LibSBGN: electronic exchange of SBGN maps
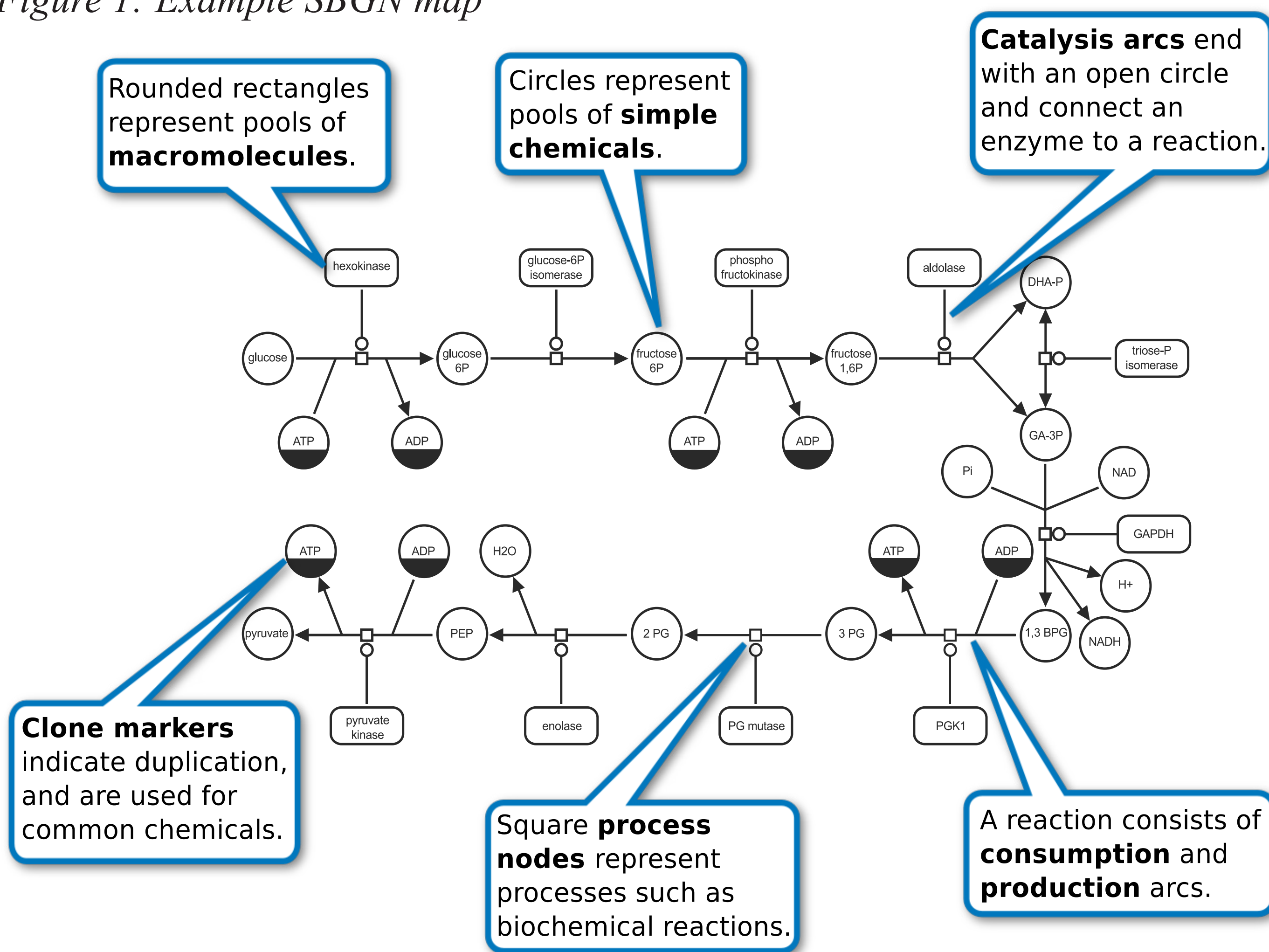
M.P. van Iersel[1,2], S.E. Boyd[3], F. Bergmann[4], S. Moodie[5], F. Schreiber[6,7], T. Czauderna[6], E. Demir[8], N. Le Novère[9], A. Sorokin[5], H. Mi[10], A. Luna[11,12], U. Dogrusoz[13], Y. Matsuoka[14], A. Funahashi[15], H. Kitano[14,16,17], M.I. Aladjem[11], M.L. Blinov[18], A.C. Villéger[19,20]

