

SOFTWARE

Open Access

# CySBGN: A Cytoscape plug-in to integrate SBGN maps

Emanuel Gonçalves, Martijn van Iersel and Julio Saez-Rodriguez\*

## Abstract

**Background:** A standard graphical notation is essential to facilitate exchange of network representations of biological processes. Towards this end, the Systems Biology Graphical Notation (SBGN) has been proposed, and it is already supported by a number of tools. However, support for SBGN in Cytoscape, one of the most widely used platforms in biology to visualise and analyse networks, is limited, and in particular it is not possible to import SBGN diagrams.

**Results:** We have developed CySBGN, a Cytoscape plug-in that extends the use of Cytoscape visualisation and analysis features to SBGN maps. CySBGN adds support for Cytoscape users to visualize any of the three complementary SBGN languages: Process Description, Entity Relationship, and Activity Flow. The interoperability with other tools (CySBML plug-in and Systems Biology Format Converter) was also established allowing an automated generation of SBGN diagrams based on previously imported SBML models. The plug-in was tested using a suite of 53 different test cases that covers almost all possible entities, shapes, and connections. A rendering comparison with other tools that support SBGN was performed. To illustrate the interoperability with other Cytoscape functionalities, we present two analysis examples, shortest path calculation, and motif identification in a metabolic network.

**Conclusions:** CySBGN imports, modifies and analyzes SBGN diagrams in Cytoscape, and thus allows the application of the large palette of tools and plug-ins in this platform to networks and pathways in SBGN format.

## Background

The illustration of information in biology is frequently made using diagrams. Therefore, having a standard notation is very beneficial as it facilitates interpretation and forgoes the need for long and detailed explanations [1]. Motivated by this, a large community of biochemists, modellers and computer scientists proposed the Systems Biology Graphical Notation (SBGN) [1]. SBGN is an established effort for a standard graphical notation in biology. It leverages previous efforts, including Molecular Interactions Maps [2,3] and Kitano process diagrams [4,5], and extracts the best of graphical standard notations in other fields, such as Unified Modelling Language in software engineering.

SBGN is subdivided into three different and complementary languages: Process Description (PD), Entity Relationship (ER) and Activity Flow (AF). This subdivision allows SBGN to be accurate and assures an unambiguous representation [1]. PD focuses on the representation

of temporal changes occurring between biochemical entities. All molecular processes and interactions and their results are depicted in PD diagrams. In contrast, the ER language depicts the effects among entities, emphasizing the influences that each entity has on each other and disregarding the transformations on itself (which are dealt with in PD diagrams). Finally, AF graphically shows the influence of activities such as inhibition or activation. To better represent these influences the diagram is simplified by ignoring biochemical details of the process and entity states, reducing the number of nodes when compared to an equivalent PD. These three languages together enable the users to represent all types of biological information from biochemical interaction maps to cellular signalling networks.

A graphical notation can only be shared and analysed if it is supported by network visualization tools. SBGN is currently supported by a list of 24 software packages (<http://www.sbg.org/>), including CellDesigner [6], PathVisio [7] and VANTED [8,9]. Cytoscape [10,11], arguably the most popular tool in bioinformatics for visualization and analysis of biological networks, has support

\*Correspondence: saezrodriguez@ebi.ac.uk  
EMBL - European Bioinformatics Institute, Cambridge, UK

for the Systems Biology Markup Language (SBML) [12] natively and via the CySBML plug-in [13], and BioPAX [14] via the BiNoM plug-in [15]. However, it currently has no support for SBGN. The ability to import SBGN diagrams into Cytoscape would be very beneficial, as it would enable the use of all the network analysis features of Cytoscape and its plug-ins in SBGN maps. These include the analysis of the network properties (e.g. shortest paths, motif discovery), visualization of expression data, and pathway modelling.

