BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 18 2010, pages 2340–2341
doi:10.1093/bioinformatics/btq407

Systems biology

Advance Access publication July 13, 2010

Editing, validating and translating of SBGN maps

Tobias Czauderna[1], Christian Klukas[1] and Falk Schreiber[1,2,*]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben and [2]Institute of Computer Science, Martin Luther University of Halle-Wittenberg, Halle, Germany

Associate Editor: Trey Ideker

ABSTRACT

Motivation: The recently proposed Systems Biology Graphical Notation (SBGN) provides a standard for the visual representation of biochemical and cellular processes. It aims to support more efficient and accurate communication of biological knowledge between different research communities in the life sciences. However, to increase the use of SBGN, tools for editing, validating and translating SBGN maps are desirable.

Results: We present SBGN-ED, a tool which allows the creation of all three types of SBGN maps from scratch or the editing of existing maps, the validation of these maps for syntactical and semantical correctness, the translation of networks from the KEGG and MetaCrop databases into SBGN and the export of SBGN maps into several file and image formats.

Availability: SBGN-ED is freely available from http://vanted.ipk-gatersleben.de/addons/sbgn-ed. The web site contains also tutorials and example files.

Contact: schreibe@ipk-gatersleben.de

Received on March 16, 2010; revised on June 18, 2010; accepted on July 6, 2010

1 INTRODUCTION

The Systems Biology Graphical Notation (SBGN; http://sbgn.org) (Le Novère *et al.*, 2009) is an emerging standard for the visual representation of biological networks. This standardization aims to improve information exchange and to support the collaborative creation of large maps based on individual modules. Such unified graphical representation should overcome the current situation in which many different styles of networks are used in biochemical, biological and medical books, articles and online resources. Especially in the growing field of systems biology, where increasingly complex and large networks have to be investigated, SBGN offers a promising approach for creating more intuitive and unambiguous visualizations.

Three different views of biological processes cover different levels of detail:

(1) *Process Description* (PD), which shows the temporal dependencies of biological interactions in a network in detail;

(2) *Entity Relationship* (ER), which displays the relationships in which a given entity participates in a network (without consideration of temporal aspects); and

(3) *Activity Flow* (AF), which shows the flow of information between biological entities in a network in an abstract way.

*To whom correspondence should be addressed.

Drawing large interaction maps is a time-consuming task. To enhance the usability of SBGN, tool support for creating, editing and validating such maps is required. Currently, several tools support or have started to consider SBGN, for an overview see http://sbgn.org/SBGN_Software.

Here we present SBGN-ED, a Vanted (Junker *et al.*, 2006) add-on for editing, validating and translating of SBGN maps. SBGN-ED contains a unique combination of features from SBGN map creation and editing to syntactic and semantic validation of SBGN maps, to automatic translation of existing non-SBGN networks from KEGG (Kanehisa *et al.*, 2010) and MetaCrop (Grafahrend-Belau *et al.*, 2008) into SBGN, and to the export of SBGN maps in different file and image formats. Furthermore, all methods of the Vanted system for integration and network-based analysis of high-throughput data, data visualization and network layout as well as constraint-based model simulation are fully accessible to the SBGN maps. The Vanted system and the SBGN-ED add-on form a comprehensive platform for SBGN-supported analysis and visualization of biological networks.

2 METHODS

SBGN-ED is an add-on for Vanted (Junker *et al.*, 2006). Vanted is an open source software that offers the loading and editing of networks, allows the integration of different *omics data and provides a variety of functions for data mapping and processing, statistical analysis and visualization. The main functions of SBGN-ED are:

(1) SBGN map creation and editing;

(2) SBGN map validation; and

(3) the translation of networks into SBGN.

These SBGN maps can then be exported into different file and image formats such as GML, GraphML, PDF, SVG, PNG, PNG/HTML image maps and JPG. The following subsections describe the methodology of important functions such as validation and translation.

2.1 Editing

SBGN maps can be created from scratch or loaded from files. The relevant functionality for editing such maps is similar to well-established editing methods for networks (see typical network analysis and graph drawing tools) and allows the addition, removal and alteration of SBGN symbols (nodes and edges), the alteration of the size, color, label and other attributes of map elements, and the alteration of the layout of the map manually or automatically.

Following the SBGN specification, SBGN-ED supports all SBGN symbols (nodes and edges) as well as special connectors defined within glyphs (e.g. ports sticking out of process nodes). These connectors are dynamically adapted during the drawing of a map according to user interaction. For each SBGN language (PD, ER and AF), a side panel contains

