

Systems Biology Graphical Notation: Process Description language Level 1

Version 2.0

Date: November 1, 2011

Disclaimer: This is a working draft of the SBGN Process Description Level 1 Version 2.0 specification. It is not a normative document.

Editors:

Stuart Moodie
Nicolas Le Novère
Emek Demir
Huaiyu Mi
Alice Viléger

Edinburgh, UK
EMBL European Bioinformatics Institute, UK
Sloan-Kettering Institute, USA
University of Southern California, USA
University of Manchester, UK

To discuss any aspect of SBGN, please send your messages to the mailing list sbgn-discuss@sbgn.org. 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://p.sf.net/sbgn/pd_tracker.



Contents

1	Introduction	1
1.1	SBGN levels and versions	1
1.2	Developments, discussions, and notifications of updates	2
1.3	Note on typographical convention	2
2	Concepts	3
2.1	Definitions and Nomenclature	4
2.2	The Language Definition	5
3	Concepts and Glyphs	6
3.1	Introduction	6
3.2	Definitions	6
3.3	Decorators	37
3.4	Controlled vocabularies	39
4	Validation Rules	42
4.1	Overview	42
4.2	Semantic rules	42
5	Layout Rules for a Process Description	46
5.1	Introduction	46
5.2	Requirements	46
5.3	Recommendations	48
5.4	Additional suggestions	49
6	Acknowledgments	51
6.1	Level 1 Release 1.0	51
6.2	Level 1 Release 1.1	51
6.3	Level 1 Release 1.2	51
6.4	Level 1 Release 1.3	51
6.5	Comprehensive list of acknowledgements	51
6.6	Financial Support	52
A	Complete examples of Process Description Maps	53
B	Reference card	57
C	Issues postponed to future levels	59
C.1	Multicompartment entities	59
C.2	Logical combination of state variable values	59
C.3	Non-chemical entity nodes	59
C.4	Generics	60
C.5	State and transformation of compartments	60
D	Revision History	61
D.1	Version 1.0 to Version 1.1	61
D.2	Version 1.1 to Version 1.2	62
D.3	Version 1.2 to Version 1.3	62

Chapter 1

Introduction

The goal of the Systems Biology Graphical Notation (SBGN) is to standardize the graphical/visual representation of 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, 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 *Process Description* visual language of SBGN. Process Descriptions are one of three views of a biological process offered by SBGN. It is the product of many hours of discussion and development by many individuals and groups.

1.1 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 need 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.

1.2 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.

1.3 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 process is encoded by the SBGN PD process”.

Chapter 2

50

Concepts

51

To set the stage for what follows, we first give a brief overview of some of the concepts in the Process Description language with the help of an example shown in Figure 2.1.

52

53

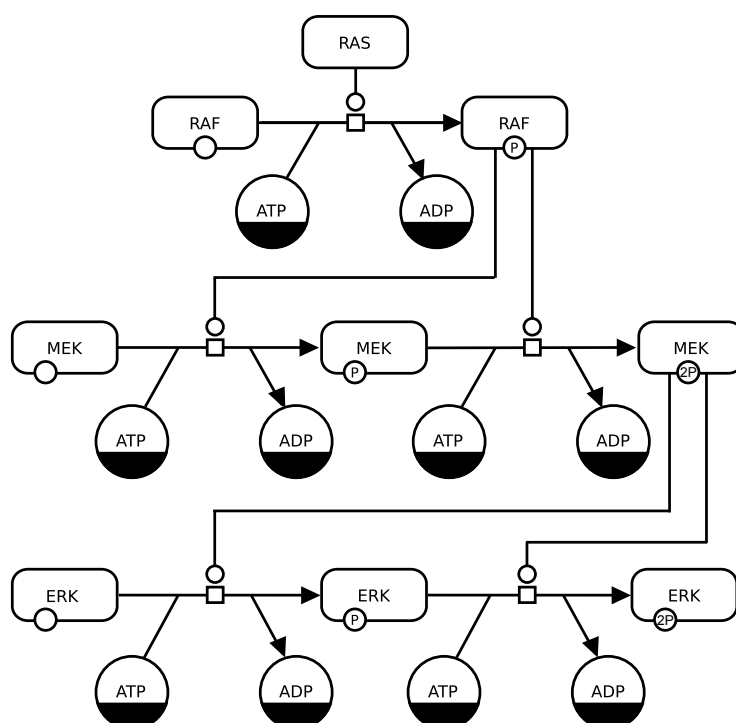


Figure 2.1: This example of a Process Description uses two kinds of entity pool nodes: one for pools of different macromolecules (Section 3.2.11) and another for pools of simple chemicals (Section 3.2.7). Most macromolecule nodes in this map are adorned with state variables (Section 3.2.17) representing phosphorylation states. This map uses one type of process node, the process node (Section 3.2.15), and three kind of connecting arc, consumption (Section 3.2.19), production (Section 3.2.19) and catalysis (Section 3.2.19). Finally, some entity pool nodes have dark bands along their bottoms; these are clone markers (Section 3.3.1) indicating that the same pool nodes appear multiple times in the map.

The map in Figure 2.1 is a simple map for part of a mitogen-activated protein kinase (MAPK) cascade. The larger nodes in the figure (some of which are in the shape of rounded rectangles and others in the shape of circles) represent biological materials—things like macromolecules and simple chemicals. The biological materials are altered via processes, which are indicated in Process Description language by lines with arrows and other decorations. In this particular map, all of the processes happen to be the same: processes catalyzed by biochemical entities. The directions of the arrows indicate the direction of the processes; for example, unphosphorylated RAF kinase processes to phosphory-

54

55

56

57

58

59

60

lated RAF kinase via a process catalyzed by RAS. Although ATP and ADP are shown as incidental to the phosphorylations on this particular graph, they are involved in the same process as the proteins getting phosphorylated. The small circles on the nodes for RAF and other entity pools represent state variables (in this case, phosphorylation sites).

The essence of the Process Descriptions is *change*: it shows how different entities in the system process from one form to another. The entities themselves can be many different things. In the example of Figure 2.1, they are either pools of macromolecules or pools of simple chemicals, but as will become clear later in this chapter, they can be other conceptual and material constructs as well. Note also that we speak of *entity pools* rather than individuals; this is because in biochemical network models, one does not focus on single molecules, but rather collections of molecules of the same kind. The molecules in a given pool are considered indistinguishable from each other. The way in which one type of entity is transformed into another is conveyed by a *process node* and links between entity pool nodes and process nodes indicate an influence by the entities on the processes. In the case of Figure 2.1, those links describe consumption Section 3.2.19, production Section 3.2.19 and catalysis Section 3.2.19, but others are possible. Finally, nodes in Process Descriptions are usually not repeated; if they do need to be repeated, they are marked with *clone markers*—specific modifications to the appearance of the node (Section 3.3.1). The details of this and other aspects of Process Description notation are explained in the following chapters.

2.1 Definitions and Nomenclature

2.1.1 Language versus notation

SBGN specifications propose symbols, ways to organise them, but also semantic rules to analyse the resulting representations. SBGN "drawings" can be translated into English, but also into computer readable formats. Those specifications really propose true languages. SBGN is therefore made up of three languages.

2.1.2 What are the languages?

PD is a language that permits the description of all the processes taking place in a biological system. The ensemble of all these processes constitute a Description. **ER** is a language that permits the description of all the relations involving the entities of a biological system. The ensemble of all these relations constitute a Relationship. **AF** is a language that permits the description of the flow of activity in a biological system.

2.1.3 Nomenclature

The three languages of SBGN should be referred to as:

- the Process Description language.
- the Entity Relationship language.
- the Activity Flow language.

Abbreviated as:

- the PD language.
- the ER language.
- the AF language.

A specific representation of a biological system in one of the SBGN languages should be referred to as:

- a Process Description map.
- an Entity Relationship map.
- an Activity Flow map.

Abbreviated as:

- a PD map.
- an ER map.
- an AF map.

The corpus of all SBGN representations should be referred to as:

- Process Descriptions.
- Entity Relationships.
- Activity Flows.

The capitalization is important. PD, ER and AF are names of languages. As such they must be capitalized in English. This is not the case of the accompanying noun (language or map).

2.1.4 Graph, diagram or map?

A graph is a very technical term that belongs to mathematics and is uncommon in biology. Diagram is a concept that encompasses more than just graph. Examples are Venn diagrams for instance. Therefore, we recommend using the term map for SBGN representations. Those representations effectively permit users to travel and orient themselves in a biological system. Map is also the term most frequently used by the different communities, whether in metabolism, signaling or genomics.

2.2 The Language Definition

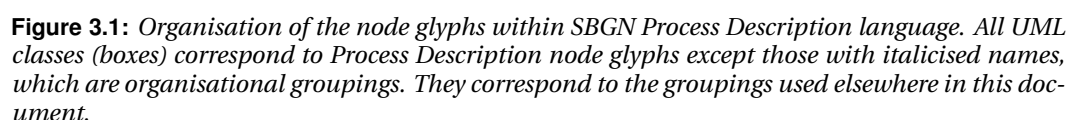
Language defined as a set of classes that represent the key concepts of the language. Done in UML. The top level classes are roughly equivalent to the grapheme, which is the basic language unit. The grapheme can then be represented by one or more glyph. For example in English the grapheme of 'a' can be represented by the glyphs *a*, *a* or *a*. In this specification we refer to a concept that organises the language as an *abstract class* and a top-level concept, i.e., a grapheme, as a *concrete class* or both generically as just a *class*. The graphical representation of the concrete class is the *glyph*.

133

134

135

136
137
138
139
140
141



142

143

144
145
146
147
148
149

Name: SBGN Element

SBO Term: Not defined

Attributes:

id:identity the implicit identity of an SBGN element ensuring all instances are unique.

Auxiliary Items: N/A

Special constraints or rules: N/A

Cloning: N/A

3.2.2 Map

The Map (figure 3.2) is a concrete class that extends the SBGNElement. It describes the visible container upon which all glyphs are drawn. The map can take any pattern or colour, but as SBGN is 'colour blind' this does not convey any meaning in itself.

Name: Map

SBO Term: Not defined

Attributes:

glyphs:SBGNGlyph The collection of glyphs held by the map.

Auxiliary Items: All SBGN glyphs that extend SBGNGlyph

Special constraints or rules: A map is valid if it is empty (although not very useful)

Cloning: N/A

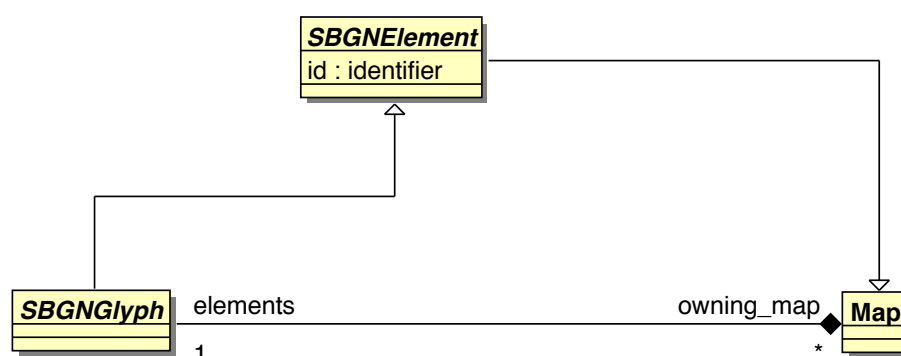


Figure 3.2: UML definition of Map and SBGNElement.

3.2.3 Glyphs, Auxiliary Units, Nodes and Arcs

In addition to the map elements within SBGN are either descendants of SBGNGlyph (glyphs) or SBGNAuxiliaryUnit (auxiliary units). See figure 3.3. Glyphs are the fundamental building blocks of the Process Description language and are the only elements that can be drawn on a map (MAP). Auxiliary units in contrast cannot and are used to adorn glyphs and providing additional information that may be useful to the reader. These can provide annotation (annotation), state information (state variable) or indicate duplication of entity pool nodes (clone marker). Glyphs are further subdivided into nodes (SBGNGlyph) and edges (SBGNArc) that define the graph structure inherent in a Process Description map.

3.2.4 Entity pool nodes

An entity pool is a population of entities that cannot be distinguished from each other, when it comes to the SBGN Process Description Level 1 map. For instance all the molecular entities that fulfill the same role in a given process form an entity pool. As a result, an entity pool can represent different granularity levels, such as all the proteins, all the instances of a given protein, only certain forms of a given protein. To belong to a different compartment is sufficient to belong to different entity pools.

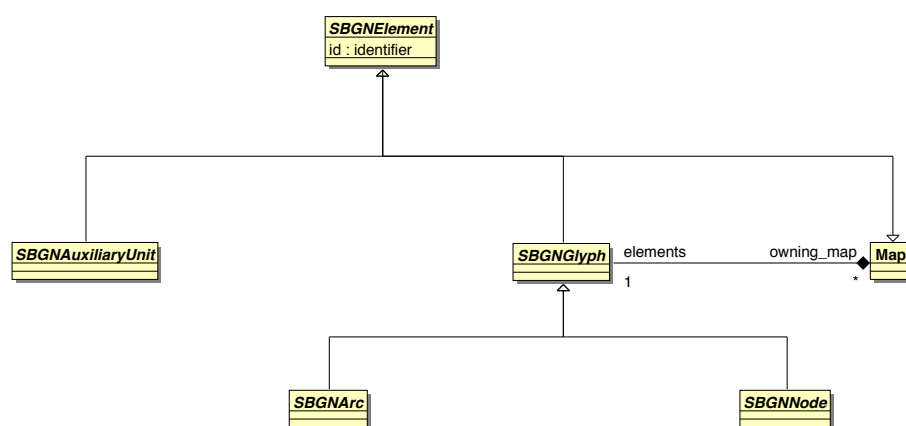


Figure 3.3: UML definition of the glyph, auxiliary unit, nodes and arcs.

Calcium ions in the endoplasmic reticulum and calcium ions in the cytosol belong to different entity pools when it comes to representing calcium release from the endoplasmic reticulum.

The Process Description contains five glyph types representing classes of material entities: unspecified entity (Section 3.2.8), simple chemical (Section 3.2.7), macromolecule (Section 3.2.11), nucleic acid feature (Section 3.2.12) and complex (Section 3.2.13). (Specific types of macromolecules, such as protein, RNA, DNA, polysaccharide, and specific simple chemicals are not defined by Process Description but may be part of future levels of SBGN.) In addition to the material entities, Process Description represents a conceptual entities: perturbing agent (Section 3.2.9).

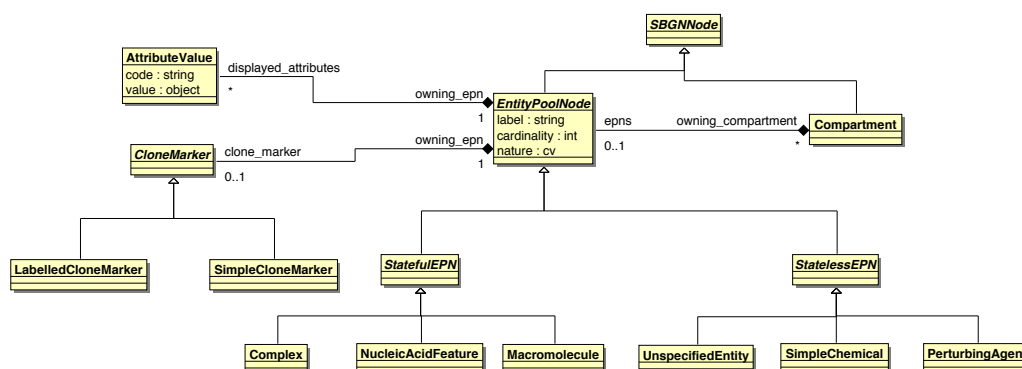


Figure 3.4: UML definition of the entity pool node and its descendant glyphs.

The EntityPoolNode (figure 3.4) is the common ancestor of all these glyphs and defines attributes that are common to all its descendants. It must belong to a compartment (c.f. section 3.2.16) and can contain a clone marker if it is cloned. Note that not all EPNs can be cloned.