Motivated by this, we have developed CySBGN, a new plug-in for Cytoscape that enables the full support of SBGN diagrams. CySBGN allows one to import and visualize SBGN diagrams stored in SBGN-ML format using the libSBGN library [16]. It is also possible to export small changes made in the SBGN diagram, such as node positions. Full integration with CySBML [13] and Systems Biology Format Converter [17] is available, allowing the user to automatically generate SBGN diagrams from SBML models. To validate the rendering of the diagrams with the plug-in, we tested it with a suite of SBGN diagrams that covers the three sub languages as well as all its shapes and connections. The plug-in applicability is also demonstrated by applying network analysis methods, in particular shortest path and motif discovery, from two different plug-ins in a PD diagram of the central plant metabolism.

## Implementation

CySBGN enables the importation of SBGN diagrams into Cytoscape and it is compatible with the latest version of Cytoscape, version 2.8.3. All the entity and relationship nodes shapes defined in SBGN specifications for each language are supported.

### Using libSBGN library and SBGN-ML format

The success of a widely used notation is also dependent on the digital storing format, since it determines how easily the notation can be shared and interpreted. Hence, a digital format should be easy to understand, capable of capturing accurately all the information and save it for further access. To face this challenge the libSBGN library and the SBGN-ML format were recently presented [16]. libSBGN is a Java library that provides an API to save, load, query and validate SBGN-ML files, that are structured as an extension of the XML file format. CySBGN makes use of these technologies to display PD, AF and ER diagrams in Cytoscape. The diagrams are loaded and then converted into Cytoscape's network structure. In order to represent the SBGN graphical notation a VizMapper visual style is applied according to the attributes of the nodes and the edges (see Figure 1).

All the information from the SBGN-ML file used to create and display the SBGN diagram is stored as

node and edge attributes. Therefore, information such as node type (e.g. phenotype, macromolecule), node width and edge type (e.g. stimulation, necessary-stimulation) is available to the user through the Data Panel in Cytoscape.

CySBGN stores in the SBGN-ML files the coordinates of all entities in the diagram, allowing the user to export a defined layout. Thus, one can use CySBGN to take advantage of the several Cytoscape automatic layout algorithms and then share the generated layout among other SBGN compliant tools.

