Systems Biology Graphical Notation:
Process Description language Level 1

Version 1.3
14 February, 2010

Editors:
Stuart Moodie
Nicolas Le Novère
Emek Demir
Huaiyu Mi
Falk Schreiber

School of Informatics, University of Edinburgh, UK
EMBL European Bioinformatics Institute, UK
Sloan-Kettering Institute, USA
SRI International, USA
IPK Gatersleben & MLU Halle, Germany

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.
5.5 Comprehensive list of acknowledgements 52
5.6 Financial Support 53

A Complete examples of Process Description Maps 54

B Reference card 58

C Issues postponed to future levels 60
C.1 Multicompartment entities 60
C.2 Logical combination of state variable values 60
C.3 Non-chemical entity nodes 60
C.4 Generics 61
C.5 State and transformation of compartments 61

D Revision History 62
D.1 Version 1.0 to Version 1.1 62
D.2 Version 1.1 to Version 1.2 63
D.3 Version 1.2 to Version 1.3 63
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:


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

Process Description Glyphs

2.1 Overview

To set the stage for what follows in this chapter, 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.

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 phosphorylated 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 on the preceding page, 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
on the previous page, those links describe consumption Section 2.9.1, production Section 2.9.2
and catalysis Section 2.9.5, 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 2.3.3). The details of this and
other aspects of Process Description notation are explained in the rest of this chapter.

Table 2.1 summarizes the different SBGN abstractions described in this chapter.

<table>
<thead>
<tr>
<th>Component</th>
<th>Abbrev.</th>
<th>Role</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity pool node</td>
<td>EPN</td>
<td>A population of entities that cannot be</td>
<td>Specific macromolecules or other chemical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>distinguished from each other</td>
<td>species</td>
</tr>
<tr>
<td>Container node</td>
<td>CN</td>
<td>An encapsulation of one or more other</td>
<td>Complexes, compartments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SBGN constructs</td>
<td></td>
</tr>
<tr>
<td>Process node</td>
<td>PN</td>
<td>A process that transforms one or more</td>
<td>Process, association, dissociation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EPNs into one or more other EPNs</td>
<td></td>
</tr>
<tr>
<td>Arc</td>
<td>—</td>
<td>Links between EPNs or CNs to PNs or CNs to</td>
<td>Production, catalysis, inhibition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>indicate influences</td>
<td></td>
</tr>
<tr>
<td>Logical operators</td>
<td>—</td>
<td>Combines one or several inputs into one</td>
<td>Boolean and, or, not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>output</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Summary of Process Description components and their roles.

2.2 Controlled vocabularies used in SBGN Process Description Level 1

Some glyphs in SBGN Process Descriptions can contain particular kinds of textual annotations
carrying information relevant to the purpose of the glyph. These annotations are units of
information (Section 2.3.1) or state variable (Section 2.3.2). 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 2.2.3). 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.

### 2.2.1 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 2.2, but others are possible. The values are to be taken from the Systems Biology Ontology (http://www.ebi.ac.uk/sbo/), specifically from the branch having identifier SBO:0000240 (material entity under entity). The labels are defined by SBGN Process Description Level 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>SBO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-macromolecular ion</td>
<td>mt:ion</td>
<td>SBO:0000327</td>
</tr>
<tr>
<td>Non-macromolecular radical</td>
<td>mt:rad</td>
<td>SBO:0000328</td>
</tr>
<tr>
<td>Ribonucleic acid</td>
<td>mt:rna</td>
<td>SBO:0000250</td>
</tr>
<tr>
<td>Deoxyribonucleic acid</td>
<td>mt:dn</td>
<td>SBO:0000251</td>
</tr>
<tr>
<td>Protein</td>
<td>mt:prot</td>
<td>SBO:0000297</td>
</tr>
<tr>
<td>Polysaccharide</td>
<td>mt:psac</td>
<td>SBO:0000249</td>
</tr>
</tbody>
</table>

**Table 2.2:** A sample of values from the material types controlled vocabulary (Section 2.2.1).

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

### 2.2.2 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 2.3, 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 SBO:0000241 (conceptual entity under entity). The labels are defined by SBGN Process Description Level 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>SBO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>ct:gene</td>
<td>SBO:0000243</td>
</tr>
<tr>
<td>Transcription start site</td>
<td>ct:tss</td>
<td>SBO:0000329</td>
</tr>
<tr>
<td>Gene coding region</td>
<td>ct:gene</td>
<td>SBO:0000335</td>
</tr>
<tr>
<td>Gene regulatory region</td>
<td>ct:grr</td>
<td>SBO:0000369</td>
</tr>
<tr>
<td>Messenger RNA</td>
<td>ct:mRNA</td>
<td>SBO:0000278</td>
</tr>
</tbody>
</table>

**Table 2.3:** A sample of values from the conceptual types vocabulary (Section 2.2.2).
2.2.3 Macromolecule covalent modifications

A common reason for the introduction of state variables (Section 2.3.2) on an entity is to allow access to the configuration of possible covalent modification sites on that entity. For instance, a macromolecule may have one or more sites where a phosphate group may be attached; this change in the site's configuration (i.e., being either phosphorylated or not) may factor into whether, and how, the entity can participate in different processes. Being able to describe such modifications in a consistent fashion is the motivation for the existence of SBGN’s covalent modifications controlled vocabulary.

Table 2.4 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 SBO:0000210 (addition of a chemical group under interaction→process→biochemical or transport reaction→biochemical reaction→conversion). The labels shown in Table 2.4 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).

