

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

**Version 2.0**

Date: January 31, 2012

## **User Manual**



# Chapter 1

## Introduction

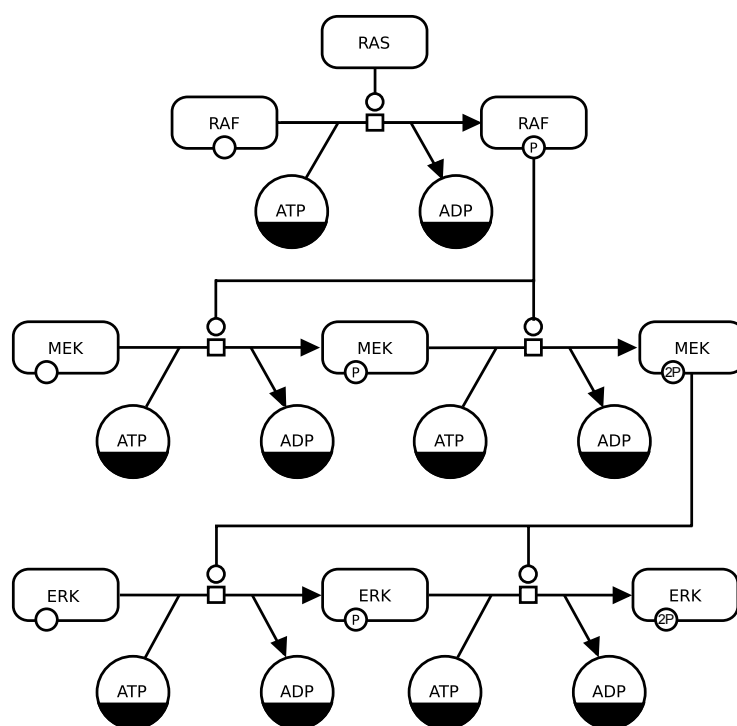
With the rise of systems and synthetic biology, the use of graphical representations of pathways and networks to describe biological systems has become pervasive. It was therefore important to use a consistent notation that would allow people to interpret those maps easily without the need of extensive legends. Furthermore, activities like synthetic biology, that reconstruct biological systems, needs to exchange their description unambiguously, as engineers exchange circuit diagrams.

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. SBGN is made up of three different and complementary languages [1]. This document describes the graphical elements composing the *Process Description language* of SBGN.

### 1.1 Overview of SBGN Process Descriptions

To quickly describe what SBGN Process Description language is about, we let's give a brief overview of some of the relevant concepts with the help of an example shown in Figure 1.1. It 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 1.1, they are either pools of macromolecules or pools of simple chemicals, but as will become clear later in this chapter, they can be other conceptual and material constructs as well. Note also that we speak of *entity pools* rather than individuals; this is because in biochemical network models, one does not focus on single molecules, but rather collections of molecules of the same kind. The molecules in a given pool are considered indistinguishable from each other. The way in which one type of entity is transformed into another is conveyed by a *process node* and arcs between entity pool nodes and process nodes indicate an influence by the entities on the processes. In the case of Figure 1.1, those arcs describe consumption Section ??, production Section ?? and catalysis Section ??, 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



**Figure 1.1:** This example of a Process Description uses two kinds of entity pool nodes: one for pools of different macromolecules (Section ??) and another for pools of simple chemicals (Section ??). Most macromolecule nodes in this map are adorned with state variables (Section ??) representing phosphorylation states. This map uses one type of process node, the process node (Section ??), and three kind of connecting arc, consumption (Section ??), production (Section ??) and catalysis (Section ??). Finally, some entity pool nodes have dark bands along their bottoms; these are clone markers (Section ??) indicating that the same pool nodes appear multiple times in the map.

appearance of the node (Section ??). The details of this and other aspects of Process Description notation are explained in the rest of this chapter.

## 1.2 SBGN levels and versions

It was clear at the outset of SBGN development that it would be impossible to design a perfect and complete notation right from the beginning. Apart from the prescience this would require (which, sadly, none of the authors possess), it also would likely 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. In addition new versions should be backwards compatible, i.e., Process Description maps that conform to an earlier version of the Process Description language within the same level should still be valid. This does not apply to a new levels.

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.

### 1.3 How to get more information

The SBGN website (<http://sbgn.org/>) is a portal for all things related to SBGN.

SBGN discussion list ([sbgn-discuss@caltech.edu](mailto:sbgn-discuss@caltech.edu)). The easiest and best way to get involved in SBGN discussions is to join the mailing list and participate.

SBGN announce list

Face-to-face meetings of the SBGN community are announced on the website as well as the mailing list.

## Chapter 2

# Symbols used in SBGN Process Descriptions

An SBGN Process Description map is mainly a bipartite graph, i.e. it is made up of two types of nodes that connect in an alternate way (some exceptions are described below, e.g. when *logical operators* or *tag* are used). The nodes are the *entity pools nodes*, representing the things that are modified by processes and the *process nodes*. The two types of nodes are connected by arcs. In addition, the *entity pools nodes* can be contained in *compartments*.

### 2.1 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 language contains six glyphs representing classes of material entities: *unspecified entity* (Section ??), *simple chemical* (Section ??), *macromolecule* (Section ??), *nucleic acid feature* (Section ??), and *complex* (Section ??). (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 ??), and *perturbing agent* (Section ??). Material and conceptual entities can optionally carry auxiliary units such as *units of information* (Section ??), *state variables* (Section ??) and *clone markers* (Section ??).

# Bibliography

- [1] N Le Novère, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, M Sarala Wimalaratne, Frank T Bergman, Ralph Gauges, Peter Ghazal, Hideya Kawaji, Lu Li, Yukiko Matsuoka, Alice Villeger, Sarah E Boyd, Laurence Calzone, Mèlanie Courtot, Ugur Dogrusoz, Tom C Freeman, Akira Funahashi, Samik Ghosh, Akiya Jouraku, Sohyoung Kim, Fedor Kolpakov, Augustin Luna, Sven Sahle, Esther Schmidt, Steven Watterson, Guanming Wu, Igor Goryanin, Douglas B Kell, Chris Sander, Herbert Sauro, Jacky L Snoep, Kurt Kohn, and Hiroaki Kitano. The systems biology graphical notation. *Nat Biotechnol*, 27(8):735–741, 2009. 10.1038/nbt.1558.