

**BioPAX Group**  
**February 13, 2004 Meeting Minutes**

**Location:** SRI, Menlo Park, California

**Participants:** Gary Bader, Michael Cary, Peter Karp, Joanne Luciano, Suzanne Paley, Imran Shah

**Overview:**

- I. Grants
  - a. Renew DOE workshop grant
  - b. Begin planning additional grant applications
- II. Roadmap
  - a. Examples
    - i. Used to generate issues
    - ii. Production quality examples will accompany Level 1 release
  - b. GKB → OWL
    - i. 100% conversion from GKB to OWL
    - ii. Will be available in next version of pathway tools
  - c. Specification document
    - i. First draft underway
    - ii. Important to synchronize spec doc with class comments
    - iii. Need to add roadmap to doc
    - iv. Likely: multiple documents in future
- III. Issues
  - a. Constraints
  - b. Participants
  - c. Interaction / control type
  - d. Entity attributes
  - e. Pathway subclasses
  - f. Inverse slots
  - g. Predecessors
  - h. XREFs
  - i. Physical entity slots
    - i. Small Molecules: STRUCTURE
    - ii. Organism: in protein, RNA
  - j. Complex assembly, complexes, molecular associations
- IV. Next steps
  - a. Next meeting: Mid-April at Argonne National Labs (tentative)

**Summary:**

**Grants**

The meeting began with a review by Joanne of the DOE grant, followed by some discussion on how we will pursue additional funding. For now, we will seek renewal of

the existing DOE workshop grant. Joanne will investigate other grant opportunities; we will discuss them at the next meeting. Imran said it might be a good idea to finish Level 1 before requesting additional funds, as it would increase our chances of getting funding. Any additional grant application will likely require letters of support; Peter offered to send an example of one to Joanne, which she will adapt to our project and, when the time comes, ask members of the community to sign.

## **Roadmap**

We next discussed the progress that had been made on the milestones listed on the roadmap. Joanne and Mike created a number of examples using the 0.9 version. These generated a number of new issues. To save time, most of these examples were minimal; production quality examples will be generated when Level 1 is finalized.

Peter's group completed the GKB→OWL converter; it will be part of the next Pathway Tools release. The OWL output loads successfully into Protégé 2.0.

Gary, Imran, and Mike started the specification document, though it was still in an early stage at the time of this meeting. Imran suggested that we may want to have additional documentation beyond the spec, e.g. a more reader-friendly (more prose) document that outlines our objectives. It was agreed that the comment on each class in the ontology should be identical to what is listed in spec document; Gary suggested using the Word document as the main version, then copying and pasting the comments into the Ocelot file if necessary.

The remainder of the meeting, save for scientific talks by Gary and Peter, was dedicated to resolving the remaining open issues in Level 1.

## **Issues**

### **Constraints**

The first issue we discussed was the degree to which we should constrain various slots in the ontology. For example, should CONTROLLER be restricted to physical entities? Rather than discussing each possible constraint in the ontology, we agreed to compile a spreadsheet showing each slot, its constraint, and the rationale behind the constraint. The spreadsheet will be sent to the pax list for review and comment.

### **Participants**

We had previously discussed adding a slot to the interaction class called PARTICIPANTS. This was agreed to, but there was an objection to the name because the *complex* class also contained a slot called PARTICIPANTS. We decided to change the slot name in *complex* to COMPONENTS.

### **Interaction / control type**

We had also previously discussed having an INTERACTION-TYPE slot in the *interaction* class. Since this slot would primarily be used to specify whether a control interaction was activating or inhibiting, we decided to call the slot CONTROL-TYPE and place it in the *control* class, instead of in the *interaction* class. Suzanne agreed to send a

list (from EcoCyc) of possible controlled vocabulary terms we could use in this slot [Addendum: This list included the following terms: INHIBITORS-ALLOSTERIC, INHIBITORS-COMPETITIVE, INHIBITORS-IRREVERSIBLE, INHIBITORS-NONCOMPETITIVE, INHIBITORS-OTHER, INHIBITORS-UNCOMPETITIVE, INHIBITORS-UNKMECH, ACTIVATORS-UNKMECH, ACTIVATORS-NONALLOSTERIC, ACTIVATORS-ALLOSTERIC]. We left as an open issue whether we still wanted to have an INTERACTION-TYPE slot in the *interaction* class, which could be used to describe specific types of interactions (e.g. “phosphorylation”, “passive transport”).