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

Table 2.4: A sample of values from the covalent modifications vocabulary (Section 2.2.3).

2.2.4 Physical characteristics

SBGN Process Description Level 1 defines a special unit of information for describing certain common physical characteristics. Table 2.5 lists the particular values defined by SBGN Process Description Level 1. It is anticipated that these will be used to describe the nature of a perturbing agent (section 2.4.8) or a phenotype (section 2.8.6).

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>SBO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>pc:T</td>
<td>SBO:0000147</td>
</tr>
<tr>
<td>Voltage</td>
<td>pc:V</td>
<td>SBO:0000259</td>
</tr>
<tr>
<td>pH</td>
<td>pc:pH</td>
<td>SBO:0000304</td>
</tr>
</tbody>
</table>

Table 2.5: A sample of values from the physical characteristics vocabulary (Section 2.2.4).
2.2.5 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 2.6 shows the way in which the values must be written. Note that the value is a 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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Label</th>
<th>SBO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>cardinality</td>
<td>N:#</td>
<td>SBO:0000364</td>
</tr>
</tbody>
</table>

Table 2.6: The format of the possible values for the cardinality unit of information (Section 2.2.5). Here, # stands for the number; for example, “N:5”.

2.3 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).

2.3.1 Glyph: Unit of information

When representing biological entities, it is often necessary to convey some abstract information about the entity’s function 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 mid-line 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 predefined 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 2.2 and summarized in the following list:

- pc: container physical characteristic
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

\text{mt} \quad \text{entity pool material type}
\text{ct} \quad \text{entity pool conceptual type}
N \quad \text{multimer cardinality}

Auxiliary items:
A \textit{unit of information} does not carry any auxiliary items.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.2.png}
\caption{The Process Description glyph for \textit{unit of information}.}
\end{figure}

2.3.2 Glyph: \textit{State variable}

Many biological entities such as molecules can exist in different \textit{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 \textit{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 \textit{state variables}, and the values of the \textit{state variables} of all its possible components, recursively.

\textbf{SBO Term:}
Not applicable.

\textbf{Container:}
A \textit{state variable} is represented by an elliptical container, as shown in Figure 2.3. The ellipse’s long axis should be tangent to the border of the glyph of the \textit{EPN} being modified by the \textit{state variable}. The center of the bounding box of a \textit{state of information} should be located on the mid-line of the border of the \textit{EPN}.

\textbf{Label:}
The identification of an instance of a \textit{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 \textit{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 \textit{state variable}. This identification should be present if confusion is possible between several state variables (e.g., several phosphorylation sites). The center of the combination of the boxes located in the container box is superposed to the center of this container box. Optionally, the identification of the \textit{state variable} can be located outside the \textit{state variable} container box. This is \textbf{strongly} discouraged. See Figure 2.4 on the next page for some examples of problems arising if the identification of a state variable is located outside the state variable. The style of labeling of \textit{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 `@', \textit{X}@\textit{Y} thus meaning \textit{value X AT variable Y}.
Auxiliary items:

A state variable does not carry any auxiliary items.

Figure 2.3: 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”.

The label of a state variable should, if possible, be displayed within the ellipse. In the top half of Figure 2.4, we show some examples of pathological cases that lead to confusion in the association between variable labels and values. Compare the discouraged examples with the recommended version in the bottom half of the figure.

Figure 2.4: (Upper part) Examples of incorrect state variables. (Lower part) Correct version.

2.3.3 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 2.7.1). 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.

2.3.3.1 Simple clone marker

As mentioned above, the simple clone marker is the unlabeled version of the clone marker. See below for the labeled version.
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

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 2.5 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 2.5: The Process Description glyph for simple clone marker applied to a simple chemical, an phenotype and a multimer of simple chemicals.

2.3.3.2 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 2.6: The Process Description glyph for labeled clone marker applied to a macromolecule, a nucleic acid feature and a multimer of macromolecules.

Figure 2.7 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.

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. 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 six glyphs representing classes of material entities: unspecified entity (Section 2.4.1), simple chemical (Section 2.4.2), macromolecule (Section 2.4.3), nucleic acid feature (Section 2.4.4), multimer (Section 2.4.5) and complex (Section 2.4.6). (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 three conceptual entities: source, sink (Section 2.4.7), and perturbing agent (Section 2.4.8). Material and conceptual entities can optionally carry auxiliary units such as units of information (Section 2.3.1), state variables (Section 2.3.2) and clone markers (Section 2.3.3).
2.4.1 Glyph: *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.)

**SBO Term:**

SBO:0000285 ! material entity of unspecified nature

**Container:**

An *unspecified entity* is represented by an elliptic container, as shown in Figure 2.8. Note that this must remain an ellipse to avoid confusion with the Simple Chemical glyph, which is a circle (c.f. 2.4.2).

**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.

**Auxiliary items:**

An *unspecified entity* may carry a *clone marker* (Section 2.3.3).

![Figure 2.8: The Process Description glyph for unspecified entity.](image)

2.4.2 Glyph: *Simple chemical*

A simple chemical in SBGN is defined as the opposite of a macromolecule (Section 2.4.3): it is a chemical compound that is *not* formed by the covalent linking of pseudo-identical residues. Examples of simple chemicals are an atom, a monoatomic ion, a salt, a radical, a solid metal, a crystal, etc.