EntityPoolNode

SBO Term: Not defined

Attributes:

- label:string The name that identifies the entity in the Process Description map. EPNs with the same label should be from the same entity. the string cannot be empty and must start and end with a non-space character. Any Unicode character is acceptable.
- cardinality:int The number of copies of the entity. Must be a positive non-zero integer.
- nature:cv The nature of the entity pool node as defined by a controlled vocabulary.
- owning_compartment:Compartment The compartment that this EPN belongs too.
- clone_marker:CloneMarker The clone marker decorator. See section 3.3.1 for its use.

displayed_attributes:AttributeValue One or more decorators used to display attribute values. 204

Auxiliary Items: Annotation 205

Special constraints or rules: 1. If cardinality > 1 then the descendant glyph must be displayed as a 206
multimer. 2. If the EPN is drawn directly on a Map then owning_compartment is assigned to an 207
invisible default compartment. nature can only use the material type (section ??), conceptual 208
type (section ??) or physical characteristics (section ??) controlled vocabularies. 209

Cloning: Defined by descendants. 210

3.2.5 Empty Set 211

It is useful to have the ability to represent the creation of an entity or a state from an unspecified 212
source, that is, from something that one does not need or wish to make precise. For instance, in a 213
model where the production of a protein is represented, it may not be desirable to represent all of 214
the amino acids, sugars and other metabolites used, or the energy involved in the protein's creation. 215
Similarly, we may not wish to bother representing the details of the destruction or decomposition of 216
some biochemical species into a large number of more primitive entities, preferring instead to simply 217
say that the species “disappears into a sink”. Yet another example is that one may need to represent 218
an input (respectively, output) into (resp. from) a compartment without explicitly representing a 219
transport process from a source (resp. to a target). 220

For these and other situations, SBGN defines a single glyph to handle these situations represent- 221
ing the involvement of an external pool of entities. The symbol used in SBGN is borrowed from the 222
mathematical symbol for “empty set”, but it is important to note that it does not actually represent a 223
true absence of everything or a physical void—it represents the absence of the corresponding struc- 224
tures in the map, that is, the fact that the external pool is conceptually outside the scope of the 225
map. 226

A frequently asked question is, why bother having an explicit symbol at all? The reason is that 227
one cannot simply use an arc that does not terminate on a node, because the dangling end could 228
be mistaken to be pointing to another node in the map. This is specially true if the map is rescaled, 229
causing the spacing of elements in the map to change. The availability and use of an explicit symbol 230
for sources and sinks is critical. 231

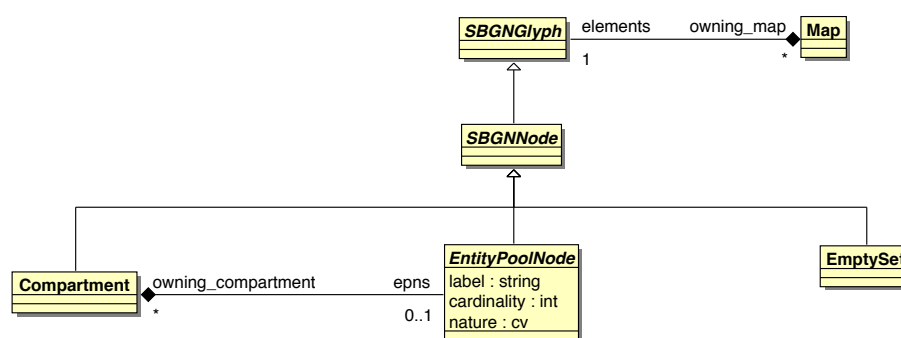
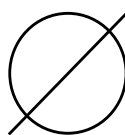


Figure 3.5: The UML definition of the EmptySet and its context in relation to other elements of the Process Description language.

The definition of the Empty Set is shown in figure 3.5. The empty set is not an EPN as it does not 232
represent a single pool of entities and does not share any of the other attributes of an EPN, nor does 233
it belong to a particular compartment. 234

Glyph: Empty Set

235

**Figure 3.6:** *The empty set glyph.***SBO Term:** SBO:0000291 ! empty set

236

Auxiliary Items: None

237

Special constraints or rules:

238

Cloning: All instances of Empty Set can be regarded as identical therefore not special decoration is used to indicate replication on the map.239
240**Container** Represented by the mathematical symbol for “empty set”, that is, a circle crossed by a bar linking the upper-right and lower-left corners of an invisible square drawn around the circle (\emptyset). Figure 3.6 illustrates this. The symbol should be linked to one and only one edge in a map.

241

242

243

Label None

244

3.2.6 Stateless Entity Pool

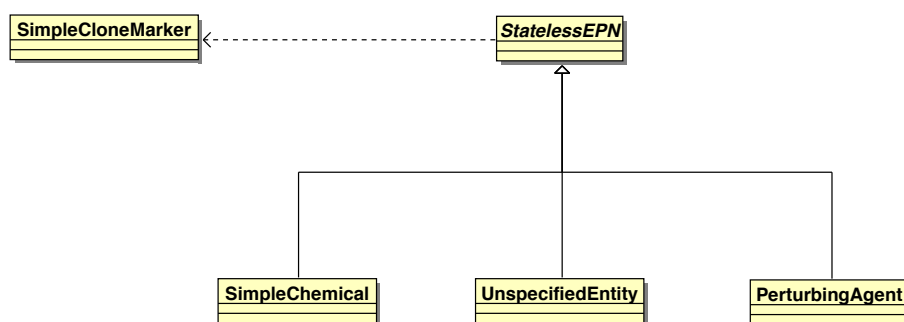
245

The StatelessEPN (figure 3.7) represents a pool where the entities do not change ‘state’. In other- words the entities do not undergo any physical change that is useful to record in a Process Descrip- tion map. Therefore these glyphs cannot be assigned a state-variable.

246

247

248

**Figure 3.7:** *UML definition of the stateless entity pool node and its descendant glyphs.***StatelessEPN**

249

SBO Term: Not defined

250

Auxiliary Items:

251

Special constraints or rules: 1. if a clone marker is used it must be of type SimpleCloneMarker.

252

Cloning: Defined by descendants.

253

3.2.7 Simple chemical

254

A simple chemical in SBGN is defined as the opposite of a macromolecule (Section 3.2.11): it is a chemical compound that is *not* formed by the covalent linking of pseudo-identical residues. Exam- ples of simple chemicals are an atom, a monoatomic ion, a salt, a radical, a solid metal, a crystal,

255

256

257

etc. The complex can be represented by a monomeric glyph (Simple chemical monomer) and a multimeric glyph (Simple chemical multimer).

Glyph: Simple chemical monomer

SBO Term: SBO:0000247 ! simple chemical

Auxiliary Items:

Special constraints or rules:

Cloning: Yes.

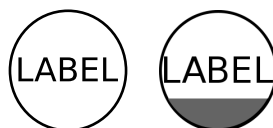


Figure 3.8: *The Process Description glyph for simple chemical.*

Container A simple chemical is represented by a circular container, as depicted in Figure 3.8. To avoid confusion with the Unspecified Entity (3.2.8), this glyph must remain a circle and cannot be deformed into an eclipse.

Label The identification of the simple chemical is carried by an unbordered box containing a string of characters. The characters may be distributed on several lines to improve readability, although this is not mandatory. The label box has to be attached to the center of the circular container. The label is permitted to spill outside the container.

Glyph: Simple chemical multimer

SBO Term: SBO:0000421 ! multimer of simple chemicals

Auxiliary Items:

Special constraints or rules:

Cloning: Yes.

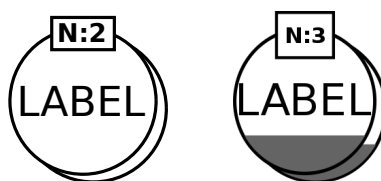


Figure 3.9: *The Process Description glyph for multimer.*

Container A simple chemical multimer is represented by two identical containers shifted horizontally and vertically and stacked one on top of the other. Figure 3.9 illustrates the glyph.

Label A multimer has no identity on its own. However, the first of the monomers carries an identifying label. The label is 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 top monomer's container. The label may spill outside of the container.

3.2.8 Unspecified entity

The simplest type of EPN is the **unspecified entity**: one whose type is unknown or simply not relevant to the purposes of the map. This arises, for example, when the existence of the entity has been inferred indirectly, or when the entity is merely a construct introduced for the needs of a map, without direct biological relevance. These are examples of situations where the *unspecified entity* glyph is appropriate. (Conversely, for cases where the identity of the entities composing the pool is known, there exist other, more specific glyphs described elsewhere in the specification.)

Glyph: Unspecified entity

SBO Term: SBO:0000285 ! material entity of unspecified nature

Auxiliary Items:

Special constraints or rules: 1. The cardinality of the unspecified entity is always 1, i.e., (there is no multimer glyph.

Cloning: Yes

Container An unspecified entity is represented by an elliptic container, as shown in 3.10. Note that this must remain an ellipse to avoid confusion with the Simple Chemical glyph, which is a circle (c.f. 3.2.7).

Label An unspecified 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 label box must be attached to the center of the container. The label may spill outside of the container.



Figure 3.10: The Process Description glyph for unspecified entity.

3.2.9 Perturbing Agent

Biochemical networks can be affected by external influences. Those influences can be the effect of well-defined physical perturbing agents, such as a light pulse or a change in temperature; they can also be more complex and not well-defined phenomena, for instance the outcome of a biological process, an experimental setup, or a mutation. For these situations, SBGN provides the **perturbing agent** glyph. It is an EPN, and represents the amount to perturbing agent applied to a process.

Glyph: Perturbing agent

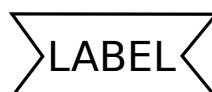


Figure 3.11: The Process Description glyph for perturbing agent.

SBO Term: SBO:0000405 ! perturbing agent

Auxiliary Items: A perturbing agent can optionally carry one or more units of information (Section 3.2.17).

Special constraints or rules: 1. A perturbing agent can optionally carry one or more units of information (Section 3.2.17). A controlled vocabulary is available to describing the physical characteristic of the perturbing agent (see Section 3.4.4).

Cloning: Yes

Container A perturbing agent is represented by a modified hexagon having two opposite concave faces, as illustrated in Figure 3.11.

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.

3.2.10 Stateful Entity Pool

Stateful entity pools can undergo physical changes, for example chemical modification or conformational change, which we wish to record in a Process Description map. This information is captured via the StateVariable, as can be seen in figure 3.12). Replicated StatefulEPNs are indicated by the Labelled Clone Marker decoration. Note that not all glyphs that are descendants of StatefulEPN can be cloned at all.

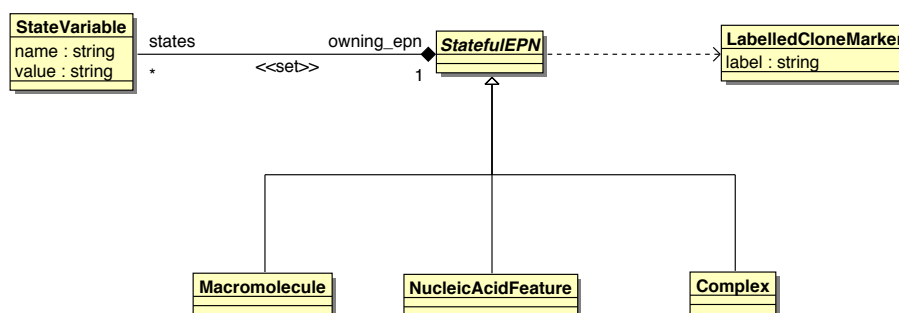


Figure 3.12: UML definition of the stateful entity pool node and its descendant glyphs and its association with state variables.

StatefulEPN

SBO Term: Not defined

Attributes:

states:StateVariable The state variables associated with this EPN.

Auxiliary Items: It can contain one or more StateVariable glyphs associated with it.

Special constraints or rules: 1. Two or more state variables with the same name are permitted. 2. State variables with no name set are permitted. 3. A LabelledCloneMaker must be used to indicate cloning. 4. StatefulEPNs that are identical and so decorated with a LabelledCloneMarker must use the same label to indicate that they are part of the same 'clone'.

Cloning:

3.2.11 Macromolecule

Many biological processes involve *macromolecules*: biochemical substances that are built up from the covalent linking of pseudo-identical units. Examples of macromolecules include proteins, nucleic acids (RNA, DNA), and polysaccharides (glycogen, cellulose, starch, etc.). Attempting to define a separate glyph for all of these different molecules would lead to an explosion of symbols in SBGN,

so instead, SBGN Process Description Level 1 defines only one glyph for all macromolecules. The same glyph is to be used for a protein, a nucleic acid, a complex sugar, and so on. The exact nature of a particular macromolecule in a map is then clarified using its label and decorations, as will become clear below. (Future levels of SBGN may subclass the macromolecule and introduce different glyphs to differentiate between types of macromolecules). This has two associated glyphs. One where

Glyph: Macromolecule monomer

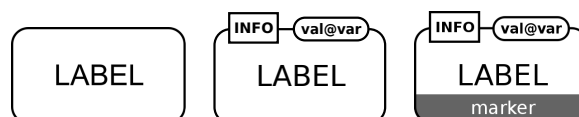


Figure 3.13: *The Process Description glyph for macromolecule, shown plain and unadorned on the left, and with with an additional state variable and a unit of information in the right and the cloned form on the right.*

SBO Term: SBO:0000245 ! macromolecule

Auxiliary Items:

Special constraints or rules: 1. cardinality must equal 1.

Cloning: Yes.

Container A macromolecule is represented by a rectangular container with rounded corners, as illustrated in Figure 3.13.

Label A macromolecule 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.

Glyph: Macromolecule multimer



Figure 3.14: *The Process Description glyph for macromolecule multimer, shown plain and unadorned on the left, and with with an additional state variable and a unit of information on the right.*

SBO Term: SBO:0000420 ! multimer of macromolecules

Auxiliary Items:

Special constraints or rules: 1. cardinality must be greater than 1.

Cloning: Yes.

Container A multimer is represented by two identical containers shifted horizontally and vertically and stacked one on top of the other. Figure 3.14 illustrates the glyph.

Label As monomer

3.2.12 Nucleic acid feature

The *Nucleic acid feature* construct in SBGN is meant to represent a fragment of a macromolecule carrying genetic information. A common use for this construct is to represent a gene or transcript. The label of this EPN and its nature are often important for making the purpose clear to the reader of a map.

Glyph: Nucleic acid feature monomer

This glyph represents a monomeric macromolecule.

SBO Term: SBO:0000354 ! informational molecule segment

Container: A nucleic acid feature is represented by a rectangular container whose bottom half has rounded corners, as shown in Figure 3.15. This design reminds that we are fundamentally dealing with a unit of information, but this information is carried by a macromolecule.

Label: The identity of a particular Nucleic acid feature is established by a label placed in an unordered box containing a string of characters. The characters may 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 nucleic acid feature can carry state variables (Section 3.2.17) that add information about its precise state. The state of a nucleic acid feature is therefore defined as the vector of all its state variables.

A nucleic acid feature can also carry one or several units of information (Section 3.2.17). These can characterize a nucleic acid feature's domain, such as a binding site, or an exon. Particular units of information carry the material type (Section 3.4.1) and the conceptual type (Section 3.4.2) of the nucleic acid feature.

Cloning: Labeled Clone Marker



Figure 3.15: The Process Description glyph for nucleic acid feature monomer, shown plain and unadorned on the left and with an additional state variable and a unit of information in the middle and the cloned form on the right.

Glyph: Nucleic acid feature multimer

This glyph represents a multimeric macromolecule.

SBO Term: SBO:0000419 ! multimer of informational molecule segments

Container: A Nucleic acid feature multimer is represented by two identical containers shifted horizontally and vertically and stacked one on top of the other. Figure 3.9 illustrates the glyph.

Label: As monomer glyph.

Auxiliary Items: A nucleic acid feature can carry state variables (Section 3.2.17) that add information about its precise state. The state of a nucleic acid feature is therefore defined as the vector of all its state variables.

A nucleic acid feature can also carry one or several units of information (Section 3.2.17). These can characterize a nucleic acid feature's domain, such as a binding site, or an exon. Particular units of information carry the material type (Section 3.4.1) and the conceptual type (Section 3.4.2) of the nucleic acid feature.

A nucleic acid feature may also carry a clone marker (Section 3.3.1).

