

**BioPAX Group**  
**June 3-4, 2004 Meeting Minutes**

**Location:** University of Chicago, Chicago, Illinois

**Participants:** Gary Bader, Michael Cary, Joanne Luciano, Natalia Maltsev, Elizabeth Marland, Suzanne Paley (via teleconference), Mustafa Syed, Jeremy Zucker

**Overview:**

1. Release Progress
  - a. Unexpected delays – brief summary by Gary and Mike of some of the issues they have encountered during release preparation
    - i. New issue: Pathway steps
    - ii. Cardinality: was inconsistent, we fixed it
    - iii. Comment changes
      1. Altering comments does not alter GKB output, but changes OWL file
      2. Class comments - use rdf:comment
      3. Instance comments - use COMMENT slot
      4. Added DB source class to entity
  - b. Need-to-do items
    - i. Finish spec document
    - ii. Glycolysis example
2. Converting WIT into BioPAX format
  - a. Started mapping WIT data model to BioPAX classes
  - b. Added CHEMICAL-FORMULA slot to smallMolecule class
  - c. Discussed generation of BioPAX output file
3. Promoting Level 1
  - a. Active promotion
    - i. PRL demo (<http://www.cbio.mskcc.org/prl>)
    - ii. Discussed criteria
  - b. Passive promotion
    - i. Discussed what resources we should provide to encourage users to adopt BioPAX on their own
4. Tools
  - a. Existing tools (45 minutes)
    - i. Using BioPAX in Protégé
    - ii. RACER and RICE
    - iii. XSLT
  - b. What tools should we develop/request/suggest?
    - i. Visualization tool
    - ii. BioPAX integration tool
    - iii. Develop biopax.lib
5. BioPAX & SBML
  - a. Reviewed draft of SBML Level 3

## **DAY 2**

1. Level 2 Planning, Roadmap
  - a. Target DBs – which DBs do we want to cover in Level 2?
    - i. BIND, MINT, DIP, HPRD, etc.
  - b. Level 2 Features – what features are needed to cover above DBs?
    - i. Molecular interactions
    - ii. Hierarchical pathways
    - iii. Complexes processes
    - iv. Generic entities
      1. Physical entity classes
      2. Markush structures
  - c. Level 2 Milestones
    - i. Decide on features
    - ii. Resolve issues
    - iii. Draft proposal
    - iv. Initial testing
    - v. Beta release
    - vi. Final release
2. Backward Compatibility (25 minutes)
  - a. How should backward compatibility effect our design decisions?
    - i. Adding slots and classes - generally OK?
    - ii. Altering existing classes and slots – typically not OK?
  - b. Converting data from BioPAX Level 1 to BioPAX Level 2
    - i. Provide a utility?
    - ii. How difficult depends on how radical the changes
3. Level 2 Issue: Molecular interactions
  - a. Integration of PSI
    - i. What changes need to be made in ontology?
4. Level 2 Issue: Hierarchical pathways
  - a. Allow pathways (along with interactions) to be parts of pathways
5. Level 2 Issue: Complex processes
  - a. Add INPUT and OUTPUT slots to pathway
6. Revise Pathway Steps
  - a. Merge INTERACTOINS and PATHWAY-STEPS into a single slot
7. Future Plans
  - a. Next meeting: ISMB

### **Summary:**

#### **Release Progress**

We began the meeting with a review of the progress being made on the release of Level 1. A number of minor bugs were discovered and fixed during the past month, including the following:

**Pathway steps** – Our old mechanism for specifying interaction order in a pathway was not efficient for storing complex steps, in which a reaction might be catalyzed by multiple enzymes or in which an enzyme's activity might be modulated by multiple compounds. The new mechanism, which replaces the interactionOrder class with the pathwayStep class, allows multiple catalyses and modulations to occur at each step of the pathway.

**Cardinality issues** – Translating the ontology into OWL and viewing in Protégé revealed several inconsistencies in how we specified the cardinality of various slots, these have been fixed.

**Comment changes** – A number of changes to comments (i.e. the text descriptions of classes and slots) in the ontology were made to clarify various points. The class and slot comments were recoded as rdf:comments, free text comments on instances will be stored in the COMMENT slot.

