

BIOPAX Work Group
9/03/03 Conference Call Minutes

Participants: Gary Bader, Michael Cary, Peter Karp, Joanne Luciano, Eric Neumann, Aviv Regev, Imran Shah.

Overview:

1. **Ontology work / status**
 - a. **Version 0.5 to be released 9/19/03**
 - i. **Will include: GKB, Protégé (OWL), XML Schema versions of ontology; documentation; announcements on web**
 - b. **Still to do:**
 - i. **Add examples to ontology**
 - ii. **Create OWL version from Protégé**
 - iii. **Update XML Schema version**
2. **External groups**
 - a. **GK – we will contact them**
 - b. **W3C pathways RDF project – we will monitor their work**
3. **Small molecule subgroup**
 - a. **We will invite Erik Brauner to next meeting**
4. **Upcoming meetings, conference calls**
 - a. **9/17/03 – next call**
 - b. **10/3/03 – next f2f meeting**
5. **Ontology issue: Activation / Inhibition classes ?**
 - a. **No, use “InteractionType” slot**
 - b. **We will gather examples to develop and test this**

Pre-call comments

Vincent Schachter said he'd be happy to review and monitoring progress more closely. We might want to consider starting the conference calls a little past the hour, to reduce the time it takes to get connected (due to reduced traffic).

Ontology work / status

Our goal is to release version 0.5 by September 19, which is two weeks before our next face-to-face meeting. That release will consist of: GKB, Protégé, and XML-Schema versions of the ontology, as well as documentation and web page support.

The following items still need to be done:

1. Examples need to be added to the GKB and Protégé files.
2. Create an OWL version from the Protégé version (via plug-in).
3. Create (updated) XML Schema version.

Report on Gordon Conference by Eric Neumann

The GK group is interested in our work. Eric will set up a conference call with them within two weeks. In the mean time, we should all look at their data model.

Eric attended the Gordon Conference last week at Oxford, England. He said it was an interesting mix because things have definitely shifted towards looking at larger systems and pathways. Shankar Subramaniam was present talking about his latest work, which was with AfCS. Eric had a chance to talk with him again and he definitely wants to be hooked up into the BioPAX conversations. Eric said he briefly presented some information about BioPAX, and said that if anyone there was interested they should follow up with Gary or any of the other people listed in the BioPAX site. Many groups expressed a strong interest in getting involved. Specifically, as you saw a note Eric sent earlier, Ewan Birney wanted to figure out how BioPAX can work with the GK initiative that he and Lincoln Stein and others are doing. Eric just received an email from them asking if something could be set up within a two-week time frame, and asking how best to do it with some of the other BioPAX members.

Gary asked the group what we thought about "how to deal with external groups like GK?" One way we could do it is like the small molecules subgroup has been doing. They take their own initiative at setting up conference calls and contacting other groups; this puts less stress on the core group. We could have a subgroup that contacts other people.

Eric said that would make sense and asked if anyone had taken a look at their web site. He hadn't spent as much time as he would have liked and wondered if others had taken a look at what their mission is. They are working around a lot of content. He said, as Ewan put it, they're realizing how they have to start to think through about adding a lot more type semantics on things. That's where he (Ewan) thought that what BioPAX is doing would be perfect ingredient for what they're up to and suggest that we discuss with them soon on how that would be structured.

Eric said they're moving quite quickly because they have an NIH grant and they have the key genomics players on it. So they are in the midst of defining semantics around gene regulation and metabolism. So it would make sense and they are open to work with where we are right now. But, they certainly have, at this point, a high bit of momentum, so sooner rather than later.

Generally, they want to organize an ontology database or knowledge base with the GO (Gene Ontology) work that's been done in an integrated model across species. So they will be driving it with data and initiatives. The Genomic database is going to have some form of GK as a source of semantics and connectivity. Which, Eric pointed out, is close to what BioPAX's original mission is too. Their drive is from the genomics centers. They've got the GO folks tied up on that. Eric said to look at the membership list to get a sense of who is playing on this; they do have major NIH funding. He would rather not see duplication of efforts on things. Both groups are quite young still, so this is the right time to do it.