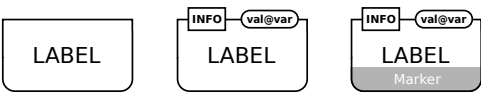


Figure 3.16: *The Process Description glyph for nucleic acid feature monomer, shown plain and unadorned on the left and with an additional state variable and a unit of information in the middle and the cloned form on the right.*

3.2.13 Complex

406

A **Complex** represents a biochemical entity composed of other biochemical entities, whether macro-
molecules, simple chemicals, multimers, or other complexes (figure 3.17). The **Complex** can de-
scribed its composition by the set of **Subunits** it contains (see figure ??). This description is entirely
optional and is their to assist the user with a visual shorthand about the composition of the complex.

407
408
409
410

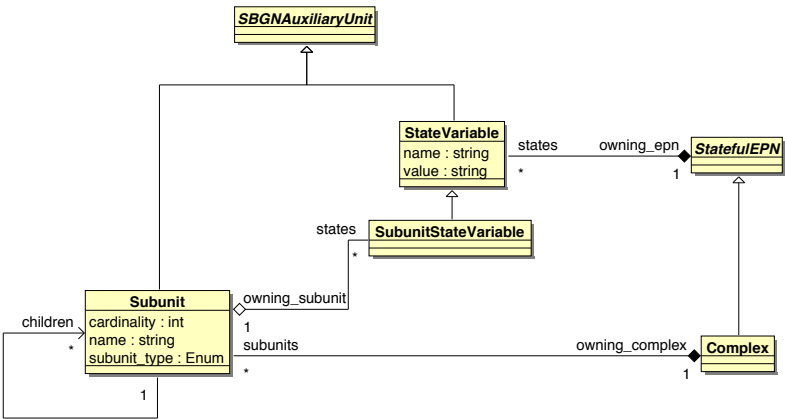


Figure 3.17: *The UML definition of the Complex and its associated subunits. In particular this describes organisation of the state variables that belong to both the subunit, but also the complex.*

Generalisations

411

- EntityPoolNode (see section ??)

412

Attributes

413

No additional attributes

414

Associations

415

subunits:Subunit (*) The subunits that describe the composition of this complex.

416

Special Rules and Constraints

417

- Once a set of subunits are defined then they are always defined.
- The set of subunits in the Complex does not identify it. One or more Complexes that contain the same set of subunits, but have different labels are **not** identical.

418

419

420

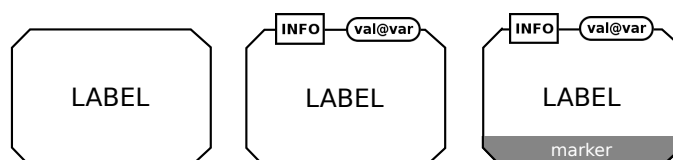
Notation

421

The **Complex** is represented by two glyphs, the **Complex Monomer** which represents a **Complex** where the cardinality is one and the **Complex Multimer** where the cardinality is greater than that.

422

423

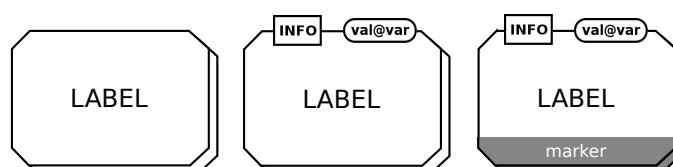
Figure 3.18: *The complex glyph.***Complex Monomer****SBO Term:** SBO:0000253 ! non-covalent complex

Auxiliary Items: A complex can carry state variables (see Section 3.2.17). The state of a complex is defined by the set of the all its state variable and all the state variables of all its components. A complex can also carry one or several units of information (see Section 3.2.17). A complex may carry a clone marker (see Section 3.3.1).

Cloning: Labeled Clone Marker

Container: A complex possesses its own container box surrounding the juxtaposed container boxes of its components. This container box is a rectangle with cut-corners (an octagonal box with sides of two different lengths). The size of the cut-corners are adjusted so that there is no overlap between the container and the components. The container boxes of the components must not overlap.

Label: The identification of a named complex is carried by an unbordered box containing a string of characters. The characters may be distributed on several lines to improve readability, although this is not mandatory. Ideally the label box should be attached to the midway between the border of the complex's container box and the border of the components' container boxes. However, if the Complex contains Subunit glyphs then the label may be positions to optimise the clarity and avoid overlapping.

Figure 3.19: *The Complex Multimer glyph.***Complex Multimer****SBO Term:** SBO:0000418 ! multimer of complexes**Auxiliary Items:** As monomer.**Cloning:** Labeled Clone Marker

Container: A Complex Multimer is represented by two identical Complex containers shifted horizontally and vertically and stacked one on top of the other. Figure 3.19 illustrates the glyph.

Label: As monomer

Examples of complex EPNs In this section, we provide examples of Entity Pool Node representations drawn using the SBGN Process Description Level 1 glyphs described above.

Figure 3.20 represents calcium/calmodulin kinase II, with phosphorylation on the sites threonine 286 and 306, as well as catalytic and autoinhibitory domains. Note the use of *units of information* and *state variables*.

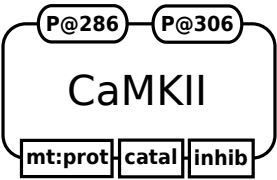


Figure 3.20: An example representation of calcium/calmodulin kinase II.

Figure 3.21 represents the glutamate receptor in the open state, with both phosphorylation and glycosylation. The entity carries two functional domains, the ligand-binding domain and the ion pore, and its chemical nature is precided.

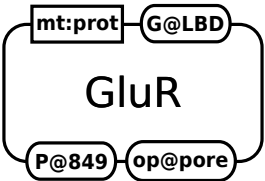


Figure 3.21: An example of a glutamate receptor in the open state.

Layout Rules and Guidelines

- The subunits inside the complex must not overlap.
- The subunits should sit above the clone marker so that they are not obscured by it.
- The label should not be obscured by subunits or obscure them.

Changes from Previous Version

- Clarified that complex must have a label and the label identifies the complex irrespective of its subunit composition.
- The label positioning does not need to be at the centre of the Complex glyph.

3.2.14 Subunits

A complex can optionally be decorated with subunit symbols (Subunit see figure 3.17) to describe the composition of the complex. The symbols available are equivalent to those used by the EPN glyphs including the complex. Therefore it is possible to describe complexes within complexes. Subunits may contain labels.

Table 3.1: Mapping between the subunit types and the glyphs used to represent it. These are essentially the EPN glyphs describe in this document.

Subunit type	Monomer Glyph	Multimer Glyph
SimpleChemical	Simple Chemical Monomer	Simple Chemical Multimer
UnspecifiedEntity	Unspecified Entity	None
PerturbingAgent	Perturbing Agent	None
Macromolecule	Macromolecule Monomer	Macromolecule Multimer
NucleicAcidFeature	Nucleic Acid Feature Monomer	Nucleic Acid Feature Multimer
Complex	Complex Monomer	Complex Multimer

Generalisation

- EntityPoolNode (see section ??)

Attributes

cardinality:int The number of copies of the subunit. 473
 name:string The name of the subunit. 474
 subunit_type:enum The type of the subunit. It can have one of the following values that correspond to 475
 the equivalent EPN class: SimpleChemical, UnspecifiedEntity, PerturbingAgent, Macromolecule,
 NucleicAcidFeature, Complex. 476

Associations

owning_complex:Complex (1) The complex that owns the subunit. 479
 states:SubunitStateVariable (*) The state variables assigned to this subunit. 480

Rules and Constraints

- Two or more state variables with the same name are permitted. 482
- State variables with no name set are permitted. 483
- Subunits can also contain subunits. There is no limit on such nesting. The namespace rules 484
 below apply. 485
- Each state variable actually belongs to the Complex. 486
- The subunit defines a namespace for its state variables, e.g. subunit “A” assigned a state vari- 487
 able “P@Ser202” and a subunit “B” assigned the same state variable can be distinguished as 488
 A:P@Ser202 and B:P@Ser202. 489

Notation

The example in figure Figure 3.22 illustrates the use of subunits in a complex shows an equivalent 491
 complex without subunits. This is an import point. For every Complex drawn with subunits it will 492
 always be possible to draw an equivalent version that does not use contains subunits. 493
 494

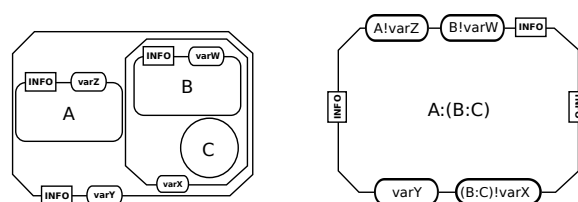


Figure 3.22: Both these complex glyphs are equivalent. The one on the left is described using sub-
 unit decorators, the one on the right describes the same thing without them.

Changes from Previous Version**3.2.15 Process**

Process nodes represent processes that transform one or several entity pools into one or several 497
 entity pools, identical or different. Note that a process may be used to represent or summarise 498
 more than one known process. SBGN Process Description Level 1 defines a generic process (Sec- 499
 tion 3.2.15), as well as five more specific ones: the omitted process (Section 3.2.15), the uncertain 500
 process (Section 3.2.15), the association (Section 3.2.15), the dissociation (Section 3.2.15), and 501
 the phenotype (Section 3.2.15). 502

Attributes:

- Instance - all entities are unique 503

glyphRules A process cannot be “orphaned”. It must always be connected to another node by one or more arcs.

Non-stoichiometric process

Process that does not necessarily result in measurable change of entity pools, and which does necessarily have a defined start and end point.

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 defines the **phenotype glyph**.

SBO Term: SBO:0000358 ! phenotype

Container: A phenotype is represented by an elongated hexagon, as illustrated in Figure 3.23.

Label: A phenotype 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.

Auxiliary Items: A phenotype may carry a **clone marker** (Section 3.3.1).



Figure 3.23: *The Process Description glyph for phenotype.*

Glyph: Submap A submap is used to encapsulate processes (including all types of nodes and edges) within one glyph. The submap hides its content to the users, and display only input terminals (or ports), linked to EPNs (Section 3.2.4). A submap is not equivalent to an **omitted process** (see Section 3.2.15). In the case of an SBGN description that is made available through a software tool, the content of a submap may be available to the tool. A user could then ask the tool to expand the **submap**, for instance by clicking on the icon representing the **submap**. The tool might then expand and show the **submap** within the same map (on the same canvas), or it might open it in a different canvas. In the case of an SBGN description made available in a book or a website, the content of the **submap** may be available on another page, possibly accessible via an hyperlink on the **submap**.

SBO Term: SBO:0000395 ! encapsulating process

Container: The submap is represented as a square box to remind the viewer that it is fundamentally a process.

Label: The identification of the submap is carried by an unbordered box containing a string of characters. The characters may be distributed on several lines to improve readability, although this is not mandatory. The label box has to be attached to the center of the container box.

Auxiliary Items: A submap carries labeled terminals. When the submap is represented folded, those terminals are linked to external EPNs (Section 3.2.4). In the unfolded view, exposing the internal structure of the submap, a set of tags point to the corresponding internal EPNs Section 3.2.4.

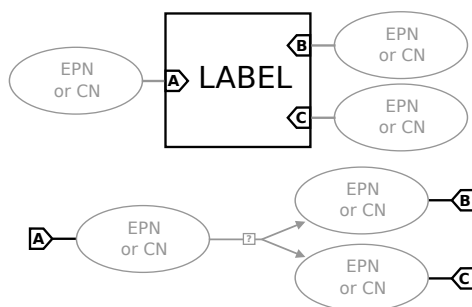


Figure 3.24: The Process Description glyph for submap. (Upper part) folded submap. (Lower part) content of the submap.

Figure 3.25 represents a submap that transforms glucose into fructose-6-phosphate. The submap carries five terminals, four linked to EPNs and one linked to a compartment. The latter is particularly important in the case of EPNs present only in a compartment enclosed in a submap, and that are not linked to terminals themselves. Note that the terminals do not define a “direction”, such as input or output. The flux of the reactions is determined by the context.

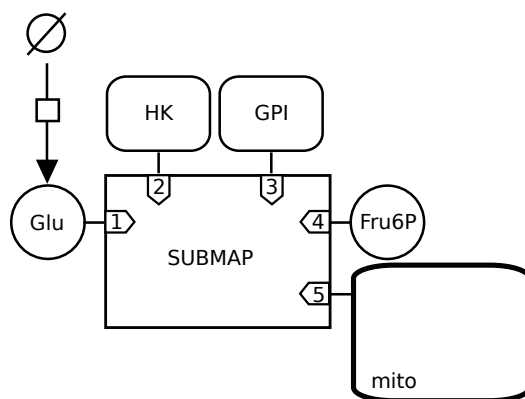


Figure 3.25: Example of a submap with contents elided.

The map in Figure 3.26 represents an unfolded version of a submap. Here, anything outside the submap has disappeared, and the internal tags are not linked to the corresponding external terminals. Note the tag 5, linking the compartment “mito” of the submap to the compartment “mito” outside the submap. The compartment containing Glu6P is implicitly defined as the same as the compartment containing Glu and Fru6P. There is no ambiguity because if Glu and Fru6P were in different compartments, one of them should have been defined within the submap.

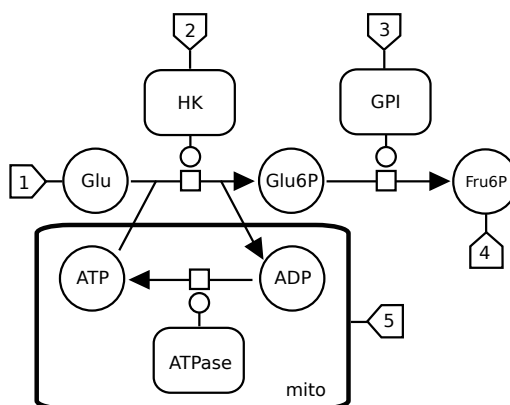


Figure 3.26: Example of an unfolded submap. The unfolded submap corresponds to the folded submap of Figure 3.25.

Logical operators

Glyph: And The glyph and is used to denote that all the EPNs linked as input are necessary to produce the output.

SBO Term: SBO:0000173 ! and.

Origin: More than one EPN (section 3.2.4) or logical operator (section 3.2.15).

Target: One modulation (section 3.2.19), stimulation (section 3.2.19), catalysis (section 3.2.19), inhibition (section 3.2.19) or necessary stimulation (section 3.2.19) arc.

Node: And is represented by a circle carrying the word “AND”.

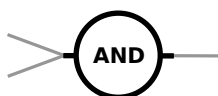


Figure 3.27: The Process Description glyph for and. Only two inputs are represented, but more would be allowed.

Glyph: Or The glyph or is used to denote that any of the EPNs linked as input is sufficient to produce the output.

SBO Term: SBO:0000174 ! or.

Origin: More than one EPN (section 3.2.4) or logical operator (section 3.2.15).

Target: One modulation (section 3.2.19), stimulation (section 3.2.19), catalysis (section 3.2.19), inhibition (section 3.2.19) or necessary stimulation (section 3.2.19) arc.

Node: Or is represented by a circle carrying the word “OR”.

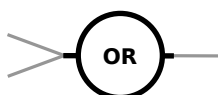


Figure 3.28: The Process Description glyph for or. Only two inputs are represented, but more would be allowed.

Glyph: Not The glyph not is used to denote that the EPN linked as input cannot produce the output.

SBO Term: SBO:0000238 ! not.

Origin: One EPN (section 3.2.4) or logical operator (section 3.2.15).

Target: One modulation (section 3.2.19), stimulation (section 3.2.19), catalysis (section 3.2.19), inhibition (section 3.2.19) or necessary stimulation (section 3.2.19) arc. 569
570
Node: Not is represented by a circle carrying the word “NOT”. 571

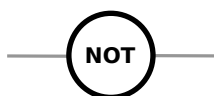


Figure 3.29: The Process Description glyph for not.

Stoichiometric process 572

A process with measurable change in the quantities of entity pools consumed and produced. 573

Glyph: Process A generic stoichiometric process that transforms a set of entity pools (represented by EPNs in SBGN Process Description Level 1) into another set of entity pools. 574
575

SBO Term: SBO:0000375 ! process 576

Origin: One or several consumption arcs (Section 3.2.19) or one or several production arcs (Section 3.2.19). 577
578

Target: One or several production arcs (Section 3.2.19). 579

Node: A process is represented by a square box linked to two connectors, small arcs attached to the centers of opposite sides. The consumption (Section 3.2.19) and production (Section 3.2.19) arcs are linked to the extremities of those connectors. The modulatory arcs (Section 3.2.19) point to the other two sides of the box. A process connected to production arcs on opposite sides is a reversible process. 580
581
582
583
584

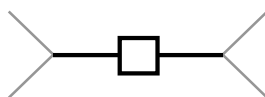


Figure 3.30: The Process Description glyph for process.

