

BioPAX Ontology Class Structure

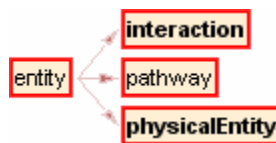
Tuesday, September 23, 2003

This document describes the current preliminary structure for the BioPAX ontology. A GKB (<http://www.ai.sri.com/~gkb/>) and a Protégé ontology file describing this has been made as well. Please comment.

Also note that full comments and documentation about how much more conceptual work is required to define specific classes is present in the GKB and Protégé ontology files.

Basic definitions

We define 4 basic concepts in the ontology: the top level entity class and three subclasses: **pathway**, **interaction** and **physicalEntity**.



Top level class

Entity (Root class of ontology)

Definition: Any concept that we will refer to as a discrete unit when describing biological pathways. e.g. a pathway, interaction or physicalEntity.

Second level classes

Pathway

Definition: A set of interactions (a pathway has interactions). A pathway is also an entity.

Example: apoptosis, glycolysis

The Pathway class may have sub-classes (e.g. all of the elements in GO Process), but this will be a topic of future discussion.

Interaction

Definition: A set or sets of entities and some relationship* between them. An interaction is an entity.

Naming rationale: We wrestled with a number of names for this concept, including “process”, “synthesis”, and “relationship” (‘relationship’ was our original choice, but we later redefined it*); we settled on “interaction” but are open to suggestions for a better name. One reason interaction was chosen is that PSI (<http://psidev.sf.net>) uses it and we would like to remain compatible with them.

physicalEntity

Definition: A building block of interactions.

Naming rationale: It's hard to find a name that encompasses all of the examples below without being too general. `physicalEntity` seems like a good choice, although `part` was suggested as well. PSI uses `interactor` here, so this may be an issue later.

Examples: protein, small molecule, RNA, DNA, photon, environment, etc.

Analogies to other conceptual areas

Linguistic

Entity ~ Noun (Subject or Object)

Relationship* ~ Verb

Interaction ~ Phrase/Sentence

Pathway ~ Paragraph

Graph representation

Entity ~ Node

Relationship* ~ Edge

Interaction ~ Either a node set by itself or a node set connected to another node set by an edge

Pathway ~ Graph

A->B->C

A, B, C ~ Entities

->, -> ~ Relationships*

A->B, B->C ~ Interactions

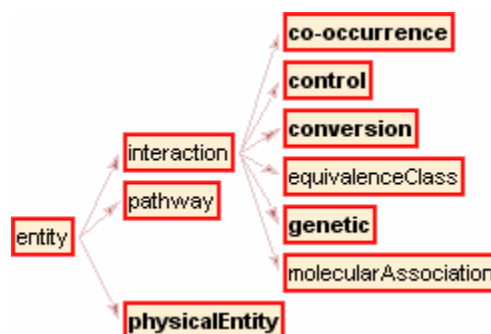
A->B->C ~ Pathway

*Relationship

Definition: Something that may only exist *between* two or more entities. Thus, a relationship is *not* an entity since it requires something external (the two or more entities) to exist. There is no class for "relationship", because relationships are defined in the ontology via slots (in classes) that refer to other classes.

Interaction sub-classes

Six terms exist under interaction: Control, conversion, molecular-association, co-occurrence, equivalenceClass and genetic



Control

Definition: The control of a process.

Examples: Enzyme catalysis controls a biochemical reaction, gene regulation controls gene expression.

Conversion

Definition: A conversion process, which converts one set of entities to another set.

Examples: A biochemical reaction converts substrates to products, the process of complex assembly converts single molecules to a complex, transport converts entities in one compartment to the same entities in another compartment.

Molecular Association

Definition: An association between a set of molecules.

Examples: Arp2-Arp3 protein-protein interaction; protein complex e.g. the result of a co-immunoprecipitation experiment; hexokinase-glucose

Co-occurrence

Definition: The co-occurrence of entities in some context. That context could be time, space, a sentence, sequence similarity space, etc.

Examples: Colocalization of a few receptors e.g. in a GPI anchored lipid raft; co-migration of cells; genes expressed at the same time.