There will be a discussion between a few key players. From Eric's group, there's going to be Brian Gillman and probably a few others that are interested in the GO annotation work. So he would think that if Gary were available that he should be on that call.

Gary said he's interested and let's set up a call and post it to the list and ask whoever would like to be on the call to be on it and report.

Is there any specific to be contacted? Eric said yes, and named Lincoln Stein, Ewan Birney, and Imre Vastrik. Imre seems to be their key developer. And he's the guy that's trying to pull together an ontology pretty much by himself.

Imran sent a link around to everyone and said it would be useful if everyone look at the data base object that they define. Eric said, "Absolutely, you can download it in either MySQL or Protege format." So, an action item for the group is that within the next week or so, we all take time to look it over and think about it, because it is quite active and is gaining a lot of recognition in the genomics community.

We should also take a look at their website (the GK team page) to see who's involved. Gary agreed to contact Ewan, because Eric said that we should talk to him first.

Eric's RDF work

Eric is involved in a group working on an RDF-based form of pathway data representation. He will send us info about joining their mailing list.

Eric said he also wanted to mention that he's also been pushing the use and discussion of an RDF-based form of pathway data representation. He said he was part of a team of about 10 people who planned to use the OWL version of BioPAX (among other things) in their work with RDF. He asked if we would like to interface with this project via a subgroup or whether we would prefer he simply report back to us with their findings. A new W3C mailing list is about to be set up (within the next few days) that will serve the project members.

Joanne asked if the email list would be high volume – she thought maybe it should be forwarded to the pax list so that we would be tuned into it. She also wondered if there would be any duplication of our work by this new RDF group. Eric said that they were not looking into using RDF as an exchange language, rather, they were looking at it from the perspective of the semantic web – i.e., using it primarily to mark-up documents (actually, any kind of scientific literature) with additional annotation.

Eric will forward to the mailing list information on how to join the email list.

Small Molecule Subgroup

Erik Brauner is not using CML, but he will send us the DTD for the format they created. We will invite Erik to talk at the October 3 meeting.

Imran and Peter had a conference call with Eric Brauner, in which they went over the items that were brought up during an email correspondence a few weeks ago. They talked about quality control, mode of submission, updates, keeping the history of various annotations, licensing terms, and the format of the data they currently have. Imran said it seems like it would be productive to invite Eric Brauner to the next face-to-face meeting, so he could give us a presentation on ChemBank. Imran said he will also write up a short summary of what was discussed and send it to the pax list. Eric agreed to send us the DTD they are currently using for their small molecules – they are not using CML.

Both Peter and Imran said the call was very positive, they said Eric has a good vision. Imran said that we should be able to point to ChemBank entries with BioPAX. Imran proposed that we invite Eric Brauner to our next meeting (Oct. 3 in NYC). Gary volunteered to invite him, and ask him to present a demo of ChemBank.

Examples subgroup

Joanne has been waiting until now to contact people, due to summer vacations. Gary said that it would be great if they could provide us with examples that test the limits of BioPAX.

Upcoming meetings, conference calls

Next call: September 17, 2003 at 2 PM EST

Next face-to-face meeting: October 3, 2003 in NYC

The earliest Aviv could attend the October 3 meeting would be late afternoon Friday, so she favored a two-day meeting. Peter did not favor a two-day meeting, especially not one that met on Saturday. Both Imran and Peter are attending a DOE meeting on microbial genomes that ends on October 2nd, so they would not be able to attend on that day. Gary proposed scheduling a full day meeting on Friday, October 3, with an optional additional work session on Saturday morning, in which we would try to get some work done but refrain from making any major decisions.

Ontology issue of the day

How should we deal with activation / inactivation information?

Summary: For now, we will use the “InteractionType” slot (present in all Interaction classes) to describe whether the interaction is activating or inhibitory, as well as to specify the context of the activation / inhibition. We will research others’ work in this area to develop a controlled vocabulary for the “InteractionType” slot. We will gather examples to test this approach.

Background: these concepts are very often used to describe the relationships between components of many types of pathways, and we should be able to describe them without too much difficulty. The basic question is “should we have activation and inhibition classes?”

