

# Systems Biology Graphical Notation: Activity Flow Diagram Level 1

Release 1.0

Date: 03 March, 2009

## Editors:

|                   |   |
|-------------------|---|
| Nicolas Le Novère | <i>EMBL European Bioinformatics Institute, UK</i> |
| Huaiyu Mi         | <i>SRI International, USA</i>                     |
| Stuart Moodie     | <i>CSBE, University of Edinburgh, UK</i>          |
| Falk Schreiber    | <i>IPK Gatersleben &amp; MLU Halle, Germany</i>   |
| Anatoly Sorokin   | <i>University of Edinburgh, UK</i> [7pt]          |

## Principal Authors:

|                   |   |
|-------------------|---|
| Nicolas Le Novère | <i>EMBL European Bioinformatics Institute, UK</i> |
| Huaiyu Mi         | <i>SRI International, USA</i>                     |
| Stuart Moodie     | <i>CSBE, University of Edinburgh, UK</i>          |
| Falk Schreiber    | <i>IPK Gatersleben &amp; MLU Halle, Germany</i>   |
| Anatoly Sorokin   | <i>University of Edinburgh, UK</i>                |
| Michael Hucka     | <i>California Institute of Technology, USA</i>    |
| Emek Demir        | <i>MSKCC Computational Biology Center, USA</i>    |
| Yukiko Matsuoka   | <i>The Systems Biology Institute, Japan</i>       |
| Katja Wegner      | <i>University of Hertfordshire, UK</i>            |
| and               |   |
| Hiroaki Kitano    | <i>The Systems Biology Institute, Japan</i>       |

To discuss any aspect of SBGN, please send your messages to the mailing list [sbgn-discuss@sbgn.org](mailto:sbgn-discuss@sbgn.org). To get subscribed to the mailing list or to contact us directly, please write to [sbgn-team@sbgn.org](mailto:sbgn-team@sbgn.org). Bug reports and specific comments about the specification should be entered in the issue tracker [http://sourceforge.net/tracker/?group\\_id=178553&atid=1082245](http://sourceforge.net/tracker/?group_id=178553&atid=1082245).



# Preface

## Acknowledgements

The authors are grateful to all the attendees of the SBGN meetings, as well as to the subscribers of the [sbgn-discuss@sbgn.org](mailto:sbgn-discuss@sbgn.org) mailing list. The authors would like to acknowledge especially the help of Frank Bergmann, Sarala Dissanayake, Ralph Gauges, Peter Ghazal, and Lu Li. SM and AS would also like to acknowledge Igor Goryanin whose financial support and encouragement enabled us to commit the necessary time to the development of this specification. A more comprehensive list of people involved in SBGN development is available in the appendix ??.

The development of SBGN was mainly supported by a grant from the Japanese *New Energy and Industrial Technology Development Organization* (NEDO, <http://www.nedo.go.jp/>). The *Okinawa Institute of Science and Technology* (OIST, <http://www.oist.jp/>), the *AIST Computational Biology Research Center* (AIST CBRC, <http://www.cbrc.jp/index.eng.html>) the British *Biotechnology and Biological Sciences Research Council* (BBSRC, <http://www.bbsrc.ac.uk/>) through a Japan Partnering Award, the European Media Laboratory (EML Research gGmbH, <http://www.eml-r.org/>), and the Beckman Institute at the California Institute of Technology (<http://bnmc.caltech.edu>) provided additional support for SBGN workshops. Some help was provided by the *Japan Science and Technology Agency* (JST, <http://www.jst.go.jp/>) and the *Genome Network Project* of the Japanese Ministry of Education, Sports, Culture, Science, and Technology (MEXT, <http://www.mext.go.jp/>) for the development of the gene regulation network aspect of SBGN, and from the *Engineering and Physical Sciences Research Council* (EPSRC, <http://www.epsrc.ac.uk>) during the redaction of the specification.

## Notes on typographical conventions

The concept represented by a glyph is written using a normal font, while a *glyph* means the SBGN visual representation of the concept.