#### Entity attributes

We briefly discussed the entity attribute model proposal, which would be one mechanism of representing location on physical entities but could also be expanded to represent states in later levels of BioPAX. We decided to leave this as an open issue, and to solicit feedback from the community to help us resolve it.

#### Pathway subclasses

We decided that subclasses of the *pathway* class were not essential for Level 1, since we will only be dealing with metabolic pathways. We left open the possibility of creating subclasses (e.g. signaling pathway, genetic network, protein-protein interaction network, etc.) in later levels.

#### Inverse slots

Many slots in the ontology point to instances of other classes. The issue at hand was whether we wanted those instances to contain slots of their own (inverse slots) that would point back to objects that refer to them (e.g. the INTERACTIONS slot in *pathway* points to the interactions in the pathway, an inverse slot called PATHWAYS could be created in *interaction* so that each *interaction* instance could point to the pathways that contain it).

The main argument for including inverse slots was that it made the ontology more amenable to direct adoption as the schema for a knowledge base. The inverse slots would increase the search speed of the KB, allow users to quickly see the relationships and pathways that entities were involved in, and would allow for easier browsing of the KB.

One argument against including inverse slots was that they are not necessary (they are 100% computable) and would therefore unnecessarily increase overhead (and possibly error). Another argument was that they would decrease the portability of physical entities, e.g. users could not extract the proteins from a biopax dataset without also extracting references to all of the interactions they were involved in.

We decided to leave this as an open issue, possibly to be resolved with the aid of community feedback.

#### Predecessors

While the interaction order for many metabolic pathways can be deduced by looking at the reactions, this is not always the case. We therefore decided to include a mechanism to indicate the order in which interactions in a pathway occur.

We decided to create a utility class called *interactionOrder*, which will contain two slots: PREDECESOR and SUCCESOR (single values for each). A slot will be added to the *pathway* class called ORDER-OF-INTERACTIONS, which will take values of type *interactionOrder*.

### XREFs

The issue at hand was how we should distinguish between various types of x-refs. We decided to create subclasses of the *xref* class for each type of x-ref that we wished to represent (*publicationXref*, *relationshipXref*, and *unificationXref*).

The *unificationXref* class will contain x-refs that point to the exact same biological entity. We decided not to add a SOURCE flag to this class, which would indicate whether or not a particular *unificationXref* pointed to the original data source of the BioPAX instance, but we agreed that the issue of data provenance will be a high priority in Level 2.

The *relationshipXref* will contain a slot that describes the nature of the relationship between the originating instance and the object pointed to by the x-ref. The slot will take a string; possible values for the string (e.g. “sequence similarity”, “sequence template”, “structural similarity”) will be suggested in the comment field but will not be dictated by a controlled vocabulary.

The *publicationXref* will point to literary references for the instance, and will contain additional slots such as AUTHORS, TITLE, SOURCE, URL, and YEAR.

### Physical entity slots

We next discussed how much detail we wished to capture in the physical entity classes. We had previously agreed that we needed to represent small molecule structures, but we hadn't agreed upon the mechanism.

We decided to capture small molecule structure via a slot in *smallMolecule* called STRUCTURE, which will take instances of the utility class *chemicalStructure* as values. These instances will consist of a format identifier (either “SMILES” or “CML”), and a string that will contain the data (i.e. the SMILES string or the entire CML file).

We also decided to add a MOLECULAR-WEIGHT slot to *smallMolecule* (but not to *protein*), and to add an ORGANISM slot to *protein* and *RNA*.

### Complex assembly, complexes, molecular associations

The last issue we discussed was whether *complexAssembly* referred only to non-covalent binding. We decided that this would be the case, and that *complexes* would be defined as non-covalent associations between a set of participants, at least one of which was a

macro-molecule. We also decided to specify in the *complexAssembly* class that one of the products had to be a *complex*.

### **Next Steps**

We agreed that we should get community feedback on what we intend to release as Level 1, and that we should create a few more examples after we finalize it. We will try to resolve the remaining open issues in the near future via email. The next meeting will likely be held in mid-April at Argonne National Labs (this is tentative – it was suggested on an earlier occasion by Natalia).