Developers Propose SBGN as 'Standard Notation' for Systems Biology Collaboration

August 14, 2009

Byline: Vivien Marx
Newsletter: BioInform
BioInform - August 14, 2009

An international group of scientists is trying to spread the word about the Systems Biology Graphical Notation, an effort to standardize the visual representation of diagrams of biochemical and cellular processes.

SBGN was launched in 2005 as an offshoot of the Systems Biology Markup Language. This week, the project published a paper describing SBGN in *Nature Biotechnology* and released the first version of three languages that underlie the standard.

Nicolas Le Novère, a researcher at the European Bioinformatics Institute and the paper's first author, told *BioInform* this week that the developers view the publication as a chance to move SBGN beyond the computational systems biology community.

"The paper is more to distribute it and make it known in a non-specialist community," he said. "People working on pathways or networks were already aware of SBGN."

Systems biology stands to gain through the use of a standardized visual language with a set of symbols and rules that define their use, by allowing "more efficient and accurate transmission of knowledge," the authors said in the paper.

They note that such a standard is sorely needed. Unlike the electronics industry's standard notation for circuit diagrams, no standard visual language exists in biology for describing interaction networks, cell signaling, or gene regulation, they wrote.

Allan Kuchinsky, principal project scientist for molecular tools at Agilent Laboratories, told *BioInform* via e-mail this week that he agrees that SBGN will help scientists "more effectively collaborate" by making it easier to exchange models and data. Kuchinsky was not involved in the development of SBGN.

"It will also help scientists reproduce and corroborate each others' work because they will have a mutually understood language for representing biological data upon which different software tools can interoperate," he said.

Kuchinsky also agreed with the SBGN developers' use of the electronics industry as a model for the systems biology community to emulate. Molecular biology "doesn't have the history of development of visual formalisms" that underlie disciplines such as electrical engineering, he said. "The SBGN undertaking is a significant step towards developing such visual formalisms."

Le Novère said that a key goal for the SBGN developers is to make the standard easy to adopt.

He said he believes in the adage that a standard will only take hold "when it is easier to use than not to use."

To draw users in, the group is developing toolkits such as "ready-made" templates for PowerPoint and graphics software.

There is also an ongoing initiative to develop software libraries based on SBGN. "This is a good sign, when people start to develop projects" based on the standard, Le Novère said.

Kuchinsky explained that it will help if "*de facto* standard software tools" that scientists already use support SBGN. Since several of the paper's authors already work on popular systems biology tools, it's likely that "SBGN will be exposed to a wide range of users," he said.

Another tool that could help SBGN gain acceptance in the community is the Cytoscape network visualization software, to which Agilent contributes, he said.

Kuchinsky said that the Cytoscape development team is tracking the progress of SBGN and plans to implement SBGN support in the Cytoscape software "at the appropriate time."

For the time being, Cytoscape is focusing upon the re-architecting for version 3.0. [*BioInform*, July 17, 2009] The team "will complete that before we undertake support of SBGN," Kuchinsky said.

When it is up and running, that functionality would allow Cytoscape to interoperate with software tools that support SBGN, for example quantitative modeling tools. "These synergies can help broaden the range of biological problems that can be addressed by computational tools, which in turn will help SBGN gain acceptance in the scientific community," Kuchinsky said.

Three Components

SBGN is made up of three different "orthogonal and complementary" diagram types, said Le Novère: process diagrams, which present molecular processes and interactions between biochemical entities; entity relationships; and activity flow.

Process diagrams might work best for researchers working on metabolic networks, he said, while entity relationships emphasize the "influences" that entities — such as molecules, networks, or even clinical conditions — have on each other. "People doing signaling use activity flow," he said.

Right now, the SBGN team is correcting the process diagram specification and is developing the next version of that spec with features such as the ability to include annotation on a graph. Specifications for entity relationships and activity flow are slated for release this year, he said.

