

Systems biology

Arcadia: a visualization tool for metabolic pathways

Alice C. Villéger^{1,2,*}, Stephen R. Pettifer² and Douglas B. Kell²

¹School of Chemistry and Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN and ²School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester M13 9PL, UK

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: Arcadia translates text-based descriptions of biological networks (SBML files) into standardized diagrams (SBGN PD maps). Users can view the same model from different perspectives and easily alter the layout to emulate traditional textbook representations.

Availability and Implementation: Arcadia is written in C++. The source code is available (along with Mac OS and Windows binaries) under the GPL from <http://arcadiapathways.sourceforge.net/>

Contact: alice.villeger@manchester.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 5, 2009; revised on March 12, 2010; accepted on April 7, 2010

1 BACKGROUND

Biological models such as metabolic pathways have traditionally been described in textbooks and journal publications via diagrams. Nowadays, these models can be stored in electronic databases and shared as electronic files, using XML-based formats such as the Systems Biology Markup Language (Hucka *et al.*, 2003). In this standardized form, the same data sets can be easily reused in different software applications (e.g. simulation tools, text mining tools, etc.). However, for a human being, the raw content of an SBML file is usually much more difficult to interpret than are traditional diagrammatic representations: even with a good understanding of SBML elements, the model still only appears as a disjointed list of isolated biochemical reactions, with no clear sense of the network they form together. To obtain a map of this network, scientists have to rely on visualization tools.

A number of existing software tools offer (among other features, e.g. model editing or network analysis) to display SBML files as diagrams: e.g. Cytoscape (Killcoyne *et al.*, 2009), CellDesigner (Funahashi *et al.*, 2008), EPE (Sorokin *et al.*, 2006) or VANTED (Junker *et al.*, 2006). Typically, they interpret the SBML model as a graph, which is transformed algorithmically into a diagram by defining a rendering style for the nodes and edges (size, shape, color, etc.), and positioning the resulting graphical objects on a 2D plane. To perform this positioning task automatically, network visualization tools apply a variety of generic layout methods developed by the graph-theory community. However, when dealing with biochemical models, the resulting network maps are often disappointing: numerous edge crossings tend to impair readability, leading in most cases to diagrams that have little in common

with traditional text-book representations of biological pathways. Users often need to perform time-consuming manual adjustments to produce a comprehensible map of their models. To remedy this problem, Arcadia recognizes that diagrams representing biological pathways are not merely generic graphs, but conform to a number of context-specific stylistic conventions that aid their legibility.

2 IMPLEMENTATION

The sole purpose of Arcadia is to display existing SBML files as diagrams. By focusing on this single task, the interface can be kept as simple as possible. Importing and exporting SBML files makes Arcadia interoperable with a large number of tools specializing in other tasks. Arcadia is packaged as a cross-platform desktop application written in C++ and powered by a number of open source libraries: Qt (Nokia Corporation, 2009) for the graphical user interface; LibSBML (Bornstein *et al.*, 2008) to handle SBML files; the Boost Graph Library (Siek *et al.*, 2002) to store the core graph model; Graphviz (Ellson *et al.*, 2002) for graph layout; and libavoid (Wybrow *et al.*, 2006) for edge routing. The source code is available on Sourceforge under the GPL (*cf.* www.gnu.org/licenses/gpl.html), along with precompiled binaries for Windows and Mac OS.

Internally, the data structure can be decomposed into three interconnected layers (Fig. 1). The first layer, or *model layer*, corresponds to the data available in the SBML file, interpreted as a directed bigraph. The last layer, or *geometry layer*, can display graphs as diagrams, according to a specific rendering style and local layout constraints. As explained before, similar mechanisms exist in other network visualization tools. However the middle layer, or *topology layer*, is specific to Arcadia. At this level, the topology of the graph representation of the model can be modified without altering the model itself. This extra layer enables unique features.

3 KEY FEATURES

Node cloning: It means replacing a single node connected to n different other nodes, by n nodes, each connected to only one node. In traditional diagrams, this operation is usually performed on highly connected currency molecules such as adenosine triphosphate (ATP) and adenosine diphosphate (ADP). This simple transformation helps reduce edge crossings and enables greater emphasis to be placed on the overall flow of the pathway.

Neighborhood visualization: An alternate way to reduce visual clutter is to focus only on chemical interactions happening around a specific network hub. In addition to the main view, Arcadia can

*To whom correspondence should be addressed.

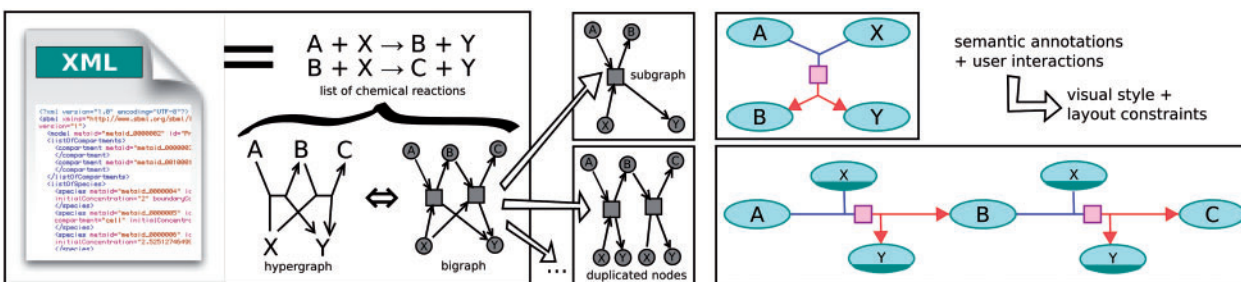


Fig. 1. Arcadia data architecture. From left to right: Model Layer, a set of biochemical reactions equivalent to a graph; Topology Layer, a set of graphs derived from the first graph and Geometry Layer, 2D layout for each graph, rendered according to a given visual style (color, size, shape, font, etc.).