A process is the basic process node in SBGN. It describes a process that transforms a given set of biochemical entities—macromolecules, simple chemicals or unspecified entities—into another set of biochemical entities. Such a transformation might imply modification of covalent bonds (conversion), modification of the relative position of constituents (conformational process) or movement from one compartment to another (translocation). 585
586
587
588
589

A cardinality label may be associated with consumption (Section 3.2.19) or production (Section 3.2.19) arcs to indicate the stoichiometry of the process. This label becomes a requirement when the exact composition of the number of copies of the inputs or outputs to a reaction are ambiguous in the map. 590
591
592
593

A process is regarded as reversible if both ‘sides’ of the process are connected to production arcs (see section 4.2.2). 594
595

The example in Figure 3.31 illustrates the use of a process node to represent the phosphorylation of a protein in a Process Description. 596
597

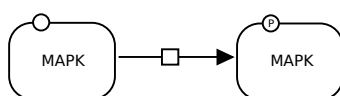


Figure 3.31: Phosphorylation of the protein MAP kinase.

The example in Figure 3.32 illustrates the use of a process node to represent a reaction between two reactants that generates three products.

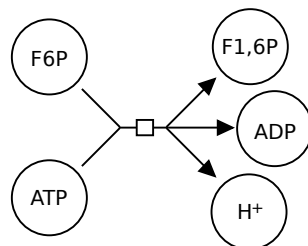


Figure 3.32: Reaction between ATP and fructose-6-phosphate to produce fructose-1,6-biphosphate, ADP and a proton.

The example in Figure 3.33 illustrates the use of a process node to represent a translocation. The large round-cornered rectangle represents a compartment border (see Section 3.2.16).

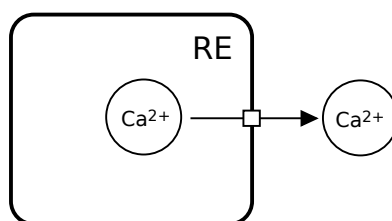


Figure 3.33: Translocation of calcium ion out of the endoplasmic reticulum. Note that the process does not have to be located on the boundary of the compartment. A process is not attached to any compartment.

The example in Figure 3.34 illustrates the use of a process node to represent the reversible opening and closing of an ionic channel in a Process Description.

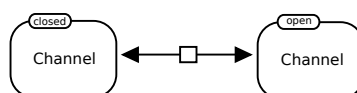


Figure 3.34: Reversible opening and closing of an ionic channel.

When such a reversible process is asymmetrically modulated, it must be represented by two different processes in a Process Description. Figure 3.35 illustrates the use of two process nodes to represent the reversible activation of a G-protein coupled receptor. In the absence of any effector, an equilibrium exists between the inactive and active forms. The agonist stabilises the active form, while the inverse agonist stabilises the inactive form.

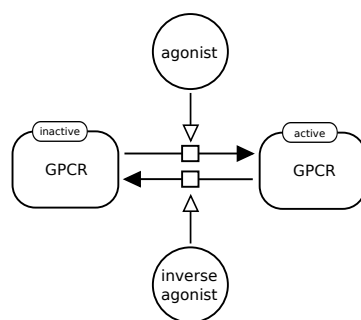


Figure 3.35: *The reversible activation of a G-protein coupled receptor.*

The example in Figure 3.36 presents the conversion of two galactoses into a lactose. Galactoses are represented by only one simple chemical, the cardinality being carried by the consumption arc.

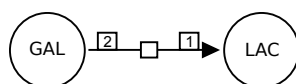


Figure 3.36: *Conversion of two galactoses into a lactose.*

Glyph: Omitted process Omitted processes are processes that are known to exist, but are omitted from the map for the sake of clarity or parsimony. A single omitted process can represent any number of actual processes. The omitted process is different from a submap. While a submap references to an explicit content, that is hidden in the main map, the omitted process does not “hide” anything within the context of the map, and cannot be “unfolded”.

SBO Term: SBO:0000397 - omitted process.

Origin: One or several consumption arcs (Section 3.2.19) or one or several production arcs (Section 3.2.19).

Target: One or several production arcs (Section 3.2.19).

Node: An omitted process is represented by a process in which the square box contains a two parallel slanted lines oriented northwest-to-southeast and separated by an empty space.

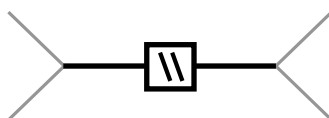


Figure 3.37: *The Process Description glyph for omitted process.*

Glyph: Uncertain process Uncertain processes are processes that may not exist. A single uncertain process can represent any number of actual processes.

SBO Term: SBO:0000396 ! uncertain process.

Origin: One or several consumption arcs (Section 3.2.19) or one or several production arcs (Section 3.2.19).

Target: One or several production arcs (Section 3.2.19).

Node: An uncertain process is represented by a process which square box contains a question mark.

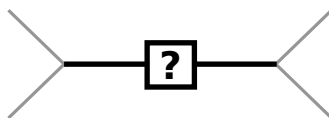


Figure 3.38: *The Process Description glyph for an uncertain process.*

Glyph: Association The association between one or more EPNs represents the non-covalent binding of the biological objects represented by those EPNs into a larger complex.

SBO Term: SBO:0000177 ! non-covalent binding.

Origin: One or more consumption arcs (Section 3.2.19).

Target: One production arc (Section 3.2.19).

Node: An association between several entities is represented by a filled disc linked to two connectors, small arcs attached on point separated by 180 degrees. The consumption (Section 3.2.19) and production (Section 3.2.19) arcs are linked to the extremities of those connectors.

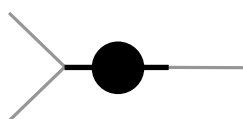


Figure 3.39: *The Process Description glyph for association.*

The example in Figure 3.40 illustrates the association of cyclin and CDC2 kinase into the Maturation Promoting Factor.

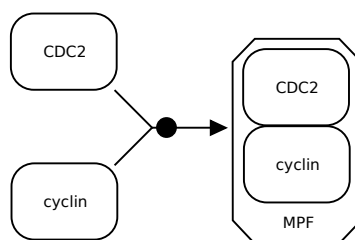


Figure 3.40: *Association of cyclin and CDC2 kinase into the Maturation Promoting Factor.*

Figure 3.41 gives an example illustrating the association of a pentameric macromolecule (a nicotinic acetylcholine receptor) with a simple chemical (the local anesthetic chlorpromazin) in an unnamed complex.

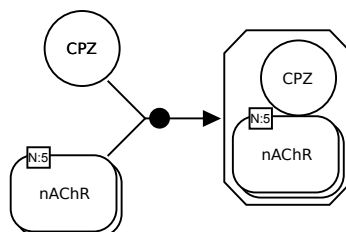


Figure 3.41: *The association of a pentameric macromolecule with a simple chemical in an unnamed complex.*

An association does not necessarily result in the formation of a complex; it can also produce a multimer, or a macromolecule (although the latter case is semantically borderline). 3.42 gives an example of this, using the formation of hemoglobin.

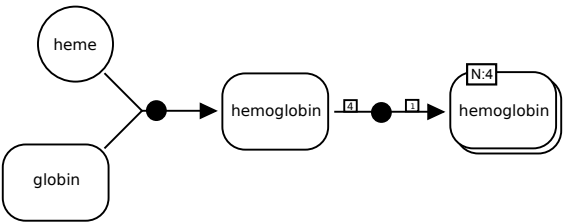


Figure 3.42: Formation of hemoglobin.

Glyph: Dissociation The dissociation of an EPN into one or more EPNs represents the rupture of a non-covalent binding between the biological entities represented by those EPNs.

SBO Term: SBO:0000180 ! dissociation.

Origin: One consumption arc (Section 3.2.19).

Target: One or more production arc (Section 3.2.19).

Node: A dissociation between several entities is represented by two concentric circles. A simple empty disc could be, in some cases, confused with the catalysis (section Section 3.2.19). Moreover, the existence of two circles reminds the dissociation, by contrast with the filled disc of the association (Section 3.2.15).

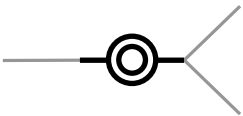


Figure 3.43: The Process Description glyph for dissociation.

The example in Figure 3.44 illustrates the dissociation of the small and large ribosomal subunits from a messenger RNA.

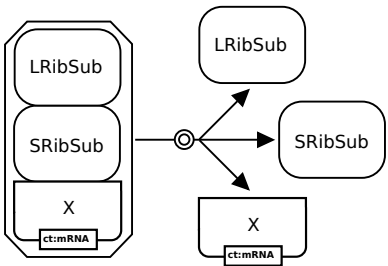


Figure 3.44: Dissociation of the small and large ribosomal subunits from a messenger RNA.

- 3.2.16 Container
- Canvas
- Identifying Attributes:
- identity
- Special constraints or rules:

If one **Compartment** is drawn on the canvas then all EPNs must be drawn in a **Compartment** and cannot be drawn directly onto the Canvas¹.

SBO Term: Not assigned

Container: None

Label: None

Auxiliary Items: The Canvas can have **Container** glyphs and any EPN, Process and Arcs as children. However all Auxiliary glyphs cannot be children of the Canvas.

Cloning: N/A

Glyph: Compartment

A compartment is a logical or physical structure that contains entity pool nodes. An EPN can only belong to one compartment. Therefore, the “same” biochemical species located in two different compartments are in fact two different pools.

SBO Term: SBO:0000290 ! physical compartment

Container: A compartment is represented by a surface enclosed in a continuous border or located between continuous borders. These borders should be noticeably thicker than the borders of the EPNs. A compartment can take **any** geometry. A compartment must always be entirely enclosed.

Label: The identification of the compartment is carried by 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 can be attached anywhere in the container box. Note that the label can spill-over from the container box.

Auxiliary Items: A compartment can carry a certain number of units of information, that will add information for instance about the physical environment, such as pH, temperature or voltage, see Section 3.2.17. The center of the bounding box of a unit of information is located on the mid-line of the border of the compartment.

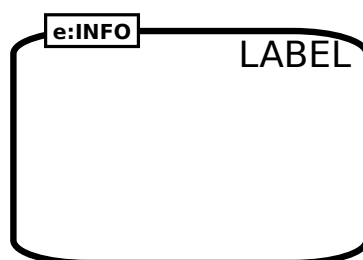


Figure 3.45: *The Process Description glyph for compartment.*

To allow more aesthetically pleasing and understandable maps, compartments are allowed to overlap each other visually, but it must be kept in mind that this does not mean the top compartment contains part of the bottom compartment. Figure 3.46 shows two semantically equivalent placement of compartments:

¹All EPNs must belong to a compartment. When no compartment is drawn then all EPNs can be thought of as belonging to an imaginary “default” compartment. But when one or more EPNs are assigned to an explicit compartment then for clarity all compartments must be explicitly drawn and EPNs assigned to them.

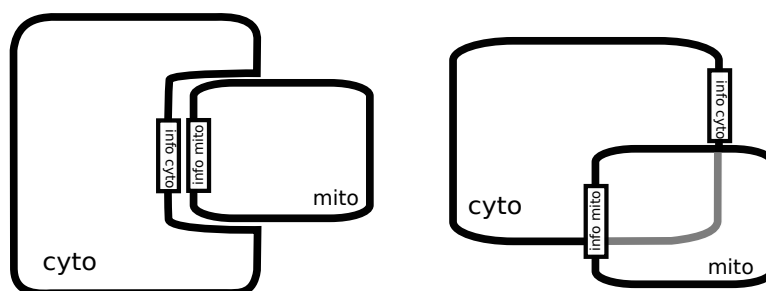


Figure 3.46: *Overlapped compartments are permitted, but the overlap does not imply containment.*

Overlapped (hidden) part of the compartment should not contain any object which could be covered by an overlapping compartment. Figure 3.47 illustrates the problem using an incorrect map.

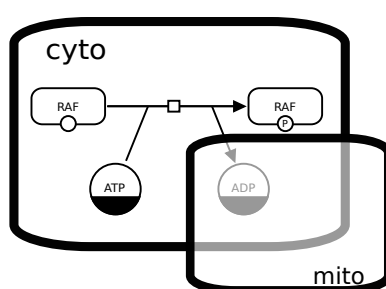


Figure 3.47: *Example of an incorrect map. Overlapped compartments must not obscure other objects.*

3.2.17 Auxiliary units

Auxiliary units are glyphs that decorate other glyphs, providing additional information that may be useful to the reader. These can provide annotation (unit of information), state information (state variable) or indicate duplication of entity pool nodes (clone marker).

Glyph: Unit of information

When representing biological entities, it is often necessary to convey some abstract information about the entity's function that cannot (or does not need to) be easily related to its structure. The unit of information is a decoration that can be used in this situation to add information to a glyph. Some example uses include: characterizing a logical part of an entity such as a functional domain (a binding domain, a catalytic site, a promoter, etc.), or the information encoded in the entity (an exon, an open reading frame, etc.). A unit of information can also convey 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 EPN being annotated by the unit of information. The center of the bounding box of a state of information should be located on the midline of the border of the EPN.

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. For certain prede-

defined types of information having controlled vocabularies associated with them, SBGN defines specific prefixes that must be included in the label to indicate the type of information in question. The controlled vocabularies predefined in SBGN Process Description Level 1 are described in Section 3.4 and summarized in the following list:

pc container physical characteristic
 mt entity pool material type
 ct entity pool conceptual type
 N multimer cardinality

Auxiliary Items: A unit of information does not carry any auxiliary items.

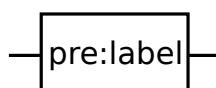


Figure 3.48: *The Process Description glyph for unit of information.*

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 whole entity (i.e., the SBGN object) can be described by the current values of all its state variables, and the values of the state variables of all its possible components, recursively.

Identifying Attributes:

- owning EPN
- name (optional)
- value

Special constraints or rules:

None.

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 3.49. The parallel segment axis should be tangent to the border of the glyph of the EPN 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 EPN.

Label: The identification of an instance of a state variable is carried by one or two unbordered boxes, each containing a string of characters. The characters cannot be distributed on several lines. One box is mandatory, and contains the value of the state variable. The value may be empty; an example of a situation where this might arise is an unphosphorylated phosphorylation site. The second box is optional and carries the identification of the state variable. The center of the combination of the boxes located in the container box is superposed to the center of this container box. In earlier version of the Process Description specification, the identification of the state variable could be located outside the state variable container box. This is now forbidden. The style of labeling of state variables encouraged by SBGN Process

Description Level 1 is to combine a prefix representing the value of the variable with a suffix representing the variable's name. Prefix and suffix should be separated by the symbol '@', X@Y thus meaning *value X AT variable Y*. The label of a `state variable` should, if possible, be displayed within the boundary of the glyph.

Auxiliary Items: A `state variable` does not carry any auxiliary items.

Cloning: Not cloned.

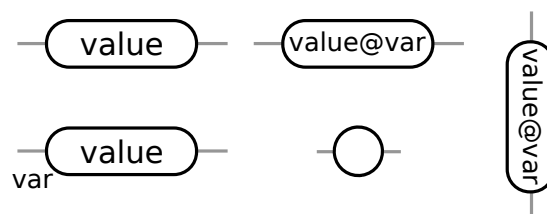


Figure 3.49: Examples of the Process Description glyph for `state variable`.

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. However, in all cases, a `state variable` on an EPN can only take *one* defined value. Further, an EPN’s `state variable` should always be displayed and always set to a value. An “empty” `state variable` is a `state variable` that is set to the value “unset”, it is not a `state variable` with no value. Note that the value “unset” is *not* synonymous to “any value” or “unknown value”.

Glyph: Annotation

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

SBO Term: SBO:NEW

Container: An annotation is represented by a rectangular container with a folded corner, as illustrated in Figure 3.50. This container is linked to the annotated element in a way that cannot be mistaken for a relationship, for instance a callout, a thick edge, a dashed line etc. 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 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 3.50: The Process Description glyph for annotation.