**SBO Term:**

SBO:0000247 ! simple chemical

**Container:**

A *simple chemical* is represented by a circular container, as depicted in Figure 2.9 on the following page. To avoid confusion with the Unspecified Entity (2.4.1), 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.
Auxiliary items:

A simple chemical may be decorated with one or more units of information (Section 2.3.1). A particular unit of information describes the material type. A simple chemical may also carry a clone marker (Section 2.3.3).

Figure 2.9: The Process Description glyph for simple chemical.

2.4.3 Glyph: 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.)

SBO Term:

SBO:0000245 ! macromolecule

Container:

A macromolecule is represented by a rectangular container with rounded corners, as illustrated in Figure 2.10 on the next page.

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.

Auxiliary items:

A macromolecule can carry state variables that can add information about its state (Section 2.3.2). The state of a macromolecule is therefore defined as the vector of all its state variables.

A macromolecule can also carry one or several units of information (Section 2.3.1). The units of information can characterize a domain, such as a binding site. Particular units of information are available for describing the material type (Section 2.2.1) and the conceptual type (Section 2.2.2) of a macromolecule.

A macromolecule may also carry a clone marker (see Section 2.3.3).
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

2.4.4 Glyph: 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 units of information are often important for making the purpose clear to the reader of a map.

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 2.11. 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 2.3.2) 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 2.3.1). 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 2.2.1) and the conceptual type (Section 2.2.2) of the nucleic acid feature.

A nucleic acid feature may also carry a clone marker (Section 2.3.3).

2.4.5 Glyph: Multimer

As its name implies, a multimer is an aggregation of multiple identical or pseudo-identical entities held together by non-covalent bonds (Thus, they are distinguished from polymers by
the fact that the later involve covalent bonds). Here, pseudo-identical refers to the possibility that the entities differ chemically but retain some common global characteristic, such as a structure or function, and so can be considered identical within the context of the SBGN Process Description. An example of this are the homologous subunits in a hetero-oligomeric receptor. SBGN Process Description accepts multimers of simple chemical (Section 2.4.2), macromolecule (Section 2.4.3), nucleic acid feature (Section 2.4.4) or complex (Section 2.4.6).

SBO Term:
- Macromolecule SBO:0000420 ! multimer of macromolecules
- Complex SBO:0000418 ! multimer of complexes
- Nucleic Acid Feature SBO:0000419 ! multimer of informational molecule segments
- Simple Chemical SBO:0000421 ! multimer of simple chemicals

Container:
A multimer is represented by two identical containers shifted horizontally and vertically and stacked one on top of the other. Figure 2.12 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.

Auxiliary items:
A multimer can carry state variables that can add information about its state (Section 2.3.2). The state of a multimer is therefore defined as the vector of all its state variables. Note that a state variable carried by a multimer actually applies to each of the constituent monomers individually. If instead the state variables are meant to apply to the whole multimeric assembly, a macromolecule (Section 2.4.3) should be used instead of multimer. An assembly containing some state variables applicable to the components, and others state variable applicable to the assembly (for instance opening of a channel and phosphorylation of each of its subunits) should be represented by a complex (Section 2.4.6).

A multimer can also carry one or several units of information (Section 2.3.1). The information can characterize a domain, such as a binding site. Particular units of information exist for describing the material type (Section 2.2.1), the conceptual type (Section 2.2.2), and the cardinality (Section 2.2.5) of the multimer. Note that a unit of information carried by a multimer actually applies to each of the constituent monomers individually. If instead a unit of information should be applicable to the whole multimeric assembly, a macromolecule should be used (Section 2.4.3). An assembly containing units of information applying to the components, and others to the assembly should be represented by a complex (Section 2.4.6).

A multimer may also carry a clone marker (Section 2.3.3).

Figure 2.12: The Process Description glyph for multimer with an additional unit of information containing the cardinality.
### 2.4.6 Glyph: Complex

A complex node represents a biochemical entity composed of other biochemical entities, whether macromolecules, simple chemicals, multimers, or other complexes. The resulting entity may have its own identity, properties and function in an SBGN map.

**SBO Term:**

SBO:0000253 ! non-covalent complex

**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. The label box has to be attached to the midway between the border of the complex’s container box and the border of the components’ container boxes.

**Auxiliary items:**

A complex can carry state variables (see Section 2.3.2). 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 2.3.1). Those units of information can characterize a domain, such as a binding site. Particular units of information carry the material type and the conceptual type of the macromolecules. A complex may carry a clone marker (see Section 2.3.3).

![Figure 2.13: An example Process Description glyph for complex.](image)

### 2.4.7 Glyph: Source and Sink

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

For these and other situations, SBGN defines two glyphs that use the same symbol for explicitly representing the involvement of an unspecified source or sink. The symbol used in SBGN is borrowed from the mathematical symbol for “empty set”, but it is important to note that it does not actually represent a true absence of everything or a physical void—it represents the absence of the corresponding structures in the model, that is, the fact that these sources or sinks are conceptually outside the scope of the map. The reason that we regard the Source and Sink as different glyphs is that they have different syntax and semantics (the former only connects to a Consumption arc and the latter a Production arc). This is mainly an issue for software tools and those mapping to and from SBGN from other notations of exchange formats. A frequently asked question is, why bother having an explicit symbol at all? The reason is that one cannot simply use an arc that does not terminate on a node, because the dangling end could be mistaken to be pointing to another node in the map. This is specially true if the map is rescaled, causing the spacing of elements in the map to change. The availability and use of an explicit symbol for sources and sinks is critical.

SBO Term:
SBO:0000291 ! empty set

Container:
A source or sink is 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 (∅). Figure 2.14 illustrates this. The symbol should be linked to one and only one edge in a map.

