

Revision History

D.1 Version 1.2 to Version 2

Below are the changes incorporated into Version 2 of the SBGN Entity Relationship Level 1 specification. The Tracker IDs correspond to the sourceforge tracker "SBGN ER L1" (http://sourceforge.net/p/sbgn/sbgn-er-l1/). The message IDs correspond to the "sbgn-discuss" mailing list (https://utils.its.caltech.edu/pipermail/sbgn-discuss/).

Description	Message or Tracker ID
Definition of outcome as a subset of instances	
User manual (RTK example with extended description) is added	
Synchronous events	
Nested entities	
Clarification of entity definition	
Clarification of outcome definition hline Add interaction outcome in the ER spec.	track: 25
Definition of outcome with rel to nested entities	track: 22
Error in SBGN-ER Specifications	track: 27
	track:

D.2 Version 1.1 to Version 1.2

Below are the changes incorporated into Version 1.2 of the SBGN Entity Relationship Level 1 specification. The Tracker IDs correspond to the sourceforge tracker "SBGN ER L1" (http://sourceforge.net/tracker/?group_id=178553&atid=1170625). The message IDs correspond to the "sbgn-discuss" mailing list (https://utils.its.caltech.edu/pipermail/sbgn-discuss/).

Description	Message or Tracker ID
The section "annotation links" of the layout chapter has been removed	
Clarification that the link between annotation and annotated must be a callout	track: 3240913
Clarification that an outcome cannot be targeted by an influence, and must carry only one influence	track: 3211399
Clarification that an interaction contour is always a circle	track: 3178637
Explanation that binary interaction can use a circle	track: 3178631
Clarification of what is an entity throughout	track: 2921526
Added caption 'Auxiliary items' for *necessary stimulation* and *absolute inhibition*	track: 3115480

continued on next page

doi:10.2390/biecoll-jib-2015-264

continued from previous page

Description	Message or Tracker ID
A section has been added to describe the *variable value* as a proper glyph	
Mentions of *non-interaction* have been removed from 6 places	track: 3115477

D.3 Version 1.0 to Version 1.1

Below are the changes incorporated into Version 1.1 of the SBGN Entity Relationship Level 1 specification. The Tracker IDs correspond to the sourceforge tracker "SBGN ER L1" (http://sourceforge.net/tracker/?group_id=178553&atid=1170625). The message IDs correspond to the "sbgn-discuss" mailing list (https://utils.its.caltech.edu/pipermail/sbgn-discuss/).

Description	Message or Tracker ID
The state variable value has been added on the reference card, as an entity node	
Auxiliary units have been moved after relationships, to avoid misunderstanding that they are only relevant for entities	track: 3051017
The description of a variable assignment with several alternative value has been clarified	track: 3004692 track: 3069103
The figures containing influences on logic arcs have been fixed	track: 2915856
The "fossil" mention a glyph *non-interaction* has been removed from the reference card	track: 2915853
The figures containing unit of informations on logic arc have been fixed	track: 2915852
Following a vote from the community, the link from an annotation to the annotated symbol is no longer undefined but is a callout. The example figure has been amended accordingly	msg: 000245
A revision history has been added at the end of the document	

doi:10.2390/biecoll-jib-2015-264

Bibliography

[1] Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, New Jersey, 1998.

[2] Michael Kaufmann and Dorothea Wagner. *Drawing Graphs: Methods and Models*, volume 2025 of *Lecture Notes in Computer Science*. Springer, 2001.

doi:10.2390/biecoll-jib-2015-264