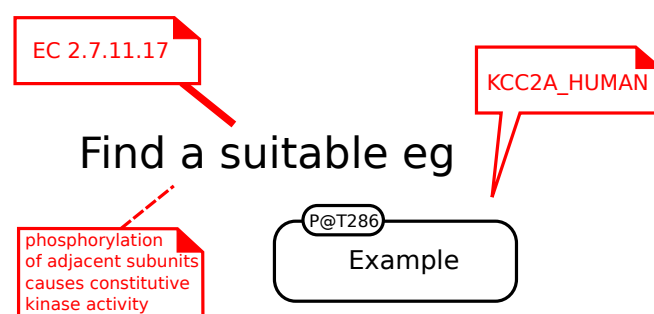


Figure 3.51: Example of annotations adding information to the description of the trans-phosphorylation of CaMKII. Note that three different types of links are used between annotation nodes and annotated elements. However, it is recommended to use a consistent scheme within a map.

3.2.18 Reference nodes

Reference nodes handle links or relationships between elements of a map and sub-map. At present there is only one reference glyph, tag, which can be used in a map referred to by a submap (Section 3.2.15) or as an auxiliary unit on the submap. The clone marker can also provide additional reference mechanisms and is discussed below (Section 3.3.1).

Glyph: Tag

A tag is a named handle, or reference, to another EPN (Section 3.2.4) or compartment (Section 3.2.16). Tags are used to identify those elements in submaps (Section 3.2.15).

SBO Term: Not applicable.

Container: A tag is represented by a rectangle fused to an empty arrowhead, as illustrated in Figure 3.52. The symbol should be linked to one and only one edge (i.e., it should reference only one EPN or compartment).

Label: A tag 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 tag does not carry any auxiliary items.



Figure 3.52: The Process Description glyph for tag.

3.2.19 Arcs

Arcs are lines that link EPNs and PNs together. The symbols attached to their extremities indicate their semantics.

Process Arcs

Glyph: Consumption Consumption is the arc used to represent the fact that an entity pool is consumed by a process, but is not produced by the process.

SBO Term: SBO:0000394 ! consumption.

Origin: Any EPN (Section 3.2.4).

Target: Any process node (Section 3.2.15).

End point: No particular symbol is used to represent a consumption.

A cardinality label may be associated with consumption (Section 3.2.19) or production (Section 3.2.19) arcs, indicating the stoichiometry of a process. This label is a number enclosed in a rectangle with one of the long sides adjacent to the consumption arc. The cardinality is required to eliminate ambiguity when the exact composition, or the number of copies, of the inputs or outputs to a reaction are ambiguous from the map. An example is a multimer of 6 subunits dissociating into 2 monomers and 2 dimers. Without stoichiometry labels another result, such as 4 monomers and 1 dimer could be inferred. Once assigned to one arc connecting to a process node, cardinality should be represented on all consumption and production arcs connected to that process node to avoid misinterpretation.

Omitted cardinality on one edge only should not be treated as cardinality of 1, but as an unspecified cardinality. In most cases, the exact value may be derived from the context, but unless cardinality is explicitly shown, it should be considered as unspecified. In the case where the stoichiometry of some part of the process is not known, or undefined, a question mark (?) should be used within the cardinality label of the corresponding arcs.



Figure 3.53: *The Process Description glyph for consumption.*

Glyph: Production Production is the arc used to represent the fact that an entity pool is produced by a process. In the case of a reversible process, the production arc also acts as a consumption arc.

SBO Term: SBO:0000393 ! production.

Origin: Any process node (Section 3.2.15).

Target: Any EPN (Section 3.2.4).

End point: The target extremity of a production carries a filled arrowhead.

A cardinality label can be associated with a production arc indicating the stoichiometry of a process.

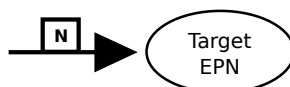


Figure 3.54: *The Process Description glyph for production.*

Figure 3.55 illustrates the use of consumption/production arc cardinality labels to represent the stoichiometry of a process.

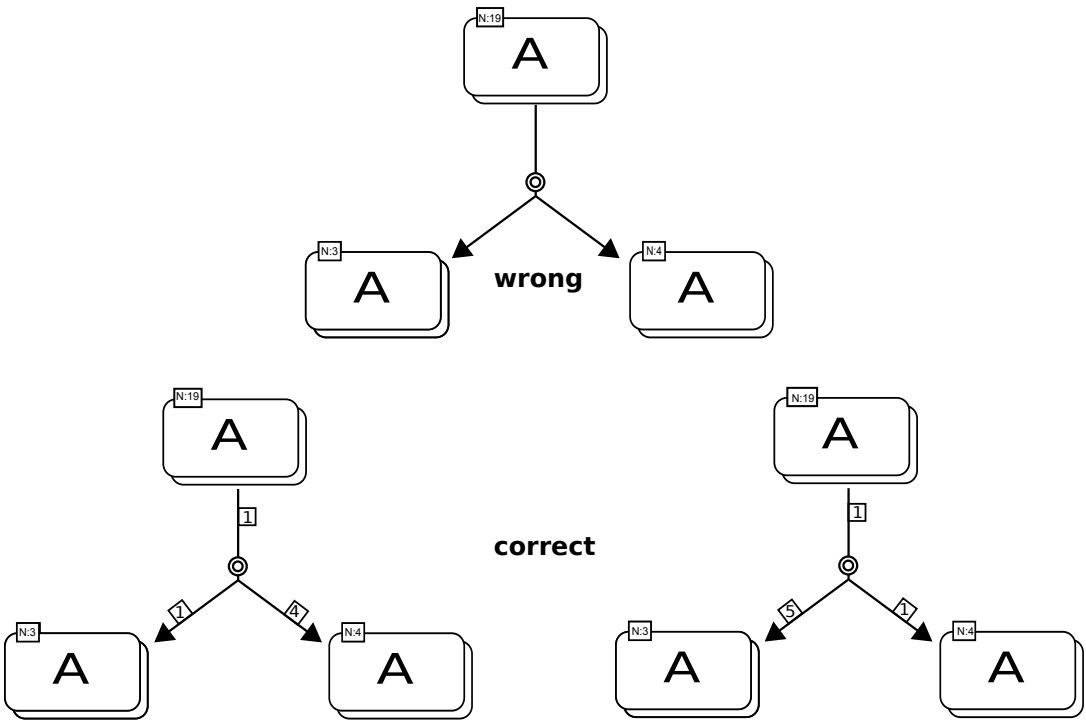


Figure 3.55: Cardinality for production arcs.

Modulation Arcs

837

Glyph: Modulation A modulation affects the flux of a process represented by the target process. Such a modulation can affect the process **positively or negatively**, or even both ways depending on the conditions, for instance the concentration of the intervening participants. A modulation can also be used when one does not know the precise direction of the effect.

838

839

840

841

SBO Term: SBO:0000168 ! control.

842

Origin: Any EPN (Section 3.2.4) or any logical operator (Section 3.2.15).

843

Target: Any process node (Section 3.2.15).

844

End point: The target extremity of a modulation carries an empty diamond.

845



Figure 3.56: The Process Description glyph for modulation.

Figure 3.57 represents the effect of nicotine on the process between closed and open states of a nicotinic acetylcholine receptor. High concentrations of nicotine open the receptor while low concentrations can desensitize it without opening.

846

847

848

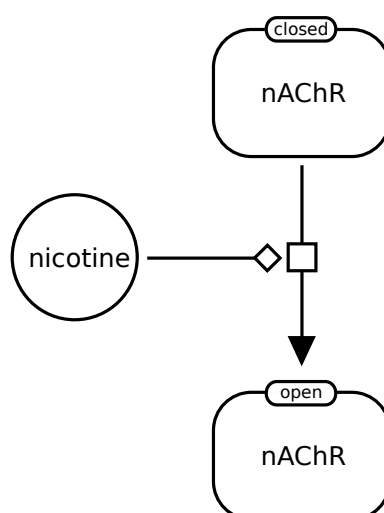


Figure 3.57: *Modulation of nicotinic receptor opening by nicotine.*

Glyph: Stimulation A stimulation affects **positively** the flux of a process represented by the target process. This stimulation can be for instance a catalysis or a positive allosteric regulation. Note that catalysis exists independently in SBGN, see Section 3.2.19.

SBO Term: SBO:0000170 ! stimulation.

Origin: Any EPN (Section 3.2.4) or any logical operator (Section 3.2.15).

Target: Any process node (Section 3.2.15).

End point: The target extremity of a stimulation carries an empty arrowhead.

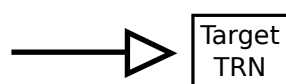


Figure 3.58: *The Process Description glyph for stimulation.*

Glyph: Catalysis A catalysis is a particular case of stimulation, where the effector affects positively the flux of a process represented by the target process. The positive effect on the process is due to the lowering of the activation energy of a reaction.

SBO Term: SBO:0000172 ! catalysis.

Origin: Any EPN (Section 3.2.4) or any logical operator (Section 3.2.15).

Target: Any process node (Section 3.2.15).

Node: The target extremity of a catalysis carries an empty circle.

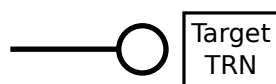


Figure 3.59: *The Process Description glyph for catalysis.*

Glyph: Inhibition An inhibition **negatively** affects the flux of a process represented by the target process. This inhibition can be for instance a competitive inhibition or an allosteric inhibition.

SBO Term: SBO:0000169 ! inhibition.

Origin: Any EPN (Section 3.2.4) or any logical operator (Section 3.2.15).

Target: Any process node (Section 3.2.15).

Node: The target extremity of an inhibition carries a bar perpendicular to the arc.

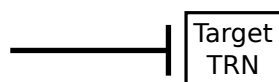


Figure 3.60: *The Process Description glyph for inhibition.*

Glyph: Necessary stimulation A necessary stimulation, is one that is necessary for a process to take place. A process modulated by a necessary stimulation can only occur when this necessary stimulation is active.

SBO Term: SBO:0000171 ! necessary stimulation.

Origin: Any EPN (Section 3.2.4) or any logical operator (Section 3.2.15).

Target: Any process node (Section 3.2.15).

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

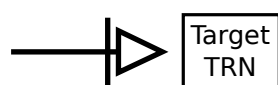


Figure 3.61: *The Process Description glyph for Necessary Stimulation.*

The example in Figure 3.62 below describes the transcription of a gene X, that is the creation of a messenger RNA X triggered by the gene X. The creation of the protein X is then triggered by the mRNA X. (Note that the same example could be represented using the gene as reactant and product, although it is semantically different.)

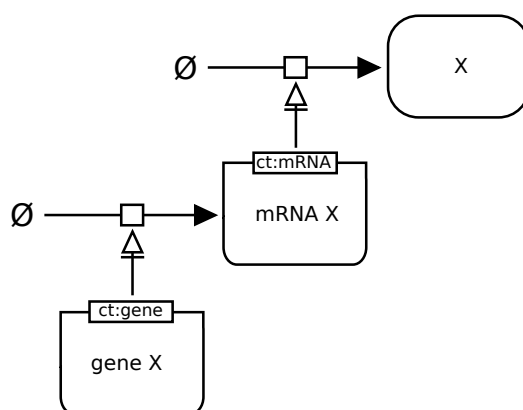


Figure 3.62: *The creation of a messenger RNA X triggered by the gene X.*

The example in Figure 3.63 below describes the transport of calcium ions out of the endoplasmic reticulum. Without IP3 receptor, there is not calcium flux, therefore, one cannot use a stimulation. The Necessary Stimulation instead represents this absolute stimulation.

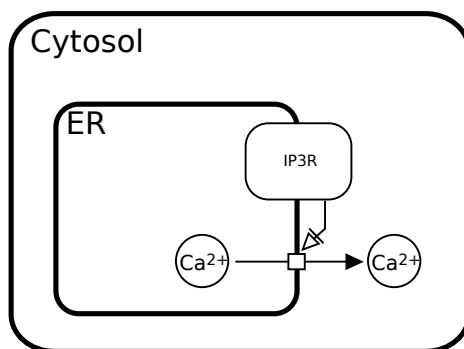


Figure 3.63: The transport of calcium ions out of the endoplasmic reticulum into the cytosol. Note that IP3R crosses both compartment boundaries. This is allowed, but the Macromolecule should only belong to one of the compartments see section C.1 for more discussion of this issue.

Others

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

SBO Term: SBO:0000398 ! logical relationship.

Origin: Any EPN (Section 3.2.4) or logical operator (Section 3.2.15).

Target: Any logical operator (Section 3.2.15).

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

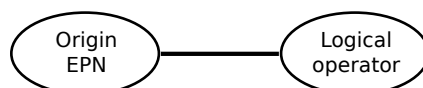


Figure 3.64: The Process Description glyph for logic arc.

Glyph: Equivalence arc Equivalence arc is the arc used to represent the fact that all entities marked by a tag are equivalent.

SBO Term: Not applicable.

Origin: Any EPN (Section 3.2.4).

Target: Tag (Section 3.2.18).

End point: No particular symbol is used to represent an equivalence arc.

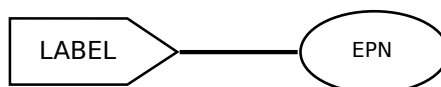


Figure 3.65: The Process Description glyph for Equivalence arc.

3.3 Decorators

3.3.1 Glyph: Clone marker

If an EPN is duplicated on a map, it is necessary to indicate this fact by using the clone marker auxiliary unit. The purpose of this marker is to provide the reader with a visual indication that this node has been cloned, and that at least one other occurrence of the EPN can be found in the map (or in a submap; see Section 3.2.15). The clone marker takes two forms, simple and labeled, depending on whether the node being cloned can carry state variables (i.e., whether it is a stateful EPN). Note that an EPN belongs to a single compartment. If two glyphs labelled “X” are located in two different

compartments, such as ATP in cytosol and ATP in mitochondrial lumen, they represent different EPNs, and therefore do not need to be marked as cloned.

Simple clone marker

As mentioned above, the simple clone marker is the unlabeled version of the clone marker. See below for the labeled version.

SBO Term: Not applicable.

Container: The simple (unlabeled) clone marker is a portion of the surface of an EPN that has been modified visually through the use of a different shade, texture, or color. Figure 3.66 illustrates this. The clone marker occupies the lower part of the EPN. The filled area must be smaller than the unfilled one.

Label: Not applicable.

Auxiliary Items: A clone marker does not carry any auxiliary items.

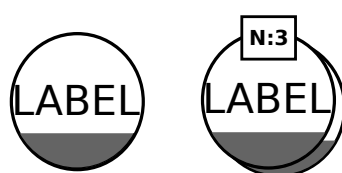


Figure 3.66: The Process Description glyph for simple clone marker applied to a simple chemical and a multimer of simple chemicals.

Labeled clone marker

Unlike the simple clone marker, the labeled clone marker includes (unsurprisingly, given its name) an identifying label that can be used to identify equivalent clones elsewhere in the map. This is particularly useful for stateful EPNs, because these can have a large number of state variables displayed and therefore may be difficult to visually identify as being identical.

SBO Term: Not applicable.

Container: The labeled clone marker is a portion of the surface of an EPN that has been modified visually through the use of a different shade, texture, or color. The clone marker occupies the lower part of the EPN glyph. The filled area must be smaller than the unfilled one, but the be large enough to have a height larger than the clone marker's label (cf below).

Label: A clone marker 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 portion of the surface of the EPN that has been modified visually). The font color of the label and the color of the clone marker should contrast with one another. The label on a labeled clone marker is mandatory.

Auxiliary Items: A clone marker does not carry any auxiliary items.

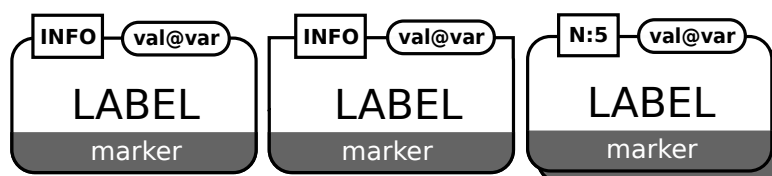


Figure 3.67: The Process Description glyph for labeled clone marker applied to a macromolecule, a nucleic acid feature and a multimer of macromolecules.