Label:
A source or sink does not carry any labels.

Auxiliary items:
A source or sink does not carry any auxiliary items.

Figure 2.14: The source and sink glyphs.

2.4.8 Glyph: 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.
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

SBO Term:
SBO:0000405 ! perturbing agent

Container:
A perturbing agent is represented by a modified hexagon having two opposite concave faces, as illustrated in Figure 2.15.

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

Auxiliary items:
A perturbing agent may carry a clone marker (Section 2.3.3).
A perturbing agent can optionally carry one or more units of information (Section 2.3.1). A controlled vocabulary is available to describing the physical characteristic of the perturbing agent (see Section 2.2.4).

![Figure 2.15: The Process Description glyph for perturbing agent.](Image)

2.4.9 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 2.16 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 2.16: An example representation of calcium/calmodulin kinase II.](Image)

Figure 2.17 on the next page 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.
2.5 Referring to other 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 2.7.1) or as an auxiliary unit on the submap. The clone marker can also provide additional reference mechanisms and is discussed below (Section 2.3.3).

2.5.1 Glyph: Tag

A tag is a named handle, or reference, to another EPN (Section 2.4) or compartment (Section 2.6.1). Tags are used to identify those elements in submaps (Section 2.7.1).

SBO Term:
Not applicable.

Container:
A tag is represented by a rectangle fused to an empty arrowhead, as illustrated in Figure 2.18. 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 2.18: The Process Description glyph for tag.

2.6 Defined sets of entity Pool nodes

2.6.1 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 2.3.1. The center of the bounding box of a unit of information is located on the mid-line of the border of the compartment.

![Figure 2.19](image.png)

*Figure 2.19: 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 2.20 shows two semantically equivalent placement of compartments:

![Figure 2.20](image.png)

*Figure 2.20: 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 2.21 on the following page illustrates the problem using an incorrect map.
2.7 Encapsulation

2.7.1 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 2.4). A submap is not equivalent to an omitted process (see Section 2.8.2). 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 2.4). In the unfolded view, exposing the internal structure of the submap, a set of tags point to the corresponding internal EPNs Section 2.4.
Figure 2.22: The Process Description glyph for submap. (Upper part) folded submap. (Lower part) content of the submap.

Figure 2.23 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.

The map in Figure 2.24 on the following page 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.
2.8 Process nodes

Process nodes represent processes that transform one or several entity pools into one or several entity pools, identical or different. SBGN Process Description Level 1 defines a generic process (Section 2.8.1), as well as five more specific ones: the omitted process (Section 2.8.2), the uncertain process (Section 2.8.3), the association (Section 2.8.4), the dissociation (Section 2.8.5), and the phenotype (Section 2.8.6). In future levels of the SBGN Process Description language, more processes may be defined. (One can even envision the development of a controlled vocabulary of processes, as is done now for EPNs; see Section 2.2.)

2.8.1 Glyph: Process

A process transforms a set of entity pools (represented by EPNs in SBGN Process Description Level 1) into another set of entity pools.

SBO Term:

SBO:0000375 ! process

Origin:

One or several consumption arcs (Section 2.9.1) or one or several production arcs (Section 2.9.2).

Target:

One or several production arcs (Section 2.9.2).

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 2.9.1) and production (Section 2.9.2) arcs are linked to the extremities of those connectors. The modulatory arcs (Section 2.9) point to the other two sides of the box. A process connected to production arcs on opposite sides is a reversible process.

Figure 2.25: 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).

A cardinality label may be associated with consumption (Section 2.9.1) or production (Section 2.9.2) 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.

A process is regarded as reversible if both ‘sides’ of the process are connected to production arcs (see section 3.5.2.5).

The example in Figure 2.26 illustrates the use of a process node to represent the phosphorylation of a protein in a Process Description.

![Figure 2.26: Phosphorylation of the protein MAP kinase.](image)

The example in Figure 2.27 illustrates the use of a process node to represent a reaction between two reactants that generates three products.

![Figure 2.27: Reaction between ATP and fructose-6-phosphate to produce fructose-1,6-biphosphate, ADP and a proton.](image)

The example in Figure 2.28 illustrates the use of a process node to represent a translocation. The large round-cornered rectangle represents a compartment border (see Section 2.6.1).

![Figure 2.28: 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.](image)

The example in Figure 2.29 on the following page illustrates the use of a process node to represent the reversible opening and closing of an ionic channel in a Process Description.
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

When such a reversible process is asymmetrically modulated, it must be represented by two different processes in a Process Description. Figure 2.30 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.

The example in Figure 2.31 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.

2.8.2 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 2.9.1) or one or several production arcs (Section 2.9.2).

Target:

One or several production arcs (Section 2.9.2).
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 2.32: The Process Description glyph for omitted process.](image)

2.8.3 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 2.9.1) or one or several production arcs (Section 2.9.2).

**Target:**
One or several production arcs (Section 2.9.2).

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

![Figure 2.33: The Process Description glyph for an uncertain process.](image)

2.8.4 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 2.9.1).

**Target:**
One production arc (Section 2.9.2).

**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 2.9.1) and production (Section 2.9.2) arcs are linked to the extremities of those connectors.
The example in Figure 2.35 illustrates the association of cyclin and CDC2 kinase into the Maturation Promoting Factor.

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). Figure 2.37 gives an example of this, using the formation of hemoglobin.
2.8.5 **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 2.9.1).