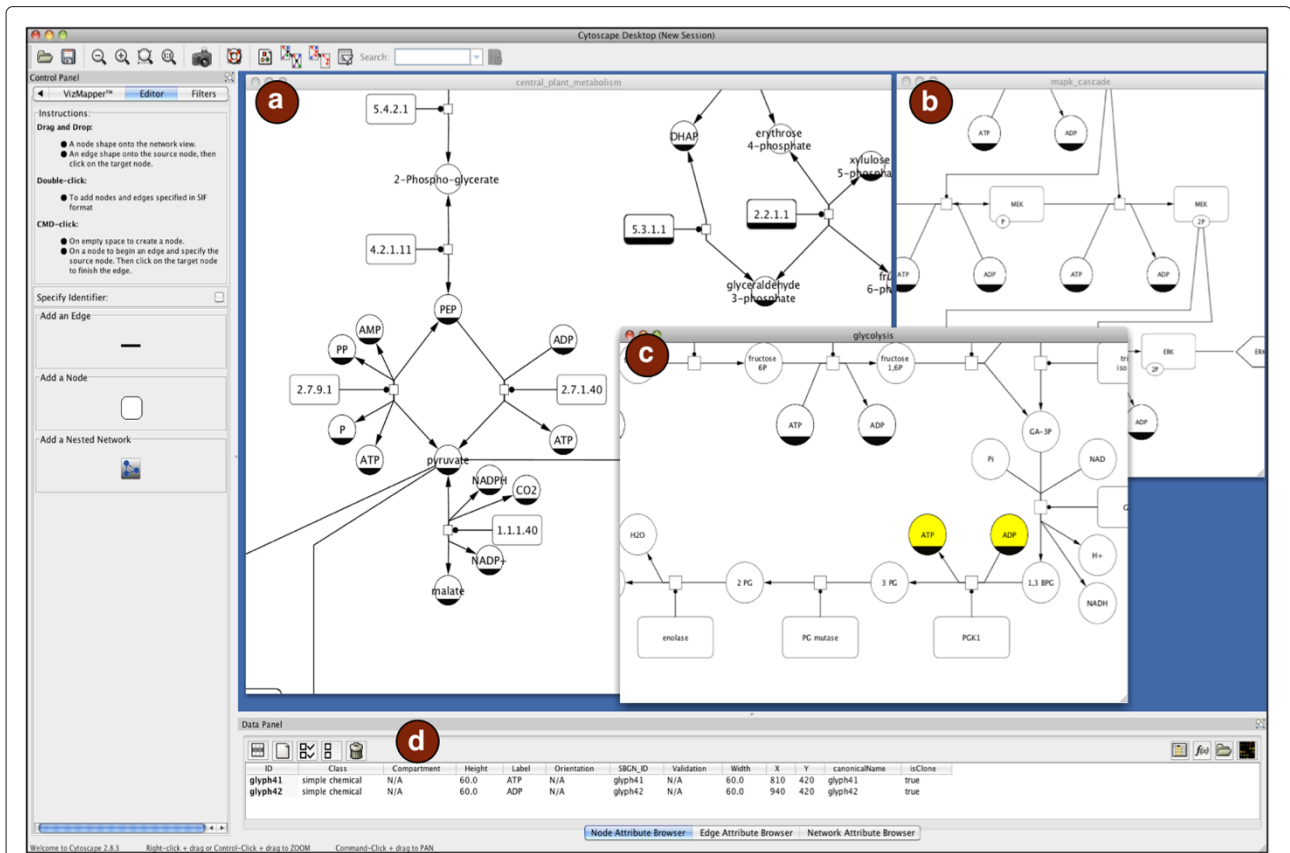
### Implementing SBGN in Cytoscape

Some of the specifications of SBGN are directly representable in Cytoscape, while some need to be represented in different ways (see Figure 2). This comes at the price of more complex networks.

Cytoscape offers nine different node shapes, and therefore it is not possible to represent all the SBGN entities based on this list alone (Figure 2a). To overcome this we use the custom node graphics feature in Cytoscape (introduced in Cytoscape 2.8 [18]), which renders any specific image above the nodes. Thus, CySBGN encompasses all the SBGN shapes by loading a previously created image in the respective type of node.

Similarly to the node shapes, not all the relationship node shapes (i.e. edges) are directly represented in Cytoscape. In particular, absolute stimulation, necessary stimulation, and absolute inhibition relationships lack a representation (Figure 2b). We noticed that the edges that are not directly supported may be split into two sub edges, e.g. the necessary stimulation edge may be subdivided into an inhibition edge followed by a stimulation edge. Taking advantage of this, CySBGN handles all unsupported relationship nodes through a composition of two supported edges linked by a small invisible auxiliary node (Figure 2b). It should be noted that adding auxiliary nodes may affect the network topology analysis methods, since they consider these nodes for analysis even if they do not contain any biological meaning. We address this issue by allowing the user to simplify the network (more details about the network simplification method can be found in the SBGN diagram simplification section). Moreover, the edges layout is not exactly the same as the one adopted by SBGN. The differences are mostly in the filling colour of the edge's shape, which in SBGN is white and in Cytoscape is black. A full comparison between the edge shapes of SBGN and the corresponding shapes in CySBGN is depicted in Figure 3.