Figure 3.68 contains an example in which we illustrate the use of clone markers to clone the species ATP and ADP participating in different reactions. This example also demonstrates the chief drawbacks of using clones: it leads to a kind of dissociation of the overall network and multiplies the number of nodes required, requiring more work on the part of the reader to interpret the result. Sometimes these disadvantages are offset in larger maps by a reduction in the overall number of line crossings, but not always. In general, we advise that cloning should be used sparingly.

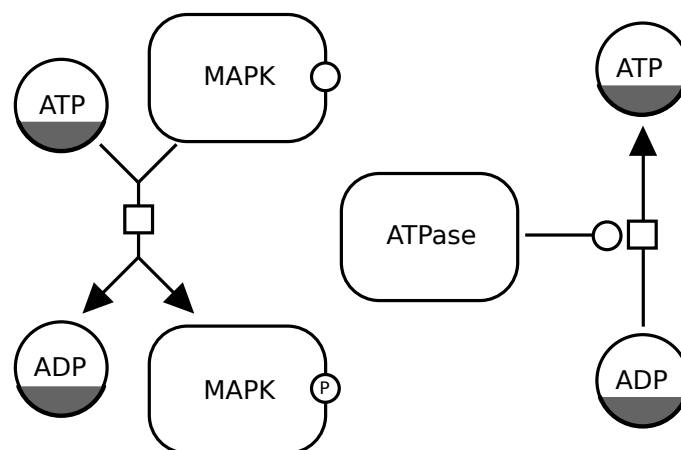


Figure 3.68: An example of using cloning, here for the species ATP and ADP.

3.4 Controlled vocabularies

Some glyphs in SBGN Process Descriptions can contain particular kinds of textual annotations conveying information relevant to the purpose of the glyph. These annotations are units of information (Section 3.2.17) or state variable (Section 3.2.17). An example is in the case of multimers, which can have a unit of information conveying the number of monomers composing the multimer. Other cases are described throughout the rest of this chapter.

The text that appears as the unit of information decorating an Entity Pool Node (EPN) must in most cases be prefixed with a controlled vocabulary term indicating the type of information being expressed. The prefixes are mandatory except in the case of macromolecule covalent modifications (Section 3.4.3). 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.

In the rest of this section, we describe the controlled vocabularies (CVs) used in SBGN Process Description Level 1. They cover the following categories of information: an EPN's material type, an EPN's conceptual type, covalent modifications on macromolecules, the physical characteristics of compartments, and cardinality (e.g., of multimers). 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 term's meanings in a figure legend or other text accompanying the map.

3.4.1 Entity pool node material types

The material type of an EPN indicates its chemical structure. A list of common material types is shown in Table 3.2, 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:0000240 (*material entity* under *entity*). The labels are defined by SBGN Process Description 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 roles. For example,

Name	Label	SBO term
Non-macromolecular ion	mt:ion	SB0:0000327
Non-macromolecular radical	mt:rad	SB0:0000328
Ribonucleic acid	mt:rna	SB0:0000250
Deoxribonucleic acid	mt:dna	SB0:0000251
Protein	mt:prot	SB0:0000297
Polysaccharide	mt:psac	SB0:0000249

Table 3.2: A sample of values from the material types controlled vocabulary (Section 3.4.1).

a strand of RNA is a physical artifact, but its use as messenger RNA is a role.

3.4.2 Entity pool node conceptual types

An EPN's *conceptual type* indicates its function within the context of a given Process Description. A list of common conceptual types is shown in Table 3.3, 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 SB0:0000241 (*conceptual entity* under *entity*). The labels are defined by SBGN Process Description Level 1.

Name	Label	SBO term
Gene	ct:gene	SB0:0000243
Transcription start site	ct:tss	SB0:0000329
Gene coding region	ct:coding	SB0:0000335
Gene regulatory region	ct:grr	SB0:0000369
Messenger RNA	ct:mRNA	SB0:0000278

Table 3.3: A sample of values from the conceptual types vocabulary (Section 3.4.2).

3.4.3 Macromolecule covalent modifications

A common reason for the introduction of state variables (Section 3.2.17) 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 3.4 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 SB0:0000210 (*addition of a chemical group* under *interaction*→*process*→*biochemical* or *transport reaction*→

biochemical reaction→*conversion*). The labels shown in Table 3.4 are defined by SBGN Process Description Level 1; for all other kinds of modifications not listed here, the author of a Process Description must create a new label (and should also describe the meaning of the label in a legend or text accompanying the map).

3.4.4 Physical characteristics

SBGN Process Description Level 1 defines a special unit of information for describing certain common physical characteristics. Table 3.5 lists the particular values defined by SBGN Process Description

Name	Label	SBO term
Acetylation	Ac	SB0:0000215
Glycosylation	G	SB0:0000217
Hydroxylation	OH	SB0:0000233
Methylation	Me	SB0:0000214
Myristoylation	My	SB0:0000219
Palmytoylation	Pa	SB0:0000218
Phosphorylation	P	SB0:0000216
Prenylation	Pr	SB0:0000221
Protonation	H	SB0:0000212
Sulfation	S	SB0:0000220
Ubiquitination	Ub	SB0:0000224

Table 3.4: A sample of values from the covalent modifications vocabulary (Section 3.4.3).

tion Level 1. It is anticipated that these will be used to describe the nature of a perturbing agent (section 3.2.9) or a phenotype (section 3.2.15).

Name	Label	SBO term
Temperature	pc:T	SB0:0000147
Voltage	pc:V	SB0:0000259
pH	pc:pH	SB0:0000304

Table 3.5: A sample of values from the physical characteristics vocabulary (Section 3.4.4).

3.4.5 Cardinality

SBGN Process Description Level 1 defines a special unit of information usable on multimers for describing the number of monomers composing the multimer. Table 3.6 shows the way in which the values must be written. Note that the value is an positive non-zero integer, and not (for example) a range. There is no provision in SBGN Process Description Level 1 for specifying a range in this context because it leads to problems of entity identifiability.

Name	Label	SBO term
cardinality	N:#	SB0:0000364

Table 3.6: The format of the possible values for the cardinality unit of information (Section 3.4.5). Here, # stands for the number; for example, “N:5”.

Chapter 4

998

Validation Rules

999

4.1 Overview

1000

The previous chapter described the glyphs and described rules specific to individual glyphs. In this chapter the rules that govern the interactions of the glyphs in SBGN Process Description Level 1 and that span an Process Description diagram as a whole.

1001

1002

1003

4.2 Semantic rules

1004

4.2.1 EPNs

1005

1. All state variables associated with a Stateful Entity Pool Node should be unique and not duplicated within that node.
2. If a state variable is used in one EPN then it must be used in all equivalent stateful EPNs¹.
3. EPNs should not be orphaned (i.e. they must be associated with at least one arc).

1006

1007

1008

1009

4.2.2 Process Nodes

1010

As described in Section 3.2.15, the consumption and production arcs converge before connecting to the process node (Figure 4.1). This defines the EPNs that are the input and outputs of an irreversible process. Since, processes can be reversible in the following rules we refer to these groupings as the “left-hand-side” (LHS) and “right-hand-side” (RHS) of the process². For convenience we will also collectively refer to the consumption and production arcs as *flux* arcs.

1011

1012

1013

1014

1015

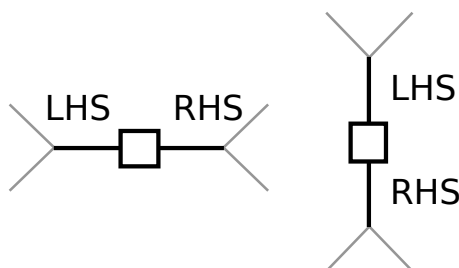


Figure 4.1: An illustration of the “sidedness” of a process. The designation of LHS and RHS is essentially arbitrary.

¹A stateful EPN is equivalent if the EPNs are identical when their state descriptions are ignored.

²Note this designation is purely for grouping and is used even when the sides of the reaction are above and below the process.

Flux Arcs

1. All process nodes (with the exception of **phenotype**) must have a LHS and RHS. 1016 1017
2. All EPNs on the LHS of a process must be unique. 1018
3. All EPNs on the RHS of a process must be unique. 1019
4. All **phenotype** glyphs must be associated with at least one modulation arc. 1020
5. The EPNs that make up the LHS of the process should be consistent with the RHS, i.e. the process should constitute a balanced biochemical reaction. 1021 1022
6. Once the stoichiometry of a flux arc is displayed in a map then all other flux arcs should display their stoichiometry make. 1023 1024
7. If the stoichiometry is undefined or unknown this should be indicated by the use of a question mark (“?”). 1025 1026
8. If more than one set of stoichiometries can be applied to the flux arcs of the process then the stoichiometry of the flux arcs must be displayed. 1027 1028

Association

1. An **Association** is always an irreversible process. 1029 1030

Dissociation

1. An **Dissociation** is always an irreversible process. 1031 1032

Modulation

As discussed in Chapter 2, it is implied, but not defined explicitly that the process has a rate at which it converts its LHS EPNs to its RHS EPNs (and vice-versa in the case of a reversible process). This concept is important in understanding how the Process Description language describes process modulation. 1034 1035 1036 1037

1. A process with no modulations has an underlying “basal rate” which describes the rate at which it converts inputs to outputs. 1038 1039
2. A **modulation** changes the basal rate in an unspecified fashion. 1040
3. A **stimulation** is a modulation that increases the basal rate. 1041
4. An **inhibition** is a modulation that decreases the basal rate. 1042
5. The above types of modulation, when assigned to the same process, are combined and have a multiplicative effect on the basal rate of the process. 1043 1044
6. Modulators that do not interact with each other in the above manner, should be drawn as modulating different process nodes. Their effect is therefore additive. 1045 1046
7. At most one **necessary stimulation** can be assigned to a process node. Two **necessary stimulations** would imply an implicit AND or OR operator. For clarity only one **necessary stimulation** can be assigned to a process, and such combinations must be explicitly expressed using logical operators. 1047 1048 1049 1050

8. At most one catalysis can be assigned to a process. Modulation by a catalysis arc implies that the exact biochemical mechanism underlying the process is known. In this context two catalysis cannot be assigned to the same process node as they represent independent re- actions. Other EPNs can be assigned to the same process as a catalysis, such as modulators, stimulators, and inhibitors, and will have a multiplicative modulation on the reaction rate de- fined by the catalysis.

Reversible Processes

A process is deemed to be reversible if it has production arcs on both the LHS and RHS of a process node Figure 4.2. Semantically, the production arc can be thought of as allowing a reversible flow of entities between the process and the EPN. A consumption arc only permits an irreversible flow from the EPN to the process. In this way, the consumption arc forces the process to be irreversible. Consumption arcs cannot be associated with both sides of a process as this would prohibit any flow through the process.

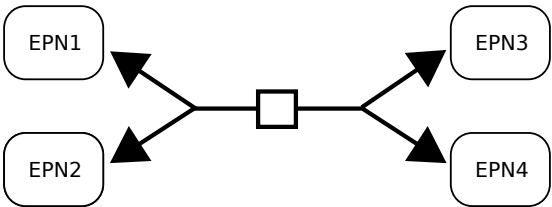


Figure 4.2: A valid reversible process. A process is reversible if its LHS and RHS contain only production arcs.

1. A mixture of consumption and production arcs on the same side of a process is not permitted.
2. A sink cannot be linked to a reversible process as it only receives entities, and so would effectively make the process irreversible³.
3. The semantics of modulation is the same as for irreversible processes, .i.e. the amount of entity in the modulation pool affects the rate of the process.

4.2.3 Cloning

SBGN allows identical nodes to be duplicated on a map if they are explicitly marked as such. This is done using a clone marker. The details are shown in table 4.1.

Table 4.1: Duplication rules.

Node	Can be duplicated	Indication	Additional Rules
Compartment	N		
SimpleChemical	Y	Simple clone marker	
UnspecifiedEntity	Y	Simple clone marker	
Source	N		
Sink	N		
Perturbing Agent	Y	Simple clone marker	
Phenotype	N		
MultimerChemicalEntity	Y	Simple clone marker	
StatefulEntityPool	Y	Labeled clone marker	
Macromolecule	Y	Labeled clone marker	
MultimerMacromolecule	Y	Labeled clone marker	

continued on next page

³A source can only be associated with a consumption arc so this rule does not apply in this case.

continued from previous page			
Node	Duplicate?	Indication	Additional Rules
NucleicAcidFeature	Y	Labeled clone marker	
Complex	Y	Labeled clone marker	
Subunit	N		
Process	N		
OmittedProcess	N		
UncertainProcess	N		
Association	N		
Dissociation	N		
LogicalOperator	Y	None	
AND	Y	None	
OR	Y	None	
NOT	Y	None	

4.2.4 Compartment spanning

An EPN cannot *belong* to more than one compartment. However, an EPN can be *drawn* over more than one compartment. In such cases, the decision on which is the owning compartment is deferred to the drawing tool or the author. A complex may contain EPNs which belong to different compartments and in this way a complex can be used to describe entities that span more than one compartment.

This restriction makes it impossible to represent in a semantically correct way a macromolecule that spans more then one compartment — for example a receptor protein. It is clearly desirable to be able to show a macromolecule in a manner that the biologist expects (i.e. spanning from the outside through the membrane to the inside). Therefore, the author is recommended to draw the macromolecule across compartment boundaries, but the underlying SBGN semantic model will assign it to only one. The assignment to a compartment may be decided by the software drawing tool or the author. Note that this has implications for auto-layout algorithms as they will only be able to treat such entity pool nodes as contained within a compartment and will have no way of knowing a macromolecule spans a compartment.

The current solution is consistent with other Systems Biology representations such as SBML and BioPAX. For more information about the problems representing membrane spanning proteins and the rationale behind the current solution see Section C.

4.2.5 Submaps

The submap is a visual device that allows the detail of an Process Description map to be exported into another Process Description map and replaced by a submap glyph, which acts as a place-holder. This is described and illustrated in Section 3.2.15. In the following discussion we will refer to the original map as the *main* map and the map containing the export detail as the submap.

1. For a valid mapping between an EPN in the map and submap to exist the identifiers in the tag and the submap terminal must be identical and their associated entity pool nodes must be identical.
2. If the same EPN is present in the map and a submap, then they must be mapped to each other.
3. Since the main map and submap share the same namespace, an EPN that is cloned in the main map must also be marked as cloned in the submap — even if there is only one copy of the EPN in the submap. The converse applies when the EPN in the submap is cloned⁴.

⁴This has the additional benefit of ensuring that main maps and submaps do not need to be modified if the submap is expanded and collapsed by a viewing or editing tool.

Chapter 5

1105

Layout Rules for a Process Description

1106

5.1 Introduction

1107

The previous chapters describe the appearance and meaning of SBGN Process Description Level 1 components. Here we describe rules governing the visual appearance and aesthetics of the Process Description language. The components of a Process Description 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*.

1108
1109
1110
1111
1112
1113
1114

SBGN Process Descriptions should be easily recognisable 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 rules for the layout of process description maps, divided into two categories:

1115
1116
1117
1118

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

1119
1120

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

1121
1122

Those layout rules 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 rules 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 [?].

1123
1124
1125
1126
1127

Please note that the color of objects do not carry any meaning in SBGN. Although one can use colors to emphasize part 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 process node may be larger than a protein node. Also the meaning of a graph should be conserved upon scaling as far as possible.

1128
1129
1130
1131
1132

5.2 Requirements

1133

Requirements are rules which **must** be fulfilled by a layout to produce a valid Process Description map.

1134
1135

5.2.1 Node-node overlaps

1136

Nodes are only allowed to overlap in two cases when they are allowed to contain other nodes — as described in Chapter 3. Otherwise, nodes are not allowed to overlap (Figure 5.1). This includes the

1137
1138

touching of nodes. Touching is not allowed apart from the case where it has a specific meaning, e.g. two macromolecules touching each other within a complex because they form the complex.

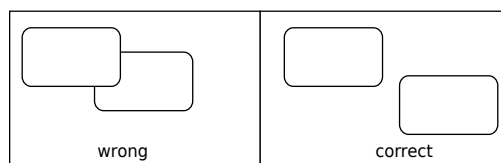


Figure 5.1: *Nodes must not overlap.*

5.2.2 Node-edge crossing

Edges must be drawn on the top of a the node (Figure 5.2). See also recommendation Section 5.3.1.