# Contents

|   |           |  |          |
|---|-----------|--|----------|
| <b>Preface</b>  | <b>ii</b> | <b>2 Activity Flow Diagram glyphs</b>                                    | <b>5</b> |
| <b>1 What is the Systems Biology Graphical Notation?</b>              | <b>1</b>  | 2.1 Overview . . . . .   | 5        |
| 1.1 History of SBGN development . . . . .                             | 1         | 2.2 Controlled vocabularies used in SBGN Activity Flow Level 1 . . . . . | 5        |
| 1.2 The three languages of SBGN . . . . .                             | 2         | 2.3 Entity pool nodes . . . . .  | 5        |
| 1.3 SBGN levels . . . . .   | 3         | 2.4 Container nodes . . . . .  | 5        |
| 1.4 Developments, discussions, and notifications of updates . . . . . | 3         | 2.5 Process nodes . . . . .  | 6        |
|   |           | 2.6 Connecting arcs . . . . .  | 6        |
|   |           | 2.7 Logical operators . . . . .  | 6        |

## Chapter 1

# What is the Systems Biology Graphical Notation?

The goal of the **S**ystems **B**iology **G**raphical **N**otation (SBGN) is to standardize the graphical/visual representation of essential biochemical and cellular processes studied in systems biology. SBGN defines a comprehensive set 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.

Standardizing graphical notations for describing biological interactions is an important step towards the efficient and accurate transmission of biological knowledge between different communities. Traditionally, diagrams representing interactions among genes and molecules have been drawn in an informal manner, using simple unconstrained shapes and edges such as arrows. Until the development of SBGN, no standard agreed-upon convention existed defining exactly how to draw such diagrams in a way that helps readers interpret them consistently, correctly, and unambiguously. By standardizing the visual notation, SBGN can serve as a bridge between different communities such as computational and experimental biologists, and even more broadly in education, publishing, and more.

For SBGN to be successful, it must satisfy a majority of technical and practical needs, and must be embraced by the community of researchers in biology. With regards to the technical and practical aspects, a successful visual language must meet at least the following goals:

1. Allow the representation of diverse biological objects and interactions;
2. Be semantically and visually unambiguous;
3. Allow implementation in software that can aid the drawing and verification of diagrams;
4. Have semantics that are sufficiently well defined that software tools can convert graphical models into formal models, suitable for analysis if not for simulation;
5. Be unrestricted in use and distribution, so that the entire community can freely use the notation without encumbrance or fear of intellectual property infractions.

This document defines the *Activity Flow* visual language of SBGN. As explained more fully in Section 1.2, Activity Flow diagrams are one of three views of a model offered by SBGN. It is the product of many hours of discussion and development by many individuals and groups. In the following sections, we describe the background, motivations, and context of Activity Flow diagrams.

### 1.1 History of SBGN development

Although problems surrounding the representation of biological pathways has been discussed for a long time, see for instance [1], the effort to create a well-defined visual notation was pioneered

by Kurt Kohn with his Molecular Interaction Map (MIM), a notation defining symbols and syntax to describe the interactions of molecules [2]. MIM is essentially a variation of the entity-relationship diagrams [3]. Kohn's work was followed by numerous other attempts to define both alternative notations for diagramming cellular processes (e.g., the work of Pirson and colleagues [4], BioD [5], Patika [6, 7], and others), as well as extensions of Kohn's notation (e.g., the Diagrammatic Cell Language of Maimon and Browning [8]).

Kitano originated the idea of having multiple views of the *same* model. This addresses two problems: no single view can satisfy the needs of all users, and a given view can only represent a subset of the semantics necessary to express biological knowledge. Kitano proposed the development of process diagrams, entity-relationship diagrams, timing charts (to describe temporal changes in a system), and abstract flow charts [9]. The Process Diagram notation was the first to be fully defined using a well-delineated set of symbols and syntax [10]. It led to a desire to establish a unified standard for graphical representation of biochemical entities, and from this arose the current SBGN effort. Separately and roughly concurrently, other groups designed similar notations, for example the Edinburgh Pathway Notation [11] or Patika [6, 7]. All of these efforts began to attract attention as more emphasis in biological research was placed on networks of interactions and not just characterization of individual entities.