**Target:**

One or more production arc (Section 2.9.2).

**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 2.9.5). Moreover, the existence of two circles reminds the dissociation, by contrast with the filled disc of the association (Section 2.8.4).

![Figure 2.38: The Process Description glyph for dissociation.](image)

The example in Figure 2.39 illustrates the dissociation of the small and large ribosomal subunits from a messenger RNA.

![Figure 2.39: Dissociation of the small and large ribosomal subunits from a messenger RNA.](image)

2.8.6 **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 2.40 on the next page.
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 2.3.3).

Figure 2.40: *The Process Description glyph for phenotype.*

### 2.9 Arcs

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

#### 2.9.1 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 2.4).

**Target:**

Any *process node* (Section 2.8).

**End point:**

No particular symbol is used to represent a consumption.

A cardinality label may be associated with *consumption* (Section 2.9.1) or *production* (Section 2.9.2) 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.
2.9.2 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 2.8).

Target:
Any EPN (Section 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 2.42: The Process Description glyph for production.

Figure 2.43 on the following page illustrates the use of consumption/production arc cardinality labels to represent the stoichiometry of a process.
2.9.3 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.

**SBO Term:**

SBO:0000168 ! control.

**Origin:**

Any *EPN* (Section 2.4) or any *logical operator* (Section 2.10).

**Target:**

Any *process node* (Section 2.8).

**End point:**

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

Figure 2.44: *The Process Description glyph for modulation.*

Figure 2.45 on the next page 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.
2.9.4 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 2.9.5.

**SBO Term:**

SBO:0000170 ! stimulation.

**Origin:**

Any EPN (Section 2.4) or any logical operator (Section 2.10).

**Target:**

Any process node (Section 2.8).

**End point:**

The target extremity of a stimulation carries an empty arrowhead.

![Figure 2.46: The Process Description glyph for stimulation.](image)

2.9.5 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 2.4) or any logical operator (Section 2.10).
Target:
Any process node (Section 2.8).

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

![Diagram of catalysis](image)

Figure 2.47: The Process Description glyph for catalysis.

### 2.9.6 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 2.4) or any logical operator (Section 2.10).

**Target:**
Any process node (Section 2.8).

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

![Diagram of inhibition](image)

Figure 2.48: The Process Description glyph for inhibition.

### 2.9.7 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 2.4) or any logical operator (Section 2.10).

**Target:**
Any process node (Section 2.8).

**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.
The example in Figure 2.50 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.)

The example in Figure 2.51 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.

2.9.8 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.
CHAPTER 2. PROCESS DESCRIPTION GLYPHS

2.9.9 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 2.4).

Target:
Tag (Section 2.5.1).

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

2.10 Logical operators

2.10.1 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 2.4) or logical operator (section 2.10).

Target:
One modulation (section 2.9.3), stimulation (section 2.9.4), catalysis (section 2.9.5), inhibition (section 2.9.6) or necessary stimulation (section 2.9.7) arc.
Node:

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

![AND](image)

**Figure 2.54:** The Process Description glyph for *and*. Only two inputs are represented, but more would be allowed.

### 2.10.2 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 2.4) or *logical operator* (section 2.10).

**Target:**

One modulation (section 2.9.3), stimulation (section 2.9.4), catalysis (section 2.9.5), inhibition (section 2.9.6) or necessary stimulation (section 2.9.7) arc.

**Node:**

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

![OR](image)

**Figure 2.55:** The Process Description glyph for *or*. Only two inputs are represented, but more would be allowed.

### 2.10.3 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 2.4) or *logical operator* (section 2.10).

**Target:**

One modulation (section 2.9.3), stimulation (section 2.9.4), catalysis (section 2.9.5), inhibition (section 2.9.6) or necessary stimulation (section 2.9.7) arc.

**Node:**

*Not* is represented by a circle carrying the word “NOT”.

![NOT](image)
Figure 2.56: The Process Description glyph for not.
Chapter 3

Process Description Language Grammar

3.1 Overview

One of the goals of SBGN Process Description Level 1 is to provide a visual description of biological systems that can accurately and unambiguously convey to the reader what the writer of a Process Description map meant. A secondary, but nevertheless important goal is to provide a visual language that can be supported by software tools. To achieve both goals we require a detailed set of rules that are universally applied to Process Description maps. In Chapter 2 the glyphs of the Process Description language were introduced and key usage rules were described; we expect that this will be sufficient for casual users of the notation. However, this chapter provides a more detailed description of the rules that apply to the Process Description language and it is anticipated that this will be required reading for “power users” of the notation and tool developers implementing SBGN Process Description viewers or editors.

In addition to understanding the rules of the notation it is also important to understand the underlying model, or abstraction, of the biological world that the Process Description language uses. By understanding the abstractions used one can also understand more clearly what the author of a map meant, and also what the author could not say given the limitations of the notation. In the next section, we describe the abstractions and concepts implicit in the SBGN Process Description language.

3.2 Concepts

The key abstraction of the SBGN Process Description language is one of processes that act on pools of entities. The entities are typically biological molecules, but need not be (such as a perturbing agent), and the process is typically a combination of one or more biochemical reactions, but again need not be. Processes are controlled, or modulated, by other entity pools and new entities can be obtained from a source and existing entities discarded to a sink.