The current ontology has a slot called InteractionType on the Interaction class. This slot takes its values from a controlled vocabulary, two terms of which could be “activation” and “inhibition”.

Eric pointed out that terms like “activation” are projected terms or functions. They are not biochemical properties of an interaction - they are descriptions of interactions that are projected back from downstream biochemical events. Therefore, terms like “activation” should be considered annotation, not as specific types of interactions (i.e. not as separate classes).

Imran agreed with Eric that activation and inhibition shouldn't be classes. He also said we need to distinguish between the biochemical level and the logical level when it comes to describing interactions. It's the difference between observed facts and what the result or effect of those facts are.

Aviv asked what we would do if the biochemistry was not known, but a general causality (A activates B) was. Imran said that perhaps such interactions should be handled differently from biochemical interactions – he did not think the two cases should be handled in the same way.

Aviv suggested that the interaction class could be enriched to contain both biochemical and logical information. Both would be optional, so that if we only knew, for example, the logical details only those would be required.

Gary said that this could be achieved by simply using the more abstract Interaction class. PSI has a very basic controlled vocabulary for InteractionType, but we would need to build on it. Gary said we could use Andrey Rzhetsky's work to add to the InteractionType c.v.; Peter said that that was one of the reasons he wanted Andrey to attend the October 3rd meeting. Aviv said that he did indeed plan to attend the October 3rd meeting.

Eric said he liked what Gary described, and said it sounded like it would be able to solve a lot of the problems he has encountered. He said an interesting notion to consider is that causal relationships require a set of rules. For example, some types of activations are transitive (if A activates B, and B activates C, then A activates C). You want the system to understand these rules so that you do not have to manually enter relationships over larger steps or distances – you don't want to build them, but want to be able to deduce them. He said you could write such rules in RDF.

Peter said he would like to have an example to work with to better illustrate the issue. For example, he wondered how one would specify which activity of a multifunctional enzyme was “activated” by a certain interaction.

Aviv pointed out that state needs to be well defined in order for terms like “activation” to have meaning; one needs to define a context when one says “activation”.

However, sometimes you don't know the biochemical details, in which case you cannot be specific about the mechanism of activation. It is slightly easier when a protein has only 1 known function, but still problematic.

Peter then made two points: 1) we need examples, 2) I thought we talked about this in terms of catalysis – that class should have slots for activators and inhibitors – should define activators and inhibitors of the pairing of an enzyme with a reaction.

Aviv said that in metabolic enzymatic reactions, you have activators and inhibitors. In signaling networks, you have an enzymatic reaction that activates the enzyme. The delicate thing about signaling is the enzyme is often the substrate.

Peter said that his group defines two forms (states) of the protein.

Aviv reiterated that if you know the details, your situation is easier. But what about the case when you know something, but not the details? What we are annotating is the reaction, not the enzyme. Peter said that he hasn't studied the imprecise cases.

Imran said it might be useful to use the same thing as enzyme catalyzed reaction. Aviv agreed, agreeing that a signaling reaction is an enzyme catalyzed reaction.

Peter said we should gather examples, and asked for a volunteer to coordinate this. Joanne and Mike volunteered.

Eric mentioned that he has come across a couple papers on action languages – higher order logic languages. These handle some notion of causality. He said we should build some form of a causal structure, we should think about supporting or incorporating a minimal causal model.

Gary asked if it would be a more specific definition. Eric replied that it would be finding a minimum set for “activation”, “inhibition” – we need to be more rigorous in the logical structure.

Miscellaneous

GKB technical question: In GKB, facets can be added to slots in either the “slot unit”, which is a frame in the knowledge base that represents that slot and its local properties, or in the slot of a specific class. Which is the best practice? E.g., where should one specify cardinality? *GKB technical answer:* Generally, put everything on the slot unit unless you want to enter class-specific information. Peter always uses the slot unit. The HTML documentation will only be generated for the slot-unit, not for class-specific modifications of slots.

Peter mentioned that his group has created a HumanCyc metabolic pathway database, which can be found on their website.