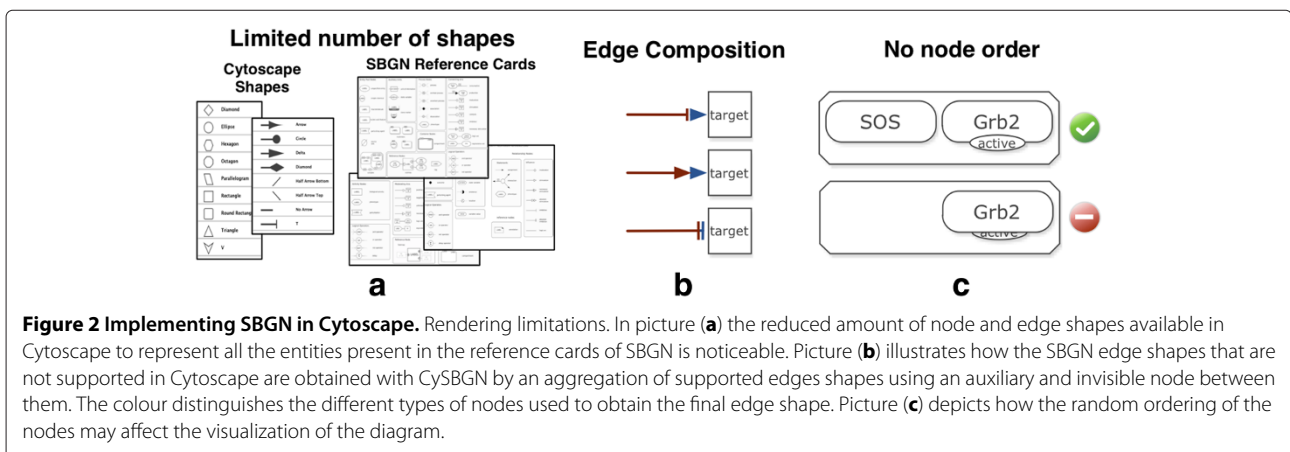
In order to increase the rendering performance Cytoscape disables the rendering order of the nodes. In other words, nodes do not have a *z* coordinate, and therefore when two nodes are overlapping each other it is not possible to identify which node is in the front or



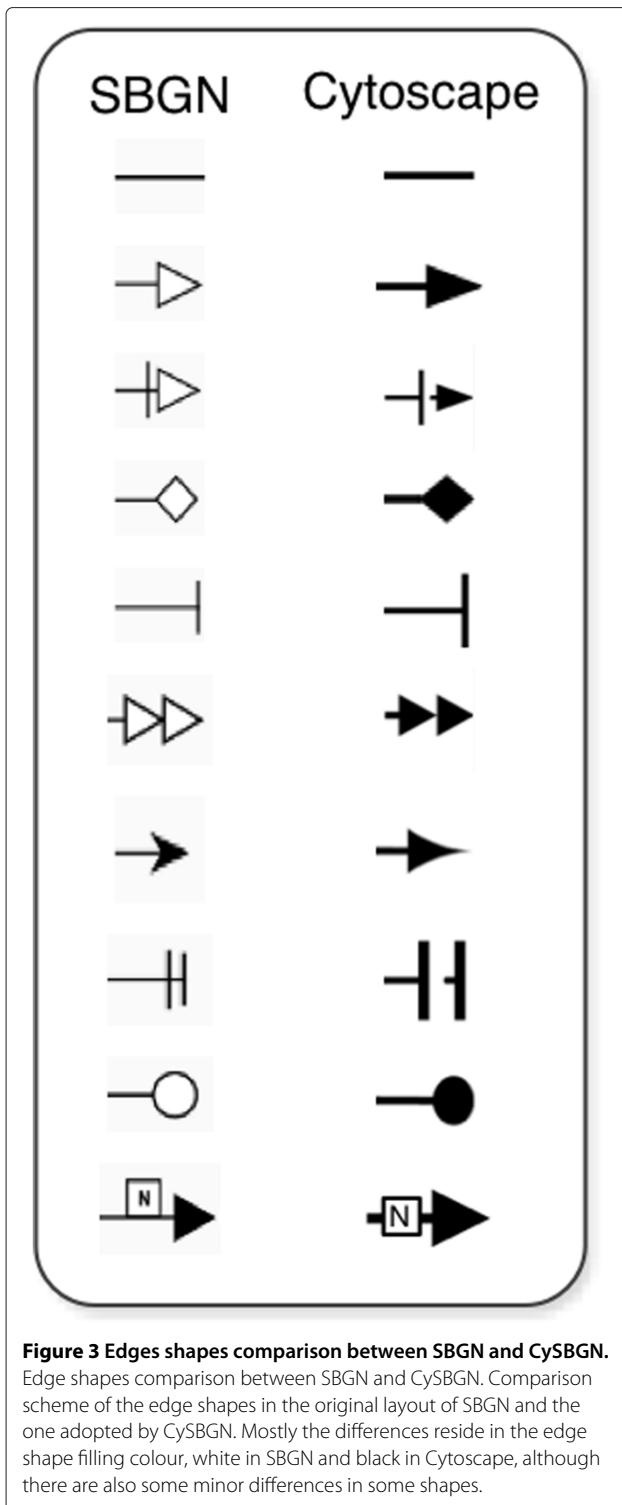
**Figure 1 CySBGN overview.** CySBGN overview. Diagram (a) shows a small part of the imported diagram of the central plant metabolism with 1324 nodes and 1322 edges. Diagram (b) depicts part of the imported MAPK cascade model. Diagram (c) represents the glycolysis metabolic pathway. Panel (d) is the Data Panel of Cytoscape where the user can check the node and edge attributes loaded from the SBGN-ML file. Both diagram (a) and (b) are part of the core validation tests of CySBGN. All these examples are part of the test examples of libSBGN and the examples of SBGN. All the test files are available on CySBGN project web page <http://sourceforge.net/projects/cysbgn/>.