In 2005, thanks to funding from the Japanese agency *The New Energy and Industrial Technology Development Organization* (NEDO, <http://www.nedo.go.jp/>), Kitano initiated the Systems Biology Graphical Notation (SBGN) project as a community effort. The first SBGN workshop was held in February 2006 in Tokyo, with over 30 participants from major organizations interested in this effort. From the in-depth discussions held during that meeting emerged a set of decisions that are the basis of the current SBGN specification. These decisions are:

- SBGN should be made up of two different visual grammars, describing Entity Relationship and Process Diagram diagrams (called *State Transition* diagrams at the time). See Section 1.2.
- In order to promote wide acceptance, the initial version(s) of SBGN should stick to at most a few dozens symbols that non-specialists could easily learn.

The second SBGN workshop was held in October, 2006, in Yokohama, Japan. This meeting featured the first technical discussions about which symbols to include in SBGN Level 1, as well as discussions about the syntax, semantics, and layout of graphs. A follow-up technical meeting was held in March, 2007, in Heidelberg, Germany; the participants of that meeting fleshed out most of the design of SBGN. The third SBGN workshop, held in Long Beach in October, 2007, was dedicated to reaching agreement on the final outstanding issues of notation and syntax. The participants of that meeting collectively realized that a third language would be necessary: the Activity Flow diagrams. The specification for the Process Diagram language was finalized and largely completed during a follow-up technical meeting held in Okinawa, Japan, in January, 2008. At this last meeting, attendees also held the first in-depth discussions about the syntax of the Entity Relationship language.

SBGN workshops are an opportunity for public discussions about SBGN, allowing interested persons to learn more about SBGN and help identify needs and issues. More meetings are expected to be held in the future, long after this specification document has been issued.

## 1.2 The three languages of SBGN

Readers may well wonder, why are there *three* languages in SBGN? The reason is that this approach solves a problem that was found insurmountable any other way: attempting to include all relevant facets of a biological system in a single diagram causes the diagram to become hopelessly complicated and incomprehensible to human readers.

The three different notations in SBGN correspond to three different *views* of the same model. These views are representations of different classes of information, as follows:

1. *Process Diagram*: the causal sequences of molecular processes and their results
2. *Entity Relationship*: the interactions between entities irrespective of sequence
3. *Activity Flow*: the flux of information going from one entity to another

In the Process Diagram view, each node in the diagram represents a given *state* of a species, and therefore a given species may appear multiple times in the same diagram if it represents the same entity in different states. Conversely, in the Entity Relationship view, a given species appears only once in a diagram. Process Diagrams are suitable for following the temporal aspects of interactions, and are easy to understand. The drawback of the Process Diagram, however, is that because the same entity appears multiple times in one diagram, it is difficult to understand which interactions actually exist for the entity. Conversely, Entity Relationship diagrams are suitable for understanding relationships involving each molecule, but the temporal course of events is difficult or impossible to follow because Entity Relationship diagrams do not describe the sequence of events.

Process Diagrams can quickly become very complex. Moreover, when diagramming a biochemical network, one often wants to ignore the biochemical basis underlying the action of one entity on the activity of another. A common desire is to represent only the flow of activity between nodes, without representing the transitions in the states of the nodes. This is the motivation for the creation of the Activity Flow view. Activity Flow diagrams permit the use of *modulation*, *stimulation* and *inhibition* and allow them to point to State/Entity nodes rather than process nodes. The Activity Flow view is thus a hybrid between Process Diagram and Entity Relationship diagrams. It is particularly convenient for representing the effect of perturbations, whether genetic or environmental in nature.

A recurring argument in SBGN development is that these three types of diagrams should be merged into one. Unfortunately, each view has such different meanings that merging them would compromise the robustness of the representation and destroy the mathematical integrity of the notation system. While having three different notations makes the overall system more complex, much of the complexity and increase in burden on learning is mitigated by reusing most of the same symbols in all three notations. It is primarily the syntax and semantics that change between the different views, reflecting fundamental differences in the underlying mathematics of what is being described.

### 1.3 SBGN levels

It was clear at the outset of SBGN development that it would be impossible to design a perfect and complete notation right from the beginning. Apart from the prescience this would require (which, sadly, none of the authors possess), it also would likely require a vast language that most newcomers would shun as being too complex. Thus, the SBGN community followed an idea used in the development of the Systems Biology Markup Language (SBML; [12]): stratify language development into levels.