It can be helpful to think of the SBGN Process Description as describing the flow of a fluid, especially when tying to understand concepts such as stoichiometry, and modulation. In this analogy, each EPN represents a tank of fluid, which may be emptied via a pipe (the consumption arc, Section 2.9.1) that is connected to a valve (the process node, Section 2.8), which in turn is connected to other pipes (the production arcs, Section 2.9.2) that fill other tanks (EPNs). The opening of the valve, and thus the rate of the process can be controlled by the volume of fluid in another tank (this is modulation)\(^1\). If there are two consumption pipes feeding fluid into the process and one is wider and allows double the flow of the first then the fluid will mix in the process with a 2:1 ratio. This corresponds to the stoichiometry of the consumption and production arcs. Finally, the system needs a source of fluid from one or more external sources (source, section Section 2.4.7) and somewhere for it to drain away (the sink).

\(^1\)The precise nature of the relationship between valve opening and amount of fluid determines the nature of the modulation.
One can see that this maps very closely to the abstraction described in the Process Description language, but it also has an interesting side-effect that it allows us to add quantities to SBGN glyphs. In particular, EPNs have an implicit amount of entities in the entity pool and processes have an implicit rate. Since SBGN Process Description is not a modelling language this will not be discussed in any further detail, however, this is an important concept that underpins our understanding of how processes are modulated in Section 3.5.2.4.

### 3.3 The conceptual model

In order to formalize the conceptual representation described above and the semantic rules that we will describe later, we have described the Process Description language using UML Domain Object Model. We have used it to define the “taxonomy” of the Process Description glyphs and their relationship to each other. Finally we describe the attributes of each node glyph that are required to uniquely identify an instance of it in a Process Description map. The concept of identity is important in some of the rules described below.

**Table 3.1:** The Identifying attributes of Process Description node glyphs. When a glyph is always unique in a map, this is indicated by the term instance. The term state values indicates that the values of all the EPN’s state variables are used in the definition of its identity.

<table>
<thead>
<tr>
<th>Glyph</th>
<th>Identifying Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Information</td>
<td>instance</td>
</tr>
<tr>
<td>State variable</td>
<td>Owning StatefulEpnNode, value</td>
</tr>
<tr>
<td>Clone maker</td>
<td>Owning EpnNode</td>
</tr>
<tr>
<td>Labelled clone maker</td>
<td>Owning EpnNode, Label</td>
</tr>
<tr>
<td>Unspecified entity</td>
<td>Owning compartment, Label</td>
</tr>
<tr>
<td>Simple chemical</td>
<td>Owning compartment, Label</td>
</tr>
<tr>
<td>Macromolecule</td>
<td>Owning compartment, Label, Material type, state values</td>
</tr>
<tr>
<td>Nucleic acid feature</td>
<td>Owning compartment, Label, Conceptual type, state values</td>
</tr>
<tr>
<td>Complex</td>
<td>Owning compartment, Label, Subunit Epns, state values</td>
</tr>
<tr>
<td>Simple chemical multimer</td>
<td>Owning compartment, Label, Cardinality</td>
</tr>
<tr>
<td>Macromolecule multimer</td>
<td>Owning compartment, Label, Material type, Cardinality, state values</td>
</tr>
<tr>
<td>Nucleic acid feature multimer</td>
<td>Owning compartment, Label, Conceptual type, Cardinality, state values</td>
</tr>
<tr>
<td>Complex multimer</td>
<td>Owning compartment, Label, Cardinality, Subunit Epns, state values</td>
</tr>
<tr>
<td>Source</td>
<td>Owning compartment</td>
</tr>
<tr>
<td>Sink</td>
<td>Owning compartment</td>
</tr>
<tr>
<td>Perturbing agent</td>
<td>Label</td>
</tr>
<tr>
<td>Tag</td>
<td>Label</td>
</tr>
<tr>
<td>Compartment</td>
<td>Label</td>
</tr>
<tr>
<td>Submap</td>
<td>Label</td>
</tr>
<tr>
<td>Process</td>
<td>Combination of all EPNs connecting to process.</td>
</tr>
<tr>
<td>Omitted process</td>
<td>As process.</td>
</tr>
<tr>
<td>Uncertain process</td>
<td>As process.</td>
</tr>
<tr>
<td>Association</td>
<td>As process.</td>
</tr>
<tr>
<td>Dissociation</td>
<td>As process.</td>
</tr>
<tr>
<td>Phenotype</td>
<td>Label</td>
</tr>
</tbody>
</table>

**Notes**

1. A complex may have a Label or be defined by its subunits. In the case where it has both then all complexes with the same label must also have the same subunit composition.

2. Generally process nodes are unique, but if all EPNs associated with a process are cloned, then the process must be a duplicate as well.
3.4 Syntax

The syntax of the SBGN Process Description language is defined in the form of an incidence matrix. An incidence matrix has arcs as rows and nodes as columns. Each element of the matrix represents the role of an arc in connection to a node. Source (S) means that the arc can begin
at that node. Target (T) indicates that the arc can end at that node. Numbers in parenthesis represent the maximum number of arcs of a particular type to have this specific connection role with the node. Empty cells means the arc is not able to connect to the node.