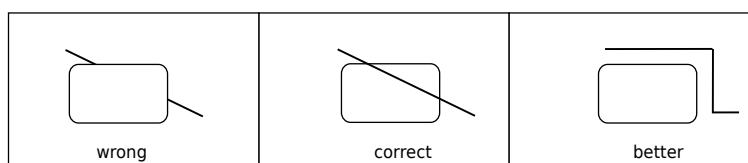


Figure 5.2: *If an edge crosses a node, the edge must be drawn on top of the node.*

5.2.3 Node border-edge overlaps

Edges are not allowed to overlap the border lines of nodes (Figure 5.3).

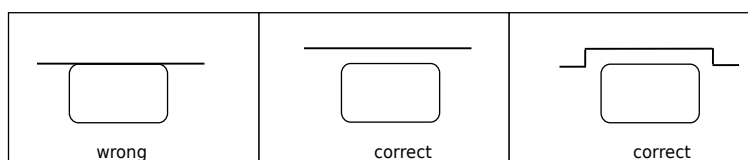


Figure 5.3: *Edges must not overlap node borders.*

5.2.4 Edge-edge overlaps

Edges are not allowed to overlap (Figure 5.4). 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.

5.2.5 Node orientation

Nodes have to be drawn horizontally or vertically, any other rotation of elements is not allowed (Figure 5.5).

5.2.6 Node-edge connection

1. The arcs linking the square glyph of a process to the consumption and production arcs are attached to the center of opposite sides (Figure 5.6).

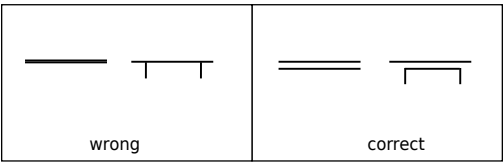


Figure 5.4: Edges must not overlap.

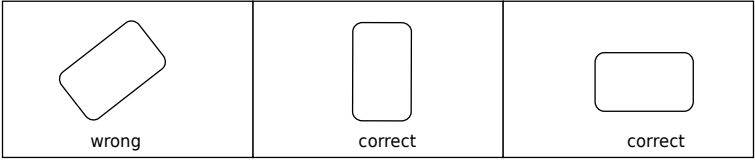


Figure 5.5: The node orientation must be horizontally or vertically.

2. The modulatory arcs are attached to the other two sides, but not necessarily all to the center, as several modifiers can affect the same process node.

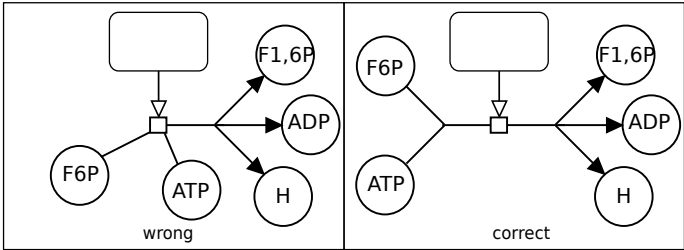


Figure 5.6: Arcs between a process and the consumption and production arcs must be attached to the center of opposite sides, modulatory arcs must be attached to the other two sides.

5.2.7 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 other nodes or other labels (this includes touching of other nodes or labels).

5.2.8 Edge labels

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

5.2.9 Compartments

If a process has all participants in the same compartment the process node and all edges/arcs should be drawn in this compartment. If a process has participants in at least two different compartments, the process node has to be either in a compartment where the process has at least one participant or in the empty space.

5.3 Recommendations

Recommendations are rules which should be followed if possible and generally should improve the clarity of the diagram.

5.3.1 Node-edge crossing

1171

Situations where edges and nodes cross should be avoided. Note that some crossings may be un-avoidable, e. g. the crossing between an edge and a compartment border or an edge and a complex (if the edge connects an element inside the complex with something outside).

1172

1173

1174

5.3.2 Labels

1175

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.

1176

1177

1178

5.3.3 Avoid edge crossings

1179

The amount of crossings between edges should be minimized.

1180

5.3.4 Branching of association and dissociation

1181

The branching points of association and dissociation nodes should be placed closed to the symbol of the process, if possible at a distance comparable than, or smaller to, the diameter of the symbol defining the process (Figure 5.7).

1182

1183

1184

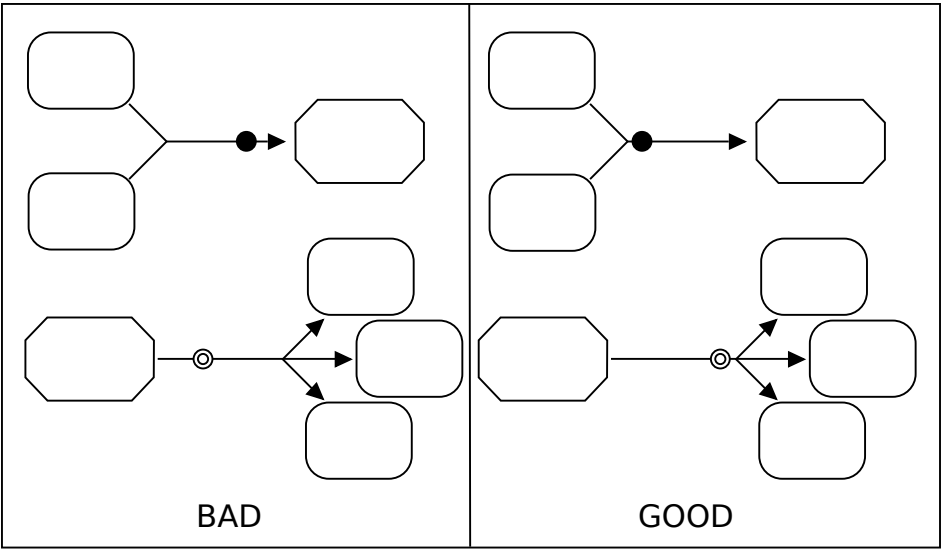


Figure 5.7: Branching points should be close to association and dissociation symbols.

5.3.5 Units of information

1185

Units of information should not hide the structure of the corresponding node and should not overlap other elements (Figure 5.8).

1186

1187

5.4 Additional suggestions

1188

Here is a list of additional layout suggestions which may help improve the aesthetics and clarity of Process Description maps.

1189

1190

- Angle of edge crossings: If edge crossing cannot be avoided then the edges should cross with an angle close to 90 degrees.
- 1191
- 1192

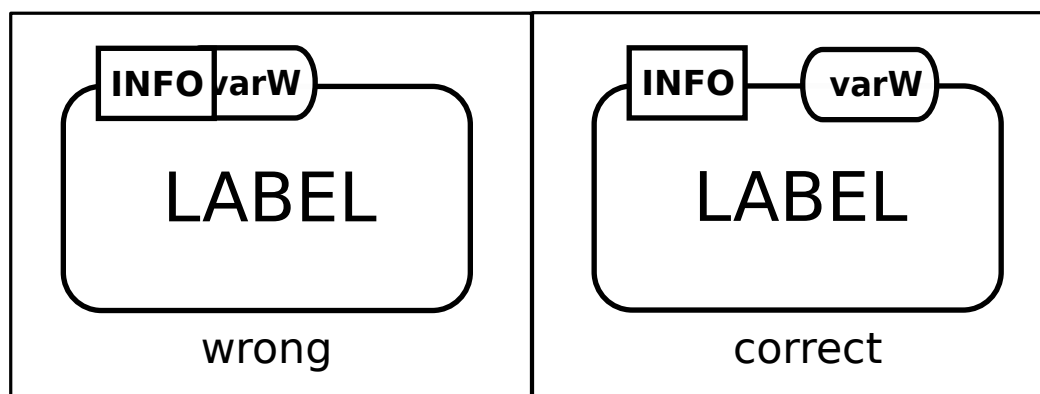


Figure 5.8: Units of information should not overlap with any other element.

- 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. 1193 1194
- Edge length: Long edges should be avoided. 1195
- Number of edge bends: Edges should be drawn with as few bends as possible. 1196
- Similar and symmetric parts: Similar parts of a map should be drawn in a similar way, and symmetric parts should be drawn symmetrically. 1197 1198
- Proximity information: Related elements (e. g. nodes connected by a process or all elements within a compartment) should be drawn close together. 1199 1200
- Directional information: Subsequent processes (e. g. a sequence of reactions) should be drawn in one direction (e. g. from top to bottom or from left to right). 1201 1202
- Compartments: It can help clarity to use a different background shade or color for each compartment. 1203 1204

Chapter 6

1205

Acknowledgments

1206

Here we acknowledge those people and organisations that assisted in the development of this and previous releases of the SBGN Process Description language specification. First we specifically acknowledge those who contributed directly to each revision of the specification document, followed by a comprehensive acknowledgement of contributors that attended workshops and forum meetings or in some other way provided input to the standard. Finally, we acknowledge the bodies that provided financial support for the development of the standard.

1207

1208

1209

1210

1211

1212

6.1 Level 1 Release 1.0

1213

The specification of was written by Nicolas Le Novère, Stuart Moodie, Anatoly Sorokin, Michael Hucka, Falk Schreiber, Emek Demir, Huaiyu Mi, Yukiko Matsuoka, Katja Wegner and Hiroaki Kitano. In addition, the specification benefited much from the help of (in alphabetical order) Frank Bergmann, Sarala Dissanayake, Ralph Gauges, Peter Ghazal, Lu Li, and Steven Watterson.

1214

1215

1216

1217

6.2 Level 1 Release 1.1

1218

The specification of SBGN PD Level 1.1 was written by Stuart Moodie and Nicolas Le Novère, with contributions from (in alphabetical order) Frank Bergmann, Sarah Boyd, Emek Demir, Sarala Wimalaratne, Yukiko Matsuoka, Huaiyu Mi, Falk Schreiber, Anatoly Sorokin, Alice Villéger.

1219

1220

1221

6.3 Level 1 Release 1.2

1222

The specification of SBGN PD Level 1.2 was modified by Stuart Moodie, with contributions from (in alphabetical order) Sarah Boyd, Nicolas Le Novère, Huaiyu Mi.

1223

1224

6.4 Level 1 Release 1.3

1225

The specification of SBGN PD Level 1.3 was modified by Stuart Moodie, with contributions from (in alphabetical order), Tobias Czauderna, Nicolas Le Novère, Anatoly Sorokin.

1226

1227

6.5 Comprehensive list of acknowledgements

1228

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 intend this list to be complete. We are very sorry if we forgot someone, and would be grateful if you could notify us of any omission.

1229

1230

1231

1232

Mirit Aladjemm, Frank Bergmann, Sarah Boyd, Laurence Calzone, Melanie Courtot, Emek Demir, Ugur Dogrusoz, Tom Freeman, Akira Funahashi, Ralph Gauges, Peter Ghazal, Samik Ghosh, Igor Goryanin, Michael Hucka, Akiya Jouraku, Hideya Kawaji, Douglas Kell, Sohyoung Kim, Hiroaki Kitano, Kurt Kohn, Fedor Kolpakov, Nicolas Le Novère, Lu Li, Augustin Luna, Yukiko Matsuoka, Huaiyu Mi, Stuart Moodie, Sven Sahle, Chris Sander, Herbert Sauro, Esther Schmidt, Falk Schreiber, Jacky Snoep, Anatoly Sorokin, Jessica Stephens, Linda Taddeo, Steven Watterson, Alice Villéger, Katja Wegner, Sarala Wimalaratne, Guanming Wu.

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

6.6 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-r.org/>), and the Beckman Institute at the California Institute of Technology (<http://bnmc.caltech.edu>) 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.

Appendix A

Complete examples of Process Description Maps

The following maps present complete examples of SBGN Process Descriptions representing Biological processes. They by no mean exhaust the possibilities of SBGN Process Description Level 1.

Figure A.1 presents an example of metabolic pathway, that exemplifies the use of the EPNs simple chemical, macromolecule, and clone marker, the PNs process, and the connecting arcs consumption, production and catalysis.

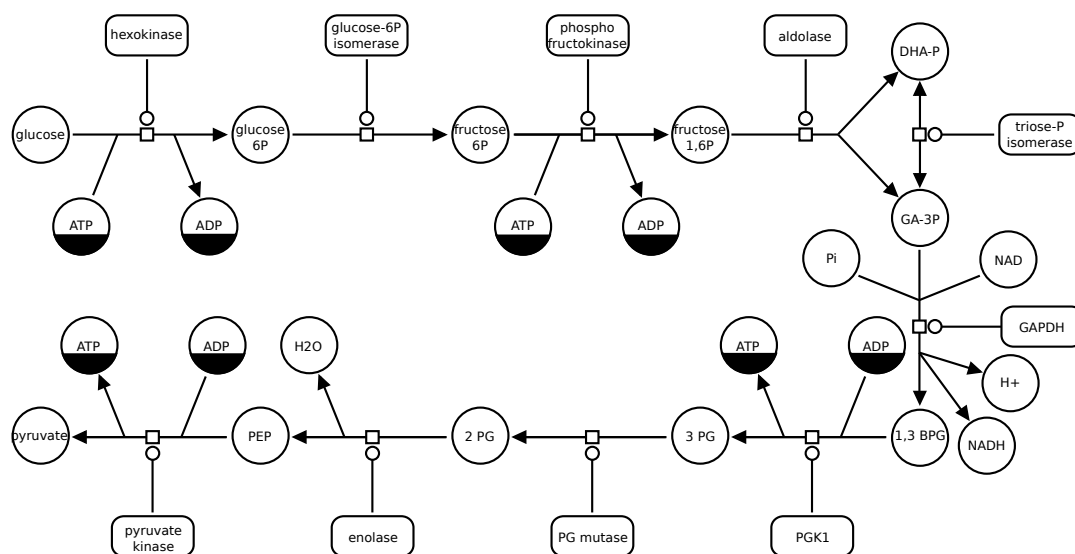


Figure A.1: Glycolysis. This example illustrates how SBGN can be used to describe metabolic pathways.

Figure A.2 presents an example of signalling pathway, that exemplifies in addition the use of the EPNs phenotype, and state variable, the containers complex, compartment and submap, the PNs association, and the connecting arcs stimulation. Note the complex IGF and IGF receptor, located on the boundary of the compartment. This position is only for user convenience. The complex has to belong to a given compartment in SBGN Process Description Level 1.

Figure A.3 is an expanded version of the submap present on the map present in Figure A.2. It shows the use of tag.