A *level* of SBGN 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. 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. In this way, SBGN development is envisioned to proceed in stages, with each higher SBGN level adding richness compared to the levels below it.

### 1.4 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@sbgn.org](mailto:sbgn-discuss@sbgn.org)) 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.

## Chapter 2

# Activity Flow Diagram glyphs

This chapter provides a catalog of the graphical symbols available for representing entities in Activity Flow diagrams. In Chapter ?? beginning on page ??, we describe the rules for combining these glyphs into a legal SBGN Activity Flow, and in Chapter ?? beginning on page ??, we describe requirements and guidelines for the way that diagrams are visually organized.

### 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 Activity Flow notation with the help of an example shown in Figure 2.1.

Figure 2.1:

### 2.2 Controlled vocabularies used in SBGN Activity Flow Level 1

#### 2.3 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 Diagram 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.

SBGN Process Diagram Level 1 contains five glyphs representing classes of material entities: *unspecified entity*, *simple chemical*, *macromolecule*, *nucleic acid feature*, and *multimer*. (Specific types of macromolecules, such as protein, RNA, DNA, polysaccharide, and specific simple chemicals are not defined by SBGN Process Diagram Level 1 but may be part of future levels of SBGN.) In addition to the material entities, SBGN Process Diagram Level 1 represents four conceptual entities: *source/sink*, *perturbation*, *observable*, and *tag*. Material and conceptual entities can optionally carry auxiliary units such as *units of information*, *state variables* and *clone markers*.

#### 2.4 Container nodes

*Containers* are SBGN constructions that contain one or several other SBGN constructs. There are three container nodes in SBGN Process Diagram Level 1: *complex*, *compartment*, and *submap*.



## 2.5 Process nodes

Process nodes represent processes that transform one or several EPNs into one or several different EPNs. SBGN Process Diagram Level 1 defines a generic *transition* (Section ??), as well as four more specific ones: the *omitted process* (Section ??), the *uncertain process* (Section ??), the *association* (Section ??) and the *dissociation* (Section ??). In future levels of the SBGN Process Diagram notation, more transitions may be defined. (One can even envision the development of a controlled vocabulary of transitions, as is done now for *EPNs*; see Section 2.2.)

## 2.6 Connecting arcs

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

## 2.7 Logical operators

# Bibliography

- [1] Gerhard Michal. On representation of metabolic pathways. *BioSystems*, 47:1–7, 1998.
- [2] Kurt W. Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell*, 10(8):2703–2734, 1999.
- [3] Peter Pin-Shan S. Chen. The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, 1976.
- [4] I. Pirson, N. Fortemaison, C. Jacobs, S. Dremier, J. E. Dumont, and C. Maenhaut. The visual display of regulatory information and networks. *Trends in Cell Biology*, 10(10):404–408, 2000.
- [5] Daniel L. Cook, J. F. Farley, and S. J. Tapscott. A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biology*, 2(4):research0012.1–research0012.10., 2001.
- [6] E. Demir, O. Babur, U Dogrusoz., A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk. Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18(7):996–1003, 2002.
- [7] E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, A. Ayaz, G. Gulesir, G. Nisanci, and R. Cetin-Atalay. An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 20(3):349–356, 2004.
- [8] Ron Maimon and Sam Browning. Diagrammatic notation and computational structure of gene networks. In Hiroaki Kitano, editor, *Proceedings of the 2nd International Conference on Systems Biology*, pages 311–317, Madison, WI, 2001. Omnipress.
- [9] Hiroaki Kitano. A graphical notation for biochemical networks. *BioSilico*, 1:169–176, 2003.
- [10] Hiroaki Kitano, Akira Funahashi, Yukiko Matsuoka, and Kanae Oda. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, 23(8):961–966, 2005.
- [11] Stuart L. Moodie, Anatoly A. Sorokin, Igor Goryanin, and Peter Ghazal. A graphical notation to describe the logical interactions of biological pathways. *Journal of Integrative Bioinformatics*, 3:36, 2006.
- [12] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.