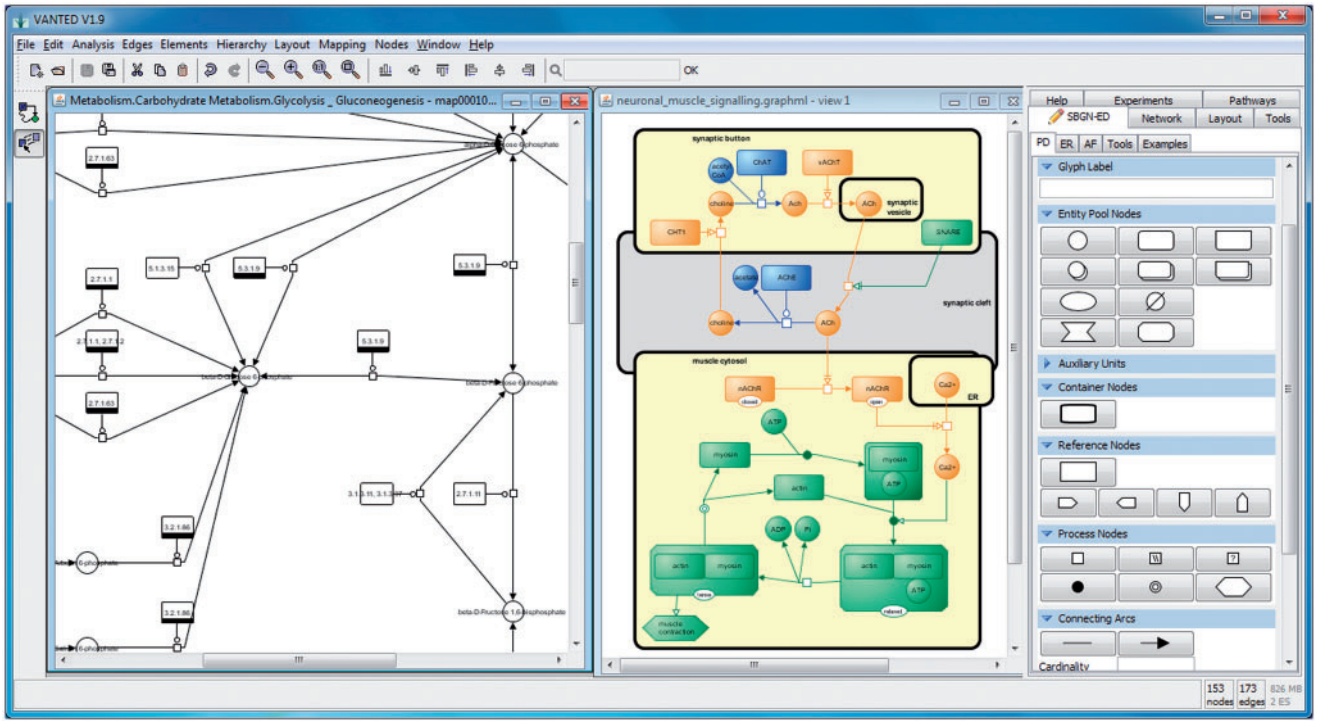

Fig. 1. Screenshot of SBGN-ED as Vanted add-on showing the glycolysis pathway from KEGG translated into SBGN notation (left), an example file for SBGN PD maps (middle) and the side panel for PD maps (right).

the relevant symbols. Figure 1 shows a screenshot of SBGN-ED with two PD maps and the respective panel for editing these type of SBGN maps.

2.2 Validating

The grammar sections of the SBGN languages describe how the SBGN glyphs can be combined and connected to produce valid SBGN maps, which symbols can be duplicated, and so on (see the SBGN specification at http://sbgn.org/Documents/Specifications). These rules are given in the form of incidence matrices with nodes as columns and connections or arcs as rows (each element represents the role of a connection), and some additional textual descriptions. SBGN-ED checks for correct connections depending on the relevant incidence matrix and emphasizes graphically invalid elements (nodes and edges) of a SBGN map for revision. This validation can be applied at any step, e.g. after loading a map or after editing elements.

2.3 Translating

A large number of networks already existing in databases use different symbols to represent biological processes. To help in reusing these networks and avoiding manual redrawing SBGN-ED provides the possibility to translate metabolic pathways from two different sources into SBGN: the KEGG (Kanehisa *et al.*, 2010) and MetaCrop (Grafahrend-Belau *et al.*, 2008) databases. The translation of diagrams from both databases follows several rules derived from the specification to create valid SBGN maps. The left panel in Figure 1 shows a part of the glycolysis pathway from KEGG which was translated into SBGN notation using SBGN-ED.

2.3.1 KEGG to SBGN Metabolic pathways are translated using the KEGG Markup Language (KGML) description of diagrams from the KEGG database. Note that KGML does not always give a complete description of the elements of the KEGG reference image. Elements which cannot be automatically translated into SBGN notation are kept in KEGG notation, and validating the map at the end of the translation is recommended. The number of nodes in the map is typically increased during the translation from KEGG to SBGN due to the SBGN specification. Therefore, node overlapping may occur despite an increased node spacing, which can be removed by node overlapping removal algorithms provided in the tool.

2.3.2 MetaCrop to SBGN Metabolic pathways are translated using GML files from the MetaCrop database.

Funding: German Ministry of Education and Research (BMBF) (under grant 0315430G, in part); Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany.

Conflict of Interest: none declared.

REFERENCES

Grafahrend-Belau,E. *et al.* (2008) MetaCrop - a detailed database of crop plant metabolism. *Nucleic Acids Res.*, **36**, D954–D958.

Junker,B.H. *et al.* (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109.1–109.13.

Kanehisa,M. *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.

Le Novère,N. *et al.* (2009) The systems biology graphical notation. *Nat. Biotechnol.*, **27**, 735–741.