Figure A.4 introduces an SBGN Process Description that spans several compartments. Note that the compartment "synaptic vesicle" is not **contained** in the compartment "synaptic button" but **overlaps** it. The simple chemical "ACh" of the "synaptic vesicle" is not the same EPN than the "ACh" of

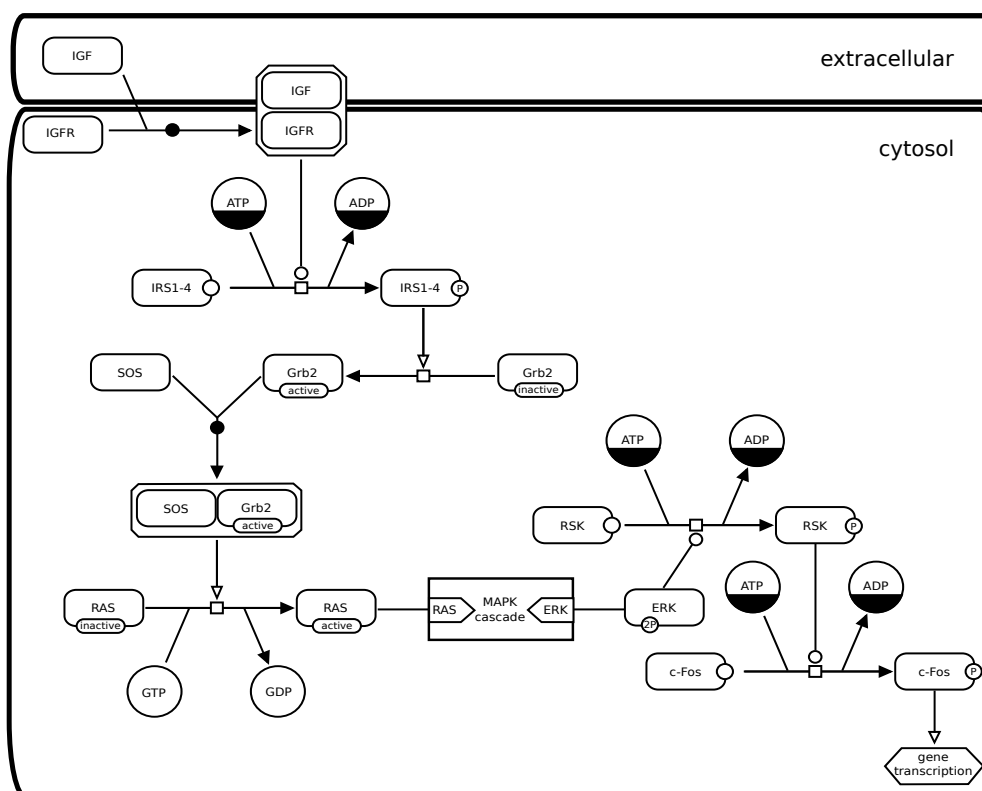


Figure A.2: Insulin-like Growth Factor (IGF) signalling. This example shows the use of compartments and how details can be hidden by using a submap. The submap is shown on Figure A.3.

the “synaptic button” and of “synaptic cleft”. The situation is similar with the compartments “ER” 1274
and “muscle cytosol”. The map exemplifies the use of the PN omitted and dissociation, and the 1275

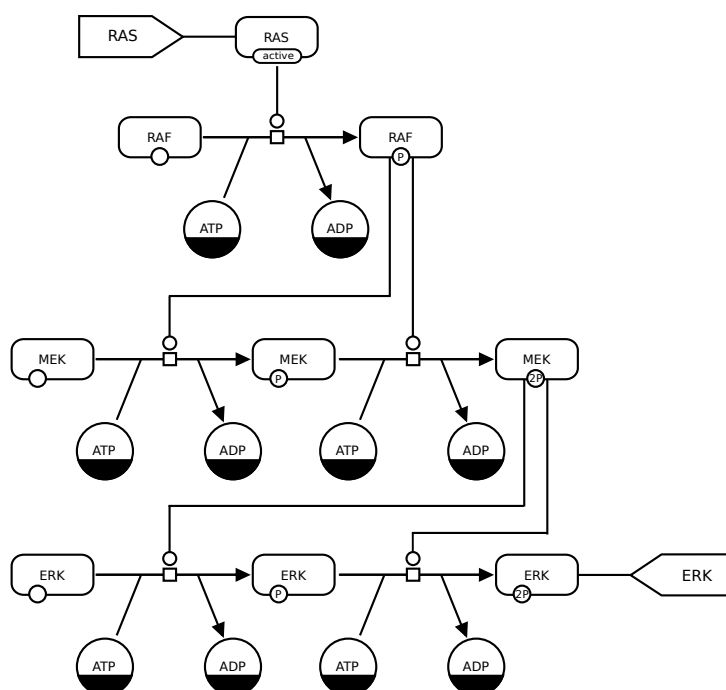


Figure A.3: A submap of the previous map showing the MAPK cascade.

connecting arc necessary stimulation.

1276

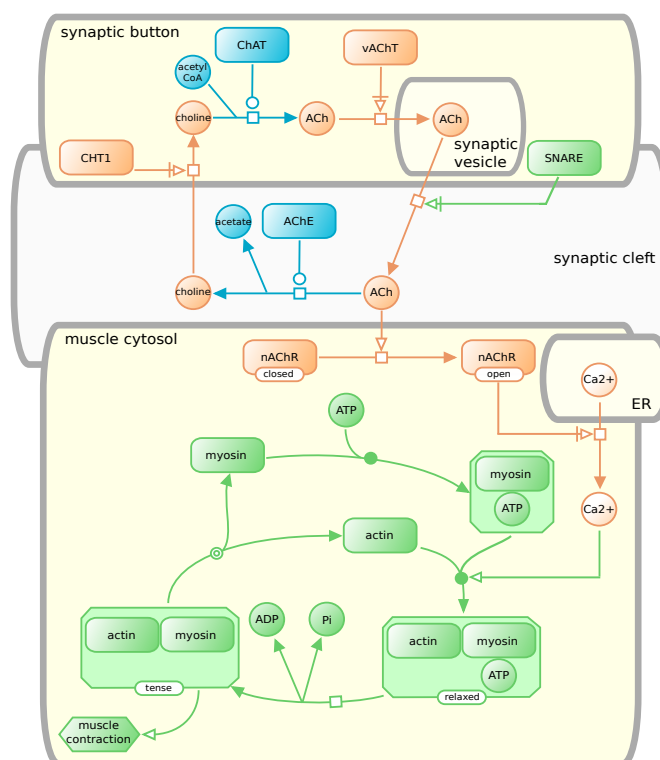


Figure A.4: Neuronal/Muscle signalling. A description of inter-cellular signalling using SBGN.

Figure A.5 introduces the use of SBGN Process Description Level 1 to encode gene regulatory networks. It also show the use of the EPNs Source and the logical operator and.

1277

1278

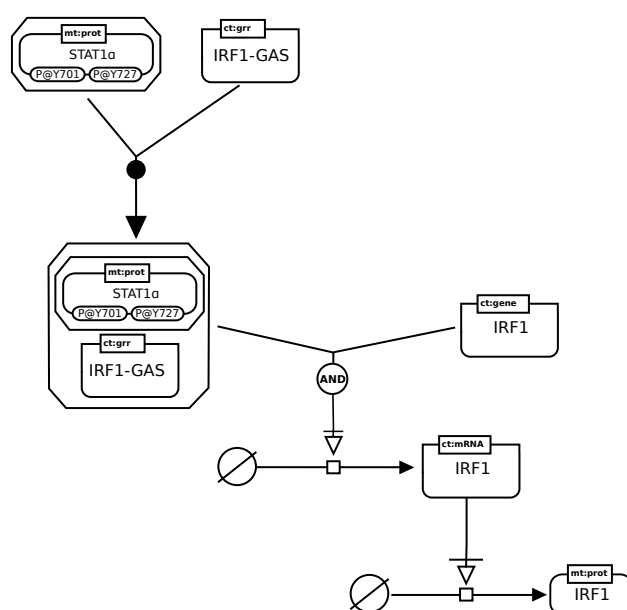


Figure A.5: Activated *STAT1α* induction of the *IRF1* gene. An example of gene regulation using logical operators.

Appendix B

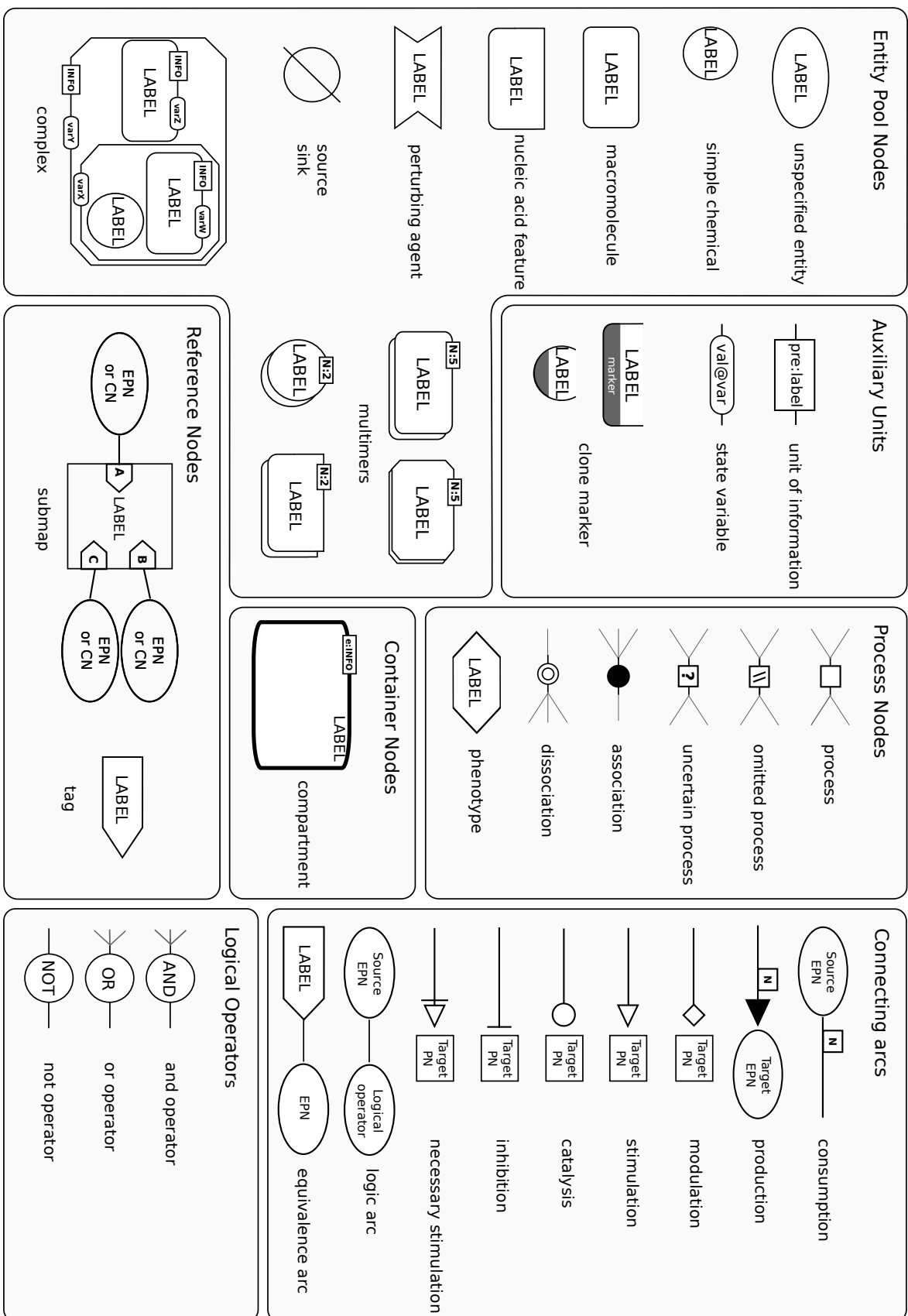
1279

Reference card

1280

Print the summary of SBGN symbols on the next page for a quick reference.

1281



Appendix C

1283

Issues postponed to future levels

1284

C.1 Multicompartment entities

1285

The problem of entities, such as macromolecules, spanning several compartments proved to be a challenge for the community involved in the development of SBGN Process Description Level 1. It was thus decided to leave it for a future Level. It turns out there is at the moment no obvious solution satisfactory for everyone. Three broad classes of solutions have been identified so far:

1286

1287

1288

1289

- One can systematically locate an EPN in a given compartment, for instance a transmembrane receptor in a membrane. However, the reactions of this entity with entities represented by EPN in other compartments, such as extracellular ligands and second messenger systems, will create artificial transport reactions.
- One can represent the domains of proteins in different compartments by macromolecules, and link all those macromolecules in a complex spanning several compartments. However, such a representation would be very confusing, implying that the domains are actually different molecules linked through non-covalent bonds.
- One can accept macromolecules that span several compartments, and represent domains as units of information. Those units of information should then be located in given compartments. To make a full use of such a representation, one should then start and end connecting arcs on given units of information, something prohibited by the current specification.

1290

1291

1292

1293

1294

1295

1296

1297

1298

1299

1300

1301

1302

C.2 Logical combination of state variable values

1303

The value of a state variable has to be perfectly defined in SBGN Process Description Level 1. If a state variable can take the alternative values 'A', 'B' and 'C', one cannot attribute it values such as 'non-A', 'A or B', 'any' or 'none'. As a consequence some biochemical processes cannot be easily represented because of the very large number of state to enumerate. The decision to forbid such a Boolean logic lies in the necessity of maintaining truth path all over an SBGN map.

1304

1305

1306

1307

1308

C.3 Non-chemical entity nodes

1309

The current specification cannot represent combinations of events and entities. For instance a variable "voltage" cannot be controlled by a difference of concentration between different entities, such as a given ion in both sides of a membrane.

1310

1311

1312

C.4 Generics

1313

SBGN Process Description Level 1 does not provide mechanisms to sub-class EPNs. There is no specific means of specifying that macromolecules or nucleic acid features X1, X2 and X2 are subclasses of X. Therefore, any process that applies to all the subtypes of X has to be triplicated. That situation can easily generate combinatorial explosions of the number of EPNs or PNs.

1314

1315

1316

1317

C.5 State and transformation of compartments

1318

In SBGN Process Description Level 1 a compartment is a stateless entity. It cannot carry state variables, and cannot be subjected to process modifying a state. As a result, a compartment cannot be transformed, moved, split or merged with another. If one want to represent the transformation of a compartment, one has to create the start and end compartments, and represent the transport of all the EPNs from one to the other. This is not satisfactory, and should be addressed in the future.

1319

1320

1321

1322

1323

Appendix D

1324

Revision History

1325

D.1 Version 1.0 to Version 1.1

1326

Below are the changes incorporated into Version 1.1 of the SBGN Process Description Level 1 specification.

1327

1328

Description	Tracker ID
Regarding modulation of reversible processes, changed “should” to “must” be represented by two process	
Removed “The connectors and the box move as a rigid entity” in the definition of process	
Changed the definition of process node to “represent processes that transform one or several EPNs into <i>one or several EPNs, identical or different</i> ”	
Changed SBO term of compartment From SBO:0000289 (functional compartment) to SBO:0000290 (physical compartment)	
Reorganised classification of glyphs	
Reorganised glyph section to reflect the above changes	
Revised reference card to reflect changes in glyph organisation	
Revised logic operators throughout spec to make sure input and output arcs meet before attaching to the glyph - as with processes.	
Added enumerated rules to grammar section. This is probably not complete, but should help the implementation of semantic validation by software tools. The hope is this will be refined as tools start validating maps.	
Updated UML maps and data dictionary to be consistent with rest of changes to spec.	
Definition of cardinality is ambiguous	2840996
Sink and source are lumped together	2726435
SBO terms are incorrect or missing.	2841261
Compartment description is confusing and contradictory.	2841122
Clone marker fill percentages unhelpful.	2841114
Use of CV for physical characteristic not clear.	2841085
Definition of Cardinality is ambiguous.	2840996
input to AND on IFN example.	2804326
more SBO terms for multimers	2803593
Legend of figure 2.20 is incorrect	2803537
legend of figure 3.2	2802990
continued on next page	

1329

<i>continued from previous page</i>	
Description	Tracker ID
Compartment colouring	2745703
Errors in diag a4.	2664912
Change name of trigger glyph.	2664908
Transition should be renamed process.	2664862
Converting arcs tautological.	2664843
Example invalid.	2545870
consumption and production.	2388317
Should require circles to be distinguishable from ellipses	2219388
Figure 2.53	2162619
Reference card: production	2104471
Figure 2.42 is wrong	2104465
Mistake in the multi-cellular example	2395488
Should not prevent processes having identical in and out	2664933
No description of linking to subunit rules.	2545810
Extensively revised the grammar section. The UML diagrams have been simplified to show glyph taxonomy, and the data dictionary has been pruned to just show glyph identity. The some syntax rules have been moved into semantics and the rules reformulated to make them easier to understand.	
Eliminated duplicate rules in layout section and revised text slightly.	
Phenotype cloning?	2989007
Perturbing agent description	2940021

1330

D.2 Version 1.1 to Version 1.2

1331

Below are the changes incorporated into Version 1.2 of the SBGN Process Description Level 1 specification.

1332

1333

Description	Tracker ID
Perturbing agent description	2940021
Members of complex touching	2849273
PD Reference card error for submap glyph	3029242

1334

D.3 Version 1.2 to Version 1.3

1335

Below are the changes incorporated into Version 1.3 of the SBGN Process Description Level 1 specification.

1336

1337

Description	Tracker ID
Incorrect editor on title page	
Typos in acknowledgements	
Fixed typo in item on catalysis in section 4.2.2.	
State variables figure 2.6 V1.2	3090543

1338

Bibliography

1339

- [1] G. Di Battista, P. Eades, R. Tamassia, and I.G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, New Jersey, 1998. 1340
1341
- [2] M. Kaufmann and D. Wagner. *Drawing Graphs: Methods and Models*, volume 2025 of *Lecture Notes in Computer Science Tutorial*. Springer, 2001. 1342
1343