back. Consequently, nodes are randomly ordered (see Figure 2c). This brings some representation problems mostly when the SBGN diagrams contain nodes that enclose other nodes, as is the case of compartments. To avoid rendering the compartment above the enclosed

nodes and therefore hiding them, CySBGN draws every compartment node with transparent filling. This ensures that the enclosed nodes are visible when the compartment is not selected. However, it is not possible to guarantee the transparency of a node when it is selected.



**Figure 2 Implementing SBGN in Cytoscape.** Rendering limitations. In picture (a) the reduced amount of node and edge shapes available in Cytoscape to represent all the entities present in the reference cards of SBGN is noticeable. Picture (b) illustrates how the SBGN edge shapes that are not supported in Cytoscape are obtained with CySBGN by an aggregation of supported edges shapes using an auxiliary and invisible node between them. The colour distinguishes the different types of nodes used to obtain the final edge shape. Picture (c) depicts how the random ordering of the nodes may affect the visualization of the diagram.



### SBGN diagram simplification

Implemented methods in Cytoscape are optimized to work with simple network structures similar to SIF files. Hence, we created a simplification method that becomes useful in cases when the SBGN diagram structure does

not fit the Cytoscape features, such as applying a layout algorithm to a SBGN diagram with cloned entities. In SBGN, a *clone marker* (dark shading in the bottom part of the entities) identifies nodes that appear duplicated in the network. Typical examples of clone entities are ADP or ATP, because they are present in more than one reaction. In these cases the hierarchical layout is wrongly affected by the presence of the cloned entities. Another example is when a necessary stimulation arc is present: due to its unsupported shape CySBGN represents it as an aggregation of two edges (see Results and discussion section for more details). This composition of edges affects the layout analysis of the network. The simplification feature duplicates the original diagram but disregards all nodes that are not supported by Cytoscape, such as compartments, tags or cloned entities. In addition, the edges that need to be drawn as a composition of other edges (i.e. necessary stimulation, absolute inhibition and absolute stimulation) are simplified into a single edge (Figure 2b). Using the simplification feature the diagram information is retained, since the edge type attribute is preserved and only the edge shape is changed. Moreover, the simplification is represented in another diagram keeping the original one unchanged. In the CySBGN tutorial (<http://www.ebi.ac.uk/saezrodriguez/cysbgn/files/tutorial.pdf>) an example of a simplification of a map kinase cascade diagram is provided.

### SBML to SBGN converter

To increase the interoperability of CySBGN we developed an integrated feature between CySBGN and CySBML [13] that enables the users to automatically generate a SBGN diagram based on a SBML model. Taking advantage of the features made available by the SBFC [17], CySBGN generates a SBGN-ML file from a previously selected SBML model and consequently imports it into Cytoscape. This allows the SBML model visualization to be complemented with the respective SBGN diagram, providing the users the possibility to generate the most convenient visualisation of the model, store it, and share it among users and tools. The CySBGN tutorial (<http://www.ebi.ac.uk/saezrodriguez/cysbgn/files/tutorial.pdf>) provides a step-by-step guide of this feature, it also provides information on how to obtain SBML models from the BioModels database [19].