Equivalence Class

Definition: A set of entities that can be considered equivalent in some context.

Examples: A set of paralogs that can replace each other as enzymes in a biochemical reaction, a set of enzymes that may not be homologs, but are functionally identical e.g. glucose-6-phosphatase.

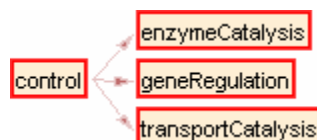
Genetic

Definition: A genetic interaction. An interaction between elements of a genotype that results in a change in phenotype.

Examples: Synthetic lethal interaction

Control sub-classes

Three types of process control exist under the control class: enzyme catalysis, gene regulation, transport catalysis.



Enzyme Catalysis

Definition: The process of enzyme catalysis. An enzyme interacts with a biochemical reaction. This class contains information about the enzyme catalyzing the reaction and enzyme kinetics. Enzyme catalysis controls biochemical reactions and together these classes can be used to describe typical metabolic pathways.

Examples: Hexokinase -> (The “Glucose + ATP -> Glucose-6-phosphate +ADP” reaction)

Gene Regulation

Definition: The process of gene regulation by a transcription factor. A transcription factor binds upstream of a gene to regulate it. This is the basis of genetic regulatory networks, where typically an arrow is drawn from the transcription factor to the regulated gene.

Examples: Transcription factor -> Regulated gene

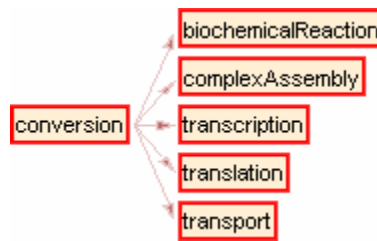
Transport Catalysis

Definition: Control of the process of transportation. An active or passive transporter can facilitate a translocation event.

Examples: Nuclear pore complex -> (The “Protein A (cytoplasm) -> Protein A (nucleus)” translocation)

Conversion sub-classes

Five types of conversion process exist under the control class: biochemical reaction, complex assembly, transcription, translation, transport.



Biochemical Reaction

Definition: A biochemical reaction is a reaction in which substrates are written in terms of sums of species. This is further defined in Peter Karp’s interactions ontology here:

<http://www.ai.sri.com/~pkarp/misc/interactions1.html>. Note: this class does not contain any reference to enzyme kinetics.

Examples: $\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_i$; all EC reactions.

Complex Assembly

Definition: A step in the process of complex assembly. This is not the same as a complex, which is a physical object, called molecular association in the interaction class hierarchy.

Examples: ProteinA + ProteinB -> ProteinComplexAB

Transcription

Definition: The process of transcription. This is useful to generally describe transcription.

Examples: Gene A -> mRNA A

Translation

Definition: The process of translation. This is useful to generally describe translation.

Examples: mRNA A -> Protein A

Transport

Definition: The process of transport from one spatial location to another, typically within or at least with respect to the cell. This does not include a transporter molecule, which would be described in the transport control class.

Examples: Protein A (cytoplasm) -> Protein A (nucleus)

Genetic sub-classes

Three types of genetic interactions exist under the genetic class: epistasis, suppression, synthetic. Genetic interactions are considered separately from interactions involving physical entities, like proteins, because genetic interactions are sufficiently different.



Epistasis

Definition: An epistatic interaction occurs when an allele at one locus renders the genotype at a second locus irrelevant - the phenotype will be dictated by the genotype of the epistatic gene alone. (This definition from

http://starklab.slu.edu/Coulter/genetics/handout3_2002.html)

Examples: Gene A -> Gene B (Gene A is epistatic to gene B)

Suppression

Definition: A suppressor is generally defined as a mutation that completely or partially restores the mutant phenotype of another mutation. (This definition from

http://dbb.urmc.rochester.edu/labs/Sherman_f/yeast/12.html#12.3)

Examples: Gene A -> Gene B (Gene A suppresses gene B)

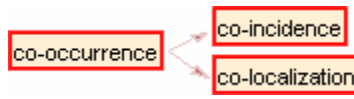
Synthetic

Definition: A synthetic genetic interaction. Two genes have a synthetic phenotype if each gene separately mutated has no phenotype, but when both genes are mutated at the same time, the phenotype exists.