We then discussed the steps we need to take to finish the release. The two major items that needed completing were the spec document and the examples. Joanne and Jeremy had previously provided feedback on the initial spec document, some of which had been integrated. For example, based on Jeremy's and Joanne's comments Gary and Mike decided to add two sections: Use Cases and Best Practices to the spec document. Gary sent the unfinished versions of these to Jeremy and Joanne for review.

The ontology changes made recently must be propagated to the examples. Gary has been updating the glycolysis pathway from EcoCyc manually; Mike has been working on a fully fleshed-out pathway snippet (containing just a few interactions but with most slots filled).

### **Mapping WIT→BioPAX**

We then reviewed some of the progress that Joanne, Elizabeth, and Natalia had made in mapping WIT to BioPAX prior to the meeting. They only had time to cover small molecules, but there were no major incompatibilities detected. One item covered in WIT but not in BioPAX was chemical formula. After some discussion, we decided to add a slot called CHEMICAL-FORMULA to the smallMolecule class. Since chemical formula could also be represented via CML in the CHEMICAL-STRUCTURE slot, we will add a comment saying that in the case of conflicts, the CML data would trump the CHEMICAL-FORMULA slot.

Later in the meeting we spent additional time discussing the WIT→BioPAX mapping, and discussing some of the software options for performing the actual translation. Since WIT can be automatically dumped to XML, one option might be to use XSLT to translate the WIT XML into BioPAX. Another solution might be to create a BioPAX database in Postgres, then use existing software to load the WIT XML into this database. It should be fairly trivial to convert a dump of this database into BioPAX format. A third solution would be to modify an existing RDF parser, e.g. a perl module available from CPAN (see <http://search.cpan.org/search?query=rdf&mode=all>) to perform the translation. We also

discussed the creation of a library of reusable code for common operations (biopax.lib) involving BioPAX. We decided that we needed to do more research on existing RDF parsers and XML converters before deciding which of the above (or other) options were most practical.

### **Promoting Level 1**

We talked briefly about various strategies to promote BioPAX within the pathway community. Mike suggested that we identify a set of databases for which we wish to pursue active promotion. We would personally contact these databases and work with them to get their data into BioPAX format. Mike also suggested we provide sufficient resources and information on the BioPAX web site to allow users to adopt BioPAX without any aid from us (Mike called this passive promotion).

To help us identify which databases we should target for active promotion, Gary demoed the Pathway Resource List (PRL, <http://www.cbio.mskcc.org/prl>), which contains information on just about every pathway database available the internet. PRL ranks each resource by the number of web pages that point to it (as determined by Google). We decided to use this ranking, along with several other important factors such as amount of data, ease of translation into BioPAX, public availability of data, and overlap of data with other BioPAX data sets, as criteria for determining which databases to target for active promotion.

### **BioPAX Tools**

We briefly discussed some of the tools currently available that might be useful for BioPAX. Mike demonstrated how to import the BioPAX ontology from the web into Protégé (so that the ontology cannot be altered by the user), and briefly discussed RACER and RICE. Gary talked a little about XSLT, and how he used it to separate out the ontology from the EcoCyc glycolysis OWL file.

We also discussed what other tools we thought might be useful for the community. Aside from translation software (discussed above), support was voiced for a BioPAX visualization tool, and a BioPAX data integrator. The integrator would take as input two separate BioPAX files, and return a single fully integrated BioPAX file as output. This would solve the problem of data integration for any user that had BioPAX read/write capacity – the user would simply have to output their data into BioPAX format, use the integrator to merge it with any new data, then read the resulting file back into their database.

Joanne suggested we create a public to-do list – a list of BioPAX software we would like to see developed by whoever wished to do it. This idea was supported by the group.

### **BioPAX & SBML**

Jeremy and Joanne talked showed their work on using BioPAX to annotate SBML documents. This would be accomplished using RDF metadata on SBML species, in a manner similar to how CellML can be used to annotate SBML.

Jeremy also discussed the major features of the current draft of the SBML Level 3 proposal. SBML Level 3 will be able to capture greater detail on binding and activity sites of species. It will also be able to store assemblies of species, e.g. protein complexes, in such a way that the details of internal binding need not be specified. Also, these can be normalized so that it can be determined if two species that are represented differently are actually the same entity (providing the species are represented in sufficient detail).