### Results and discussion

The accuracy of diagram rendering in CySBGN was validated by running a varied and extensive number of test cases containing maps from the three complementary languages of SBGN. For each case, we imported a SBGN-ML file and compared the generated diagram with the expected layout. The AF language was tested with 8 diagrams, the ER with 18 and the PD with 27, leading to

a total of 53 different maps, covering almost all shapes and connection types. This suite of test cases is shared among other SBGN compliant tools, such as PathVisio [7] or SBGN-ED [20] (a VANTED [8,9] plug-in which creates and edits SBGN diagrams). Hence, it is possible to compare the results obtained by CySBGN with the original (expected) diagrams and the ones obtained in other tools.

To perform automatically this comparison we developed a python script that generates a html page containing a side-by-side comparison showing the expected pictures of the SBGN maps and those obtained by CySBGN (see Figure 4). The resulting maps are displayed in a matrix structure where the columns correspond to different applications and the rows to the different test cases. The first column refers to the original and correct drawing of the map and the last column is the drawing obtained by CySBGN. Due to its large size the comparison table is available as a supplementary file (Additional file 1) and can also be visualized at [http://libsbn.sourceforge.net/render\\_comparison/](http://libsbn.sourceforge.net/render_comparison/).

CySBML allows users to import SBML models directly from BioModels database [19], visualize and analyse them. We used this feature to test the automatic generation

of SBGN diagrams from the SBML models. This allows one to pick a pathway from BioModels, download the respective SBML model and then generate the SBGN diagram that can be further used as a publication diagram or simply shared with other users. The plug-in's tutorial is available on the CySBGN web page (<http://www.ebi.ac.uk/saezrodriguez/cysbgn/files/tutorial.pdf>) and provides detailed step-by-step guides to all CySBGN features. Additionally, the plug-in's applicability is illustrated by using two different Cytoscape plug-ins to apply network analysis methods to SBGN diagrams. The first example shows the identification of the shortest path between pyruvate and succinate in a large metabolic network using the CytoHubba plug-in (<http://hub.iis.sinica.edu.tw/cytoHubba/>). The second example demonstrates how users can identify network motifs using the NetMatch [21] plug-in. Both examples use a large SBGN diagram of the plant central metabolism (1324 nodes and 1322 edges).

Future steps in CySBGN will focus on: (i) further integration with other Cytoscape plug-ins (e.g. BiNoM [15] integrates a wide variety of structure analysis methods and ways to convert the CellDesigner SBML extension to BioPAX and BioPAX to SBML), and (ii) improving the