Examples: Gene A and gene B are synthetic lethal if a double gene A and gene B mutant results in death, but neither mutation is lethal independently.

Co-occurrence sub-classes

Two types of interactions exist under the co-occurrence class: co-incidence and co-localization.



Co-incidence

Definition: The co-occurrence of entities in time.

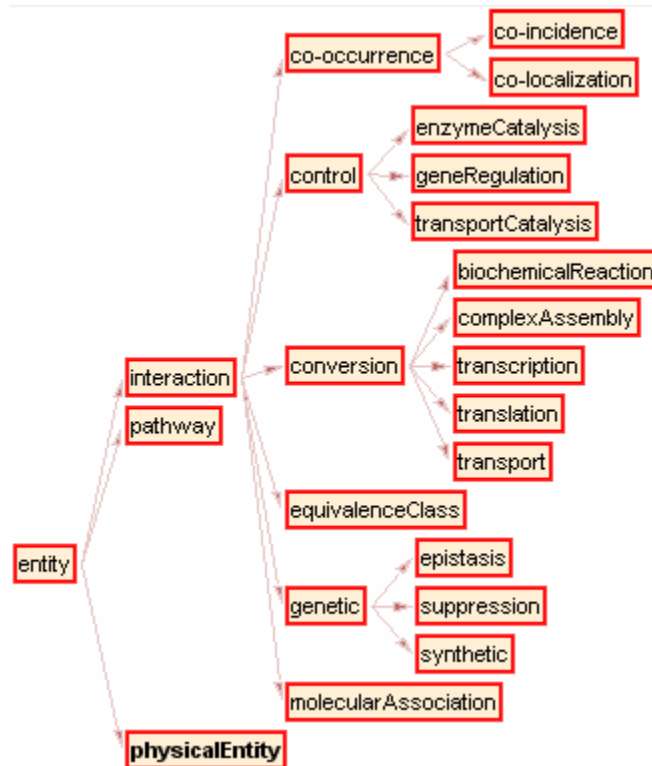
Examples: co-migration of cells; genes expressed at the same time.

Co-localization

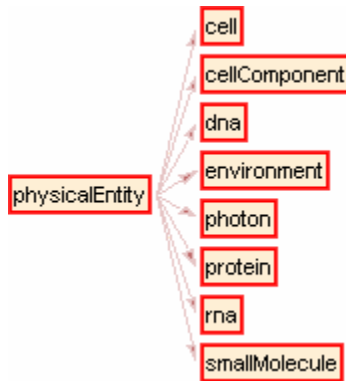
Definition: The co-occurrence of entities in space.

Examples: Colocalization of a few receptors e.g. in a GPI anchored lipid raft

Summary of Interaction Class Structure



physicalEntity leaf classes



DNA

Definition: Deoxyribonucleic acid

Examples: The EGFR DNA sequence. See GenBank for more examples.

RNA

Definition: Ribonucleic acid

Examples: messengerRNA, microRNA, ribosomalRNA

Protein

Definition: A protein

Examples: The EGFR protein sequence. See Swiss-Prot for more examples.

Small Molecule

Definition: A non-polymeric biomolecule. Generally, any bioactive molecule that is not a peptide, protein, DNA, RNA or possibly not a complex carbohydrate.

Examples: glucose, penicillin

Cell Component

Definition: A part of a cell. The Gene Ontology contains a large list in the ‘cellular component’ ontology.

Examples: nucleus, mitochondrion

Cell

Definition: A specific type of cell.

Examples: cardiac myocyte, B lymphocyte.

Photon

Definition: Light at some intensity and wavelength

Examples: UV light

Environment

Definition: A physical or environmental effect.

Naming rationale: This could also be called physical events as Demir et al. (Bioinformatics, submitted) call it.

Examples: Calcium wave, electric shock, heat, mechanical stress.

Summary of BioPAX Class Structure