## **DAY 2**

### **Level 2 Planning**

We began Day 2 of the meeting by planning the development of Level 2. We revisited the Roadmap and agreed that the target databases should include those we identified previously (i.e. those covered by the PSI-MI format). We also discussed a number of other features that we would like to consider supporting in level 2, including:

**Hierarchical pathways** – pathways that contain or interact with other pathways.

**Complex processes** – common multi-step processes such as transcription and translation; we need a mechanism of representing these in a clear and simple way that is compatible with the rest of the ontology.

**Generic entities** - protein families, generic pathways (i.e. pathways that are not organism-specific), Markush structures, etc.

### **Backward Compatibility**

Mike raised the issue of backward compatibility, stating that it should be a major consideration but that there may be cases where the benefits of a new feature outweigh the benefits of maintaining backward compatibility. In those cases, we should consider providing translators that convert BioPAX Level 1 into Level 2. In general, Mike stated that adding classes or slots to the ontology would generally not disrupt backward compatibility, but removing or changing them would.

### **Molecular Interactions**

Gary discussed recent progress made in PSI-MI, specifically, their progress toward their own Level 2 and their expanded scope. PSI-MI Level 2 will add support for other types of molecular interactions, such as protein-DNA and protein-RNA interactions. They have defined an interaction hierarchy that is very similar to our own and it should be easy to modify BioPAX to cover PSI-MI.

### **Hierarchical Pathways**

The proposed solution toward implementing hierarchical pathways was relatively straight-forward: allow pathways to occupy the INTERACTIONS (in the pathway class) and the STEP-INTERACTIONS (in the pathwayStep class) slots. This proposal did not

generate any objections on its own, but combining it with the following proposal for complex processes led to a lengthy discussion.

### **Complex Processes**

The proposed solution for representing complex processes was to add two slots to the pathway class: INPUT and OUTPUT. These two slots would give the pathway class the ability to represent a complex multi-step process without having to specify all of the interactions involved in that process (though users could of course still do so if they desired). For example, a pathway called “translation of PTEN mRNA” would have the PTEN mRNA in the INPUT slot and the PTEN protein in the OUTPUT slot.

[Addendum: another slot, not discussed, could point to the GO biological process entry for translation to clarify the nature of the pathway].

This feature, combined with the hierarchical pathways proposal, would also allow black-box pathways to be used more meaningfully within other pathways. For example, if a user wanted to represent the glycolysis pathway within a larger metabolic framework, but did not want to specify all of the gory details, they could simply create a pathway with INPUT=glucose and OUTPUT=pyruvate. This would describe the overall function of the pathway without the need for specific interactions.

Suzanne asked that, in that scenario, what the difference would be between a conversion and a black-box pathway - both would specify an INPUT and OUTPUT without any intervening interactions. Mike said that the difference was that a conversion describes a single step process, but a black-box pathway describes a multi-step process. Suzanne said that even so-called single step processes, such as biochemical reactions, could be broken down into multiple steps (e.g. enzyme forms complex with substrate, complex undergoes conversion, enzyme separates from product). We discussed these points at length and agreed that we would need to do more work on both the hierarchical pathway and the complex process issues.

### **Pathway Steps**

Suzanne then asked why we had both an INTERACTIONS slot and a PATHWAY-STEPS slot in the pathway class. Since they specify much of the same information, why not just have a single slot that could take either interactions or pathwaySteps? We debated the merits of this proposal and by the end of the meeting the majority of the group favored Suzanne’s proposal.

We also discussed at length alternative forms of the pathway and pathwayStep classes. Jeremy sought to define both in such a way that multiple representations of the same pathway (e.g. representations that use different internal hierarchical pathways) could be normalized and shown to be equivalent.

### **Future Plans**

We did not set a date for the next conference call but we agreed that the next face-to-face meeting would not occur before ISMB. We briefly discussed our ISMB plans – we will present talks in the BioPathways and BioOntologies sessions, and a poster in the main

poster session. We will also organize a BOF meeting with the goal of getting additional people involved in BioPAX.