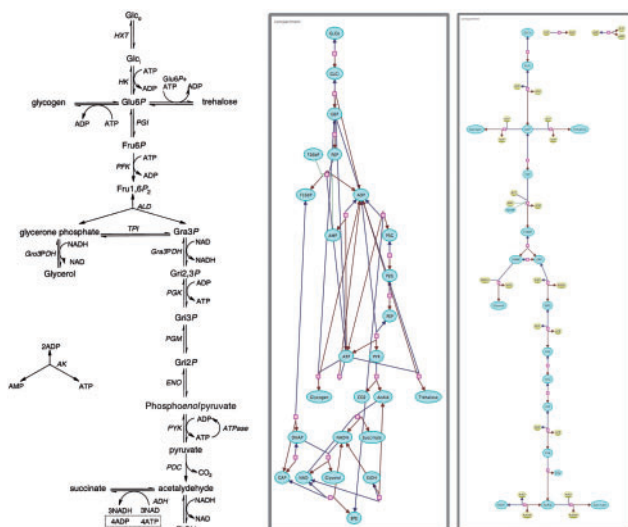


Fig. 2. Comparison of manual, automated and semi-automated layouts. From left to right: hand-drawn diagram, initial layout in Arcadia using Graphviz dot algorithm, and final layout obtained after duplicating highly connected chemicals and branching out secondary chemicals and reactions.

generate complementary views of the core model, displaying all the species located one reaction away from a species of interest.

Graph constraints: At the *geometry* level, it is possible to alter the automatic placement of species and reactions by attaching specific layout rules to certain parts of the graph. This can be used to emphasize particular aspects of the pathway: e.g. the central flux can be made to stand out as a main vertical axis, by placing secondary reactions and species perpendicularly to the overall layout.

Each of the above operations can be performed in a single step. In term of rendering, Arcadia uses the systems biology graphical notation (Le Novère *et al.*, 2009) for process description. By default, Arcadia represents SBML species and reactions as SBGN unspecified entities and transitions. When these SBML objects are annotated with relevant systems biology ontology terms (Le Novère, 2006), Arcadia automatically translates them into more specific SBGN glyphs.

4 RESULTS AND FUTURE PLANS

Figure 2 illustrates results obtained on a published model of yeast glycolysis (Pritchard and Kell, 2002). More generally, in its current

form, Arcadia can deal with networks of up to a few 100 nodes. Diagrams can be saved as standard SBML annotations (Gauges *et al.*, 2006) or exported as vector graphic files (PDF, PS, SVG). Annotated SBML files can be reused in tools supporting the SBML layout extension, e.g. COPASI (Hoops *et al.*, 2006).

Future efforts will focus on supporting more input and output standards, dealing with genome scale networks, and the visual comparison of more than one model.

ACKNOWLEDGEMENTS

We thank the BBSRC for financial support, and members of the MIB (in particular the MCISB) for ideas, test cases and feedback.

Funding: This work was supported by the Biotechnology and Biological Sciences Research Council [BBE0160651 to S.P.].

Conflict of Interest: none declared.

REFERENCES

- Bornstein,B.J. *et al.* (2008) LibSBML: an API library for SBML. *Bioinformatics*, **24**, 880–881.
- Ellson,J. *et al.* (2002) Graphviz - open source graph drawing tools. *Lect. Notes Comput. Sci.*, **2265**, 594–597.
- Funahashi,A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. In *Proceedings of the IEEE*, Vol. 96. pp. 1254–1265.
- Gauges,R. *et al.* (2006) A model diagram layout extension for SBML. *Bioinformatics*, **22**, 1879–1885.
- Hoops,S. *et al.* (2006) COPASI - a CComplex Pathway Simulator. *Bioinformatics*, **22**, 3067–3074.
- Hucka,M. *et al.* (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Junker,B.H. *et al.* (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinfo.*, **7**, 109.
- Killcoyne,S. *et al.* (2009) Cytoscape: a community-based framework for network modeling. *Methods Mol. Biol.*, **563**, 219–239.
- Le Novère,N. (2006) Model storage, exchange and integration. *BMC Neurosci.*, **7** (Suppl. 1), S11.
- Le Novère,N. *et al.* (2009) The Systems Biology Graphical Notation. *Nat. Biotechnol.*, **27**, 735–741.
- Pritchard,L. and Kell,D.B. (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *Eur. J. Biochem.*, **269**, 3894–3904.
- Siek,J. *et al.* (2002) The Boost Graph Library: user guide and reference manual. Addison-Wesley Professional, Boston.
- Sorokin,A. *et al.* (2006) The pathway editor: a tool for managing complex biological networks. *IBM J. Res. Dev.*, **50**, 561–573.
- Wybrow,M. *et al.* (2006) Incremental connector routing. *Lect. Notes Comput. Sci.*, **3843**, 446–457.