### 3.4.1 Node connectivity

<table>
<thead>
<tr>
<th>Arc\EPN</th>
<th>macromolecule</th>
<th>simple chemical</th>
<th>unspecified entity</th>
<th>multimer</th>
<th>complex</th>
<th>nucleic acid</th>
<th>tag</th>
<th>source</th>
<th>sink</th>
<th>perturbing agent</th>
<th>submap</th>
</tr>
</thead>
<tbody>
<tr>
<td>consumption</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td>S(1)</td>
</tr>
<tr>
<td>production</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td>T</td>
<td></td>
<td></td>
<td>T(1)</td>
</tr>
<tr>
<td>modulation</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stimulation</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catalysis</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inhibition</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>necessary stimulation</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logic arc</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>equivalence arc</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td>T</td>
</tr>
</tbody>
</table>

**Additional rules**

1. An EPN that is a subunit of a complex can only be linked to a modulation arc.

2. With the above exception, glyphs that are subunits cannot be connect to any arc glyph.
### 3.4.2 Containment definition

By containment we mean that a glyph can be drawn inside the other glyph. This does not necessarily mean that the glyph “belongs” the the containing node, although in some cases it does. In this section the concept of “belonging” is referred to as ownership and you can find node ownership in Section 3.3. There are two glyphs that allow containment: *compartment* and *complex*. The next table describes the relationship between Process Description glyphs and these containers. A + means that the element is able to be drawn within a container. A − means containment is not allowed.

<table>
<thead>
<tr>
<th>Glyph \ Containers</th>
<th>complex</th>
<th>compartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>unspecified entity</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>simple chemical</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>macromolecule</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>nucleic acid feature</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>multimer</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>source</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>sink</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>perturbing agent</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>phenotype</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>tag</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>complex</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>compartment</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>subplot</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>process</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>omitted process</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>uncertain process</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>association</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>dissociation</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>consumption</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>production</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>modulation</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>stimulation</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Continued on next page
CHAPTER 3. PROCESS DESCRIPTION LANGUAGE GRAMMAR

continued from previous page

<table>
<thead>
<tr>
<th>Glyph</th>
<th>Containers</th>
<th>complex</th>
<th>compartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>catalysis</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>inhibition</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>necessary stimulation</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>logic arc</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>equivalence arc</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>and</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>or</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>not</td>
<td>-</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

3.5 Semantic rules

3.5.1 EPNs

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 EPNs\(^2\).

3. EPNs should not be orphaned (i.e., they must be associated with at least one arc.

3.5.2 Process Nodes

As described in Section 2.8.1, the consumption and production arcs converge before connecting to the process node (Figure 3.3). 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 process\(^3\). For convenience we will also collectively refer to the consumption and production arcs as flux arcs.

![Figure 3.3: An illustration of the “sidedness” of a process. The designation of LHS and RHS is essentially arbitrary.](image)

3.5.2.1 Flux Arcs

1. All process nodes (with the exception of phenotype) must have a LHS and RHS.

2. All EPNs on the LHS of a process must be unique.

3. All EPNs on the RHS of a process must be unique.

4. All phenotype glyphs must be associated with at least one modulation arc.

---

\(^2\) A stateful EPN is equivalent if the EPNs are identical when their state descriptions are ignore.

\(^3\) Note this designation is purely for grouping and is used even then the sides of the reaction are above and below the process.
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.

6. Once the stoichiometry of a flux arc is displayed in a map then all other flux arcs should display their stoichiometry make.

7. If the stoichiometry is undefined or unknown this should be indicated by the use of a question mark ("?").

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.

3.5.2.2 Association

1. An Association is always an irreversible process.

3.5.2.3 Dissociation

1. An Dissociation is always an irreversible process.

3.5.2.4 Modulation

As discussed in Section 3.2, 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.

1. A process with no modulations has an underlying “basal rate” which describes the rate at which it converts inputs to outputs.

2. A modulation changes the basal rate in an unspecified fashion.

3. A stimulation is a modulation that increases the basal rate.

4. An inhibition is a modulation that decreases the basal rate.

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.

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.

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.

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 reactions. 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 defined by the catalysis.
3.5.2.5 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 3.4. 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 3.4: A valid reversible process. A process is reversible if its LHS and RHS contain only production arcs.](image)

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 irreversible\(^4\).
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.

3.5.3 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 3.2.

<table>
<thead>
<tr>
<th>Node</th>
<th>Can be duplicated</th>
<th>Indication</th>
<th>Additional Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compartment</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimpleChemical</td>
<td>Y</td>
<td>Simple clone marker</td>
<td></td>
</tr>
<tr>
<td>UnspecifiedEntity</td>
<td>Y</td>
<td>Simple clone marker</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sink</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perturbing Agent</td>
<td>Y</td>
<td>Simple clone marker</td>
<td></td>
</tr>
<tr>
<td>Phenotype</td>
<td>Y</td>
<td>Simple clone marker</td>
<td></td>
</tr>
<tr>
<td>MultimerChemicalEntity</td>
<td>Y</td>
<td>Simple clone marker</td>
<td></td>
</tr>
<tr>
<td>StatefulEntityPool</td>
<td>Y</td>
<td>Labeled clone marker</td>
<td></td>
</tr>
<tr>
<td>Macromolecule</td>
<td>Y</td>
<td>Labeled clone marker</td>
<td></td>
</tr>
<tr>
<td>MultimerMacromolecule</td>
<td>Y</td>
<td>Labeled clone marker</td>
<td></td>
</tr>
<tr>
<td>NucleicAcidFeature</td>
<td>Y</td>
<td>Labeled clone marker</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>Y</td>
<td>Labeled clone marker</td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td>Y</td>
<td>None</td>
<td>Duplication is implied when all EPNs linked to the Process are marked as clones.</td>
</tr>
</tbody>
</table>

\(^4\)A source can only be associated with a consumption arc so this rule does not apply in this case.
### 3.5.4 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 than 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.

### 3.5.5 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 2.7.1. 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 cloned\(^\text{5}\).