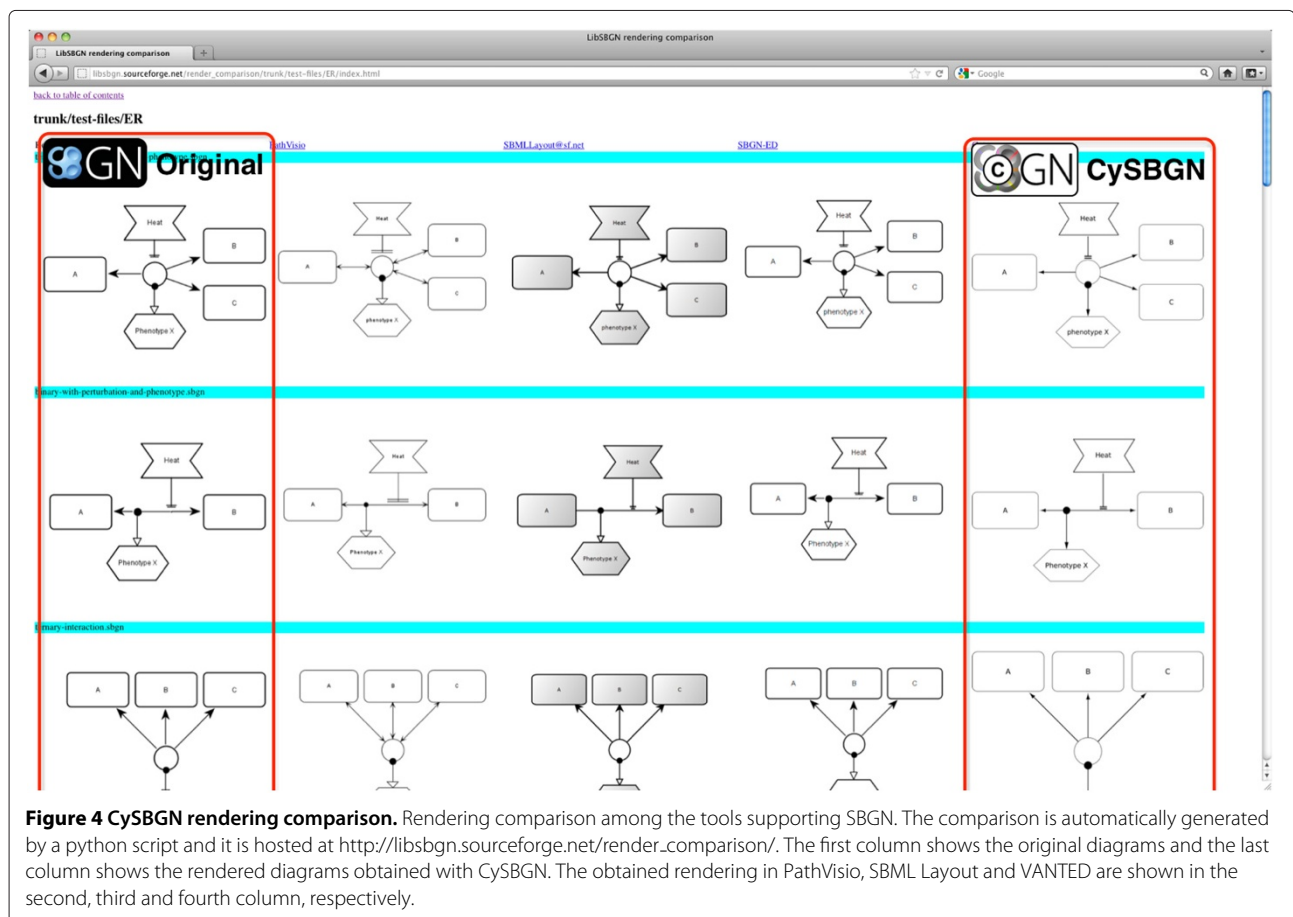


diagram export feature to allow a full mapping of the changes made to the SBGN diagram (i.e. allow the user to add/remove entities and arcs and also allow the storage of any other modifications made to the diagram). Furthermore, we will upgrade CySBGN compatibility to the upcoming version of Cytoscape, Cytoscape 3, when it is released.

## Conclusions

Here we presented a new plug-in for Cytoscape, CySBGN, that provides support to SBGN diagrams. CySBGN allows one to load and visualize SBGN-ML models in Cytoscape and leverage all its repertoire of features and plug-ins. Among other features it also establishes a connection with the CySBML plug-in and SBFC, allowing the automated generation of SBGN diagrams of any imported SBML model. A detailed tutorial containing several step-by-step guides covering all CySBGN features is made available in the plug-in web page (<http://www.ebi.ac.uk/saezrodriguez/cysbgn/files/tutorial.pdf>).

Cytoscape's active community allows the development of the plug-ins to be processed smoothly. Thus, features like extending Cytoscape's supported edges and nodes shapes may be added in near future, consequently improving CySBGN. Moreover, there is an increasing number of plug-ins available [22], thus we expect more synergies of CySBGN with other plug-ins in the future.

## Availability and requirements

- **Project name:** CySBGN
- **Project home page:** <http://www.ebi.ac.uk/saezrodriguez/cysbgn/>
- **Operating system(s):** Platform independent
- **Programming language:** Java
- **Other requirements:** Java 1.6 or higher, Cytoscape 2.8.3
- **License:** GNU GPL v3
- **Any restrictions use by non-academics:** Only those imposed already by the license

## Additional file

**Additional file 1: Supplementary materials 1 CySBGN rendering validation and comparison.** Exhaustive multi-page table containing the rendering comparison of CySBGN with other SBGN compliant tools and the original drawing of the 53 validation test cases proposed in the manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

JSR and MVI proposed the idea, initiated the project and evaluated the software by pointing out errors and improvements. EG and MVI designed the software architecture. EG implemented CySBGN. EG and JSR wrote the paper. All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank Nicolas Rodriguez for helping integrating SBFC library into CySBGN and Francesco Iorio for drawing the CySBGN logo. We acknowledge the financial aid received from the EU through project "BioPreDyn" (ECFP7-KBBE-2011-5 Grant number 289434), and from the LLP/Erasmus Programme.

