Use Case Descriptions / Comments

Name - Elgar Pichler
Organization - AstraZeneca
Project - Interaction Database, Gene List Analyis Platform

“We would like to see data consolidation of gene/protein association data in as few canonical reference sources as possible. use case 1: update AZ proprietary interaction database with non-redundant, cleaned-up, semantically typed gene/protein associations from the public domain use case 2: exploring a signaling/metabolic pathway obtain a reacting network topology and a null hypothesis set of reaction parameters from BioPAX system; feed this into reaction network modeling software; iteratively validate/update model based on experimental results; possibly store and manage versions of model in BioPAX system.”

Name - Kam Dahlquist
Organization - Vassar College
Project - GenMAPP

The paragraph below was taken from a NSRA Postdoctoral Fellowship application by Alex Pico in the Conklin GenMAPP group. We are now planning to computationally define interaction types according to standards compatible with GenMAPP, Cytoscape and BioPAX. The Simple Interaction Format (SIF) is implemented in Cytoscape as a basic input format that provides a single field to describe the type of interaction. Most interaction data available in SIF are limited to two types, commonly denoted “pp” (protein-protein) and “pd” (protein-DNA). The main advantage of SIF is that it is simple to build, parse, and maintain, even by hand, making it feasible to translate the archive of MAPP pathways at GenMAPP.org into SIF interactions. The variety of interaction types represented on our pathways can be captured by defining additional interaction codes. Using the roadmap for BioPAX as general guide, we will define new codes that will cover current GenMAPP pathways and facilitate future conversion to a mature BioPAX standard. The newly defined interaction types will include activation and inhibition by modulation (e.g., phosphorylation) or binding (e.g., protein or ligand), activation and inhibition of transcription (e.g., transcription factor binding), reaction catalysis (e.g., cAMP production), and groupings (e.g., complexes, paralogs, or isomers). In addition, SIF (and eventually BioPAX) versions of other pathway resources will be made, providing GenMAPP users with the building blocks for generating new MAPPs.

Name - Frank Gibbons
Organization - Harvard Medical School
Project - BioGraphNet

We're engaged in building infrastructure to facilitate distribution of "interaction" data, where the quoted term is defined as loosely as possible: any relationship between two or more entities, encompassing physical interactions, genetic interactions (e.g., epistasis), but also sequence homology, co-expression, cellular co-localization, literature cocitation, etc. We focus on fulfilling two needs: 1) the middle-ground in data distribution, between the web-based one-gene-at-a-time and the whole-database approach; and 2) automating data distribution to the maximum extent possible. The two overlap, of course. We see web-services technology as being the best available solution, BioPAX as a potential common language for data interchange, particularly with the development of Level 3. We have a working prototype that is NOT based on BioPAX, but with plans to re-factor it based on what we've learned, BioPAX is of great interest to us.

Name - Michael Hucka
Organization - Caltech
Project - SBML

One of the things we've talked about is finding ways of connecting/linking/cross-referencing SBML and BioPAX representations. That's what I would like to work towards.

Name - Andrea Splendiani
Organization - Currently: Univ. Milano Bicocca / Genopolis
Project - Pathway profiling. To be defined.

My aim is to used BioPAX as a standard/representation on which to build a pathway analysis process.

Name - Emek Demir
Organization - Bilkent University
Project - PATIKA

I expect that a biologist who wants to adopt a system-level approach starts by constructing a large network of cellular processes shown to be related to lung cancer. This can span multiple databases and information sources. They do this by specifying queries from a single common interface and obtain a BioPAX model. They then add their knowledge and data into it, and visualize and analyze the resulting model using BioPAX compliant editors. Ontology of the model should be supporting concepts that are common in current biology literature, such as molecular and cellular states, pathways and generalizations. The ontology must also support multiple levels of detail so that the users can add their own data without discarding or obscuring the existing model. Users will like to share and integrate their model with their colleagues, especially if they are in a large distributed research community (e.g. European NoEs or AFCS of United States). This can be done in two different levels, in increasing difficulty. At the first level, they simply check-in to CVS. At the second level, there is a BioPAX compliant server, which provides identity and concurrency control services at the graph level, using optional BioPAX fields. They often will couple this model with high-throughput data, trying to figure out how the changes in the genotype led to the phenotype they are observing, specify complex queries to test their hypothesis or come up with new targets for drugs, or identifying drug combinations. There will be tools/services for such tasks, where BioPAX models are coupled with custom data, and each tool/service runs their own analysis algorithm using these custom data fields. Finally they publish and share their model.

Name - Mirit I. Aladjem
Organization - NCI
Project - molecular interaction maps

Molecular interaction maps in BioPAX

Need for diagrammatic language: Researchers interested in using modern systems biology approaches to understand signal transduction in normal and diseased states need tools to organize a large collection of facts, including descriptions of bioregulatory molecules, their modifications (for example, phosphorylation), and the complexes they form. To address this need, we require a standard language that will to integrate data in a clear, standardized, and preferably computer-readable format. Because enzymes in bioregulatory networks are often substrates of other enzymes, and molecules are often subject to modifications (by other molecules) that change their binding or enzymatic capabilities, the interactions that form a network are hard to describe with a linear series of binary interactions. In addition, regulatory proteins can form multi-molecular complexes, which have different activities, depending on their composition and modifications. These interactions can therefore be best represented in the form of diagrams.

MIMs as an unambiguous standard: Metabolic diagrams typically describe biochemical pathways in cartoon-like diagrams, but diagramatic representations of molecular interactions are often incomplete and ambiguous. For example, an arrow between two components could signify an increase in quantity, an increase in activity, or a modification of one molecule by the other. For efficient data exchange that will facilitate a full understanding of a biological system, the graphic representation needs to be complete and unambiguous. The molecular interaction map (MIM) language is designed as a graphical language that encodes molecular information in the form of diagrams in the same way as circuit diagrams are used to trouble-shoot electronic devices.

Typical case study: A researcher interested in understanding a biological process under normal and disease situation will first encode experimental data regarding molecular interactions (e.g. binding, modifications, degradation) using the graphic symbols suggested by the MIM language or a derivative of the language used in BioPAX. We already have a set of publicly available maps (http://discover.nci.hin.gov) that can be incorporated, whole or in part, as modules in other maps. This knowledge can be shared with other researchers familiar with the conventions of the MIM language, can be linked to outside databases and other sources (see examples on our web site), and can be easily updated as more information becomes available. The researcher could then derive specific situations (e.g. stages of the cell cycle, mutations in specific bioregulatory molecules, diseases) and predict possible outcomes from the MIMs. Predictions, which can be simply deducted from simple situation or derived from MIM-guided computer simulations, will be validated or refuted by experimental data. This process is likely to generate hypotheses regarding the roles of particular components of bioreglatory networks in specific signal transduction pathways, which will be further refined by a reiterative process of updating molecular interaction data and experimental validation.

Sander group use case:

The Sander group at MSKCC (Chris Sander, Mike Cary, Gary Bader) is interested in collecting and integrating human cancer relevant pathway information (metabolic, signaling, gene regulations and other more general associations) from all possible sources (including related processes in model organisms) for the purposes of:

1. Browsing and visualizing (resource)
2. Gene expression and other molecular profiling and gene copy-number (e.g. ArrayCGH) data analysis in the context of pathways
3. Pathway mathematical modeling/simulation