---

\(^5\)This 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.
4.1 Introduction

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.

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:

1. requirements, i.e. rules which must be fulfilled by a layout, and

2. recommendations, i.e. rules which should be followed if possible.

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

Those layout 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 [1] and Kaufmann and Wagner [2].

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.

4.2 Requirements

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

4.2.1 Node-node overlaps

Nodes are only allowed to overlap in two cases when they are allowed to contain other nodes — as described in Section 3.4.2. Otherwise, nodes are not allowed to overlap (Figure 4.1 on the following page). This includes the 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 4.1: Nodes must not overlap.](image)

### 4.2 Node-edge crossing

Edges must be drawn on the top of a the node (Figure 4.2). See also recommendation Section 4.3.1.

![Figure 4.2: If an edge crosses a node, the edge must be drawn on top of the node.](image)

### 4.2.2 Node-edge crossing

Edges must be drawn on the top of a node (Figure 4.2). See also recommendation Section 4.3.1.

### 4.2.3 Node border-edge overlaps

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

### 4.2.4 Edge-edge overlaps

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

### 4.2.5 Node orientation

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

### 4.2.6 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 4.6 on page 50).

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.

### 4.2.7 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).
4.2.8 Edge labels

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

4.2.9 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.

4.3 Recommendations

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

4.3.1 Node-edge crossing

Situations where edges and nodes cross should be avoided. Note that some crossings may be unavoidable, 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).

4.3.2 Labels

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

4.3.3 Avoid edge crossings

The amount of crossings between edges should be minimized.

Figure 4.5: The node orientation must be horizontally or vertically.
4.3.4 Branching of association and dissociation

The branching points of association and dissociation nodes should be placed close 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 4.7).

4.3.5 Units of information

Units of information should not hide the structure of the corresponding node and should not overlap other elements (Figure 4.8 on the next page).

4.4 Additional suggestions

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

- Angle of edge crossings: If edge crossing cannot be avoided then the edges should cross with an angle close to 90 degrees.
- Drawing area and width/height ratio: The drawing should be compact and the ratio between the width and the height of the drawing should be close to 1.
- Edge length: Long edges should be avoided.
Figure 4.8: Units of information should not overlap with any other element.

- Number of edge bends: Edges should be drawn with as few bends as possible.
- Similar and symmetric parts: Similar parts of a map should be drawn in a similar way, and symmetric parts should be drawn symmetrically.
- Proximity information: Related elements (e.g. nodes connected by a process or all elements within a compartment) should be drawn close together.
- 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).
- Compartments: It can help clarity to use a different background shade or color for each compartment.
Chapter 5

Acknowledgments

Here we acknowledge those people and organisations the 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.

5.1 Level 1 Release 1.0

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.

5.2 Level 1 Release 1.1

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.

5.3 Level 1 Release 1.2

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.

5.4 Level 1 Release 1.3

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.

5.5 Comprehensive list of acknowledgements

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.

Mirit Aladjemm, Frank Bergmann, Sarah Boyd, Laurence Calzone, Melanie Courtot, Emek Demir, Ugur Dogrusoz, Tom Freeman, Akira Funahashi, Ralph Gauges, Peter Ghazal, Samik
CHAPTER 5. ACKNOWLEDGMENTS

5.6 Financial Support

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 examplifies the use of the EPNs simple chemical, macromolecule, and clone marker, the PNs process, and the connecting arcs consumption, production and catalysis.

![Glycolysis diagram]

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

Figure A.2 on the next page presents an example of signalling pathway, that examplifies 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 on page 56 is an expanded version of the submap present on the map present in Figure A.2 on the next page. It shows the use of tag.

Figure A.4 on page 57 introduces an SBGN Process Description that spans several compart-
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 on the next page.

ments. 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 the “synaptic button” and of “synaptic cleft”. The situation is similar with the compartments “ER” and “muscle cytosol”. The map exemplifies the use of the PN omitted and dissociation, and the connecting arc necessary stimulation.

Figure A.5 on page 57 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.
Figure A.3: A submap of the previous map showing the MAPK cascade.
**Figure A.4:** Neuronal/Muscle signalling. A description of inter-cellular signalling using SBGN.

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

Reference card

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

Issues postponed to future levels

C.1 Multicompartment entities

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:

- 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.

C.2 Logical combination of state variable values

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.

C.3 Non-chemical entity nodes

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.
C.4 Generics

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.

C.5 State and transformation of compartments

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.
Appendix D

Revision History

D.1 Version 1.0 to Version 1.1

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

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

*continued on next page*
APPENDIX D. REVISION HISTORY

---

D.2 Version 1.1 to Version 1.2

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

<table>
<thead>
<tr>
<th>Description</th>
<th>Tracker ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perturbing agent description</td>
<td>2940021</td>
</tr>
<tr>
<td>Members of complex touching</td>
<td>2849273</td>
</tr>
<tr>
<td>PD Reference card error for submap glyph</td>
<td>3029242</td>
</tr>
</tbody>
</table>

D.3 Version 1.2 to Version 1.3

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

<table>
<thead>
<tr>
<th>Description</th>
<th>Tracker ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect editor on title page</td>
<td></td>
</tr>
<tr>
<td>Typos in acknowledgements</td>
<td></td>
</tr>
<tr>
<td>Fixed typo in item on catalysis in section 3.5.2.4.</td>
<td>3090543</td>
</tr>
<tr>
<td>State variables figure 2.6 V1.2</td>
<td>3090543</td>
</tr>
</tbody>
</table>
Bibliography