Received: 11 September 2012 Accepted: 27 December 2012

Published: 16 January 2013

## References

1. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, et al.: **The Systems Biology Graphical Notation.** *Nat Biotechnol* 2009, **27**(8):735–741.
2. Pirson I, Fortemaison N, Jacobs C, Dremier S, Dumont JE, Maenhaut C: **The visual display of regulatory information and networks.** *CELL BIOL* 2000, **10**:1–5.
3. Cook DL, Farley JF, Tapscott SJ: **A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems.** *Genome Biol* 2001, **2**(4):1–11.
4. Kitano H, Funahashi A, Matsuoka Y, Oda K: **Using process diagrams for the graphical representation of biological networks.** *Nat Biotechnol* 2005, **23**(8):961–966.
5. Kitano H: **A graphical notation for biochemical networks.** *BIOFILICO* 2004, **1**(5):1–8.
6. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H: **CellDesigner 3.5: A versatile modeling tool for biochemical networks.** *Proc IEEE* 2008, **96**(8):1254–1265.
7. van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, Conklin BR, Evelo C: **Presenting and exploring biological pathways with PathVisio.** *BMC Bioinf* 2008, **9**:399.
8. Junker B, Klukas C, Schreiber F: **VANTED: A system for advanced data analysis and visualization in the context of biological networks.** *BMC Bioinf* 2006, **7**:109.
9. Rohn H, Junker A, Hartmann A, Grafarend-Belau E, Treutler H, Klapperstück M, Czauderna T, Klukas C, Schreiber F: **VANTED v2: a framework for systems biology applications.** *BMC Syst Biol* 2012, **6**:139.
10. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.** *Genome Res* 2003, **13**(11):2498–2504.
11. Davis DA, Chawla NV: **Exploring and Exploiting Disease Interactions from Multi-Relational Gene and Phenotype Networks.** *PLoS ONE* 2011, **6**(7):e22670.
12. Finney A, Hucka M: **Systems biology markup language: Level 2 and beyond.** *Biochem Soc Trans* 2003, **31**:1–2.
13. König M, Dräger A, Holzhütter HG: **CySBML: a Cytoscape plugin for SBML.** *Bioinformatics* 2012, **28**(18):2402–2403.
14. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, et al.: **The BioPAX community standard for pathway data sharing.** *Nat Biotechnol* 2010, **28**(9):935–942.
15. Zinovyev A, Viara E, Calzone L, Barillot E: **BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks.** *Bioinformatics* 2008, **24**(6):876–877.
16. van Iersel MP, Villéger AC, Czauderna T, Boyd SE, Bergmann FT, Luna A, Demir E, Sorokin A, Dogrusoz U, Matsuoka Y, et al.: **Software support for SBGN maps: SBGN-ML and LibSBGN.** *Bioinformatics* 2012, **28**(15):2016–2021.
17. **Systems Biology Format Converter.** [<http://sbfc.sourceforge.net/>].
18. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2011, **27**(3):431–432.
19. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laipe C: **BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.** *BMC Syst Biol* 2010, **4**:92.
20. Czauderna T, Klukas C, Schreiber F: **Editing, validating and translating of SBGN maps.** *Bioinformatics* 2010, **26**(18):2340–2341.

21. Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader GD, Shasha D: **NetMatch: a Cytoscape plugin for searching biological networks.** *Bioinformatics* 2007, **23**(7):910–912.
22. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T: **A travel guide to Cytoscape plugins.** *Nat Publishing Group* 2012, **9**(11):1069–1076.

doi:10.1186/1471-2105-14-17

**Cite this article as:** Gonçalves et al.: CySBGN: A Cytoscape plug-in to integrate SBGN maps. *BMC Bioinformatics* 2013 **14**:17.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