1) Department of Bioinformatics BiGCaT, Maastricht University, Maastricht, The Netherlands. 2) Netherlands Consortium for Systems Biology (NCSB), The Netherlands. 3) AgriBio, La Trobe University, Melbourne, Australia. 4) Department of Bioengineering, University of Washington, Seattle, Washington, USA. 5) School of Informatics, University of Edinburgh, Edinburgh, UK. 6) Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. 7) Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany 8) Memorial Sloan Kettering Cancer Center - Computational Biology Center, New York, NY, USA. 9) EMBL European Bioinformatics Institute, Hinxton, UK. 10) SRI International, Menlo Park, California, USA. 11) National Cancer Institute, Bethesda, Maryland, USA. 12) Department of Bioinformatics Boston University, Boston, MA, USA 13) Bilkent Center for Bioinformatics, Bilkent University, Ankara, Turkey. 14) The Systems Biology Institute, Tokyo, Japan. 15) Department of Biosciences and Informatics, Keio University, Yokohama, Japan. 16) Okinawa Institute of Science and Technology, Okinawa, Japan. 17) Sony Computer Science Laboratories, Tokyo, Japan. 18) Department of Computational Biology, University of Pittsburgh School of Medicine, USA. 19) School of Computer Science, University of Manchester, Manchester, UK. 20) Manchester Interdisciplinary Biocentre, Manchester, UK.

What is SBGN?

The Systems Biology Graphical Notation (SBGN, See also *Le Novère et al., 2009*) is a standard notation for the visual representation of biochemical and cellular processes. SBGN maps can be used to exchange complex biological knowledge in a clear, concise and unambiguous manner. See figure 1 for an example of the SBGN Process Description (PD) language.

Figure 1: Example SBGN map

Rounded rectangles represent pools of **macromolecules**.

Circles represent pools of **simple chemicals**.

Catalysis arcs end with an open circle and connect an enzyme to a reaction.

Clone markers indicate duplication, and are used for common chemicals.

Square **process nodes** represent processes such as biochemical reactions.

A reaction consists of **consumption** and **production** arcs.

What is LibSBGN?

As SBGN is becoming adopted more widely, there is a need for exchanging maps electronically between the various systems biology tools. Exchange using graphics-only file formats (such as SVG) is often insufficient, because biological meaning of elements is lost. There is a need for a toolset to exchange maps while preserving biological meaning and relations.

Figure 2: SBGN-ML example

There are only two top level elements: **glyph** and **arc**.

Both arcs and glyphs have a **class attribute**, which determines both the biological meaning and the appearance of that glyph.

Glyphs have an optional **label**.

In the current version, only the **bounding box** of glyphs is stored. Further details, such as e.g. the roundness of corners, are left for future versions.

The connection between arcs and glyphs is stored using a system of **identifiers** and **references**.

Implementation

To meet this need, we are defining a new markup language named SBGN-ML. This file format is based on an XML Schema definition (XSD), and currently only covers the Process Description language. See figure 2 for an example.

In addition we are developing a supporting software library called LibSBGN. Besides reading and writing SBGN-ML files, this library will also be used to validate SBGN maps against the specifications, and convert to- and from related systems biology standards, such as SBML (with layout extension) and BioPAX. LibSBGN is currently implemented in Java, a parallel C++ version is planned for the future. LibSBGN is still under development, but it is already being adopted by several tools (see the project wiki for an up-to-date list).

A test suite has been created; it consists of dozens of reference maps covering every aspect of SBGN PD. Corresponding reference documents have been encoded in XML. To test the adherence of supporting tools to the SBGN-ML specification, a rendering comparison pipeline was set up (see figure 3); it automatically renders SBGN maps through different programs, thus enabling comparison against the reference map.

Figure 3: Renderer comparison

Currently, **Three pathway tools** support conversion to the LibSBGN file format, namely PathVisio, Render extension and SBGN-ED.

Reference maps have been made for all aspects of SBGN.

By comparing three implementations, we can root out **ambiguities** and tool-specific **artifacts**.

Some features are **optional**. For example, not all tools can do edge routing. The biological meaning of the map is the same regardless.

Converting in each tool automatically helps detect **errors**.

Community & Infrastructure

LibSBGN is a community driven effort, involving people from institutes all over the world, representing a wide selection of tools. To organize this community, we have set up the following infrastructure:

- A wiki for documentation, bug tracker and code repository available at SourceForge: http://libsbgn.sourceforge.net
- Mailing list: sbgn-libsbgn@lists.sourceforge.net
- Monthly online meetings, open to any interested party.

Roadmap

Milestone 1:
- Implement semantics of Process Description language Level 1 Version 1.1.
- Only high-level graphics specification.
- XML Schema available for validation.

Milestone 2:
- Implement semantics for all three languages: PD, ER and AF

Milestone 3:
- Support for detailed graphics specification

Future plans
- Complete validation
- Conversion to and from SBML (with layout extension)
- Conversion to and from BioPAX
- Implementation of LibSBGN (object model & API) in C++

Acknowledgements

References

N. Le Novère et al., **The Systems Biology Graphical Notation**, *Nature Biotechnology* 2009, 27(**8**):735-41