A large use case for SBGN is the Human Metabolic Reconstruction Jamboree, a venture to create a "community consensus reconstruction" of the human metabolic network, he said. It is a follow-on project to Yeastnet, a consensus yeast metabolic network reconstruction led by the Manchester Centre for Integrative Systems Biology, which was created in Systems Biology Markup Language, SBML. That format is computer readable and developed to help represent models of biological processes.

This model is "huge," said Le Novère, who is part of the project. "There is no way a single researcher can write down this model" and explain it to another researcher, he said. With a notation such as SBGN, scientists will be able to study another team's model without encountering symbol ambiguities and other questions, which would require extensive personal communication with the original diagram makers months or even years after it was drawn up.

Hack and Bash

The SBGN group meets twice a year for a working meeting and hackathon, "where we develop, bash the language, and try to break it, [to] find the problems," he said.

The language is intended for human reading, but can be used to help computer-based information exchange as well, he said.

Hiroaki Kitano, a co-author of the paper, is "the initiator of the whole project," Le Novère said. He first proposed such a standard in 2005 in a paper in *Nature Biotechnology* that described how "process diagrams" could be used to graphically annotate biological networks.

Le Novère cited Kitano's CellDesigner, a diagram editor for biological networks, as an example of how SBGN could be used in practice.

Le Novère said that CellDesigner has had such "great success" not because it is "so pleasing to the eye" or "well-designed," but because it is "really, really easy to use."

SBGN builds on other efforts in the visualization tool domain, such as Molecular Interaction Maps, or MIM, developed by paper co-author Kurt Kohn of the National Cancer Institute's Laboratory of Molecular Pharmacology.

Filling a Gap

The authors of the *Nature Biotechnology* paper note that a standard for biological notation is long overdue. "The closest to a standard is the notation long used in many metabolic and signaling pathway maps, but in reality, even that lacks uniformity between sources and suffers from undesirable ambiguities," they write. "Moreover, the existing tentative representations, however well crafted, were ambiguous, and only suitable for specific needs, such as representing metabolic networks or signaling pathways or gene regulation."

While scientists can draw *ad hoc* graphical notations, for example an arrow between two elements, this introduces "greater ambiguity" into the diagram, the scientists said. The biggest drawback is inconsistency for the reader as well as problems for any kind of automated diagram verification.

A directional arrow, for example, can mean many things, including "associates into," "disassociates into," "is transformed into," or "catalyzes the formation of," the scientists said. As an example, they document eight different meanings that could be associated with the same symbol in a chart about cell cycle regulation by antibody and reagent supplier Abcam.

In response to this problem, the SBGN developers propose a standardized set of glyphs or symbols, such as representations of macromolecules or phosphorylation states, as well as logical operators.

Symbols that work in SBGN may not be obvious to some biologists used to traditional methods of representation. For example, SBGN does not depend on relative positions of symbols or certain visual features — such as colors, patterns, shades, and thickness of edges — which could make rescaling problematic or disappear in a fax or a photocopy.

Colors are fine, but only if they do not change the semantics of a given diagram, Le Novère said. For example, he said it's common to use a green arrow for stimulation and red arrows for inhibition. But if that information is lost when a graph is photocopied, then "it's not a good language," he said.

The difference between the use of diagrams in biology and engineering is not just related to tradition, Kuchinsky said. An engineer's diagram specifies a "system to be built according to a set of constraints and abstractions that are universally agreed to, developed over time, and backed by statistics," he said, whereas in biology it's a model for a living system for which many mechanisms are not yet fully understood.

Diagrammatic representations of biological pathways serve as representations for an increasingly diverse community of practitioners, including biologists, biochemists, computer scientists, mathematicians, Kuchinsky said.

"Different disciplines bring different perspectives to the problem, in particular the way computer scientists and biologists think about network representations may be quite different at times," he said. "The co-authors of this article are representative of this diversity of perspectives and they have taken a community-based approach to bring together these points of view."

Genomeweb system

These settings are generally managed by the web site so you rarely need to consider them.
Issue Order: 1

-->

Drupal