Role-based representation and inference of biochemical processes

Christian Bölling[1,*] Michel Dumontier[2], Michael Weidlich[3] and Hermann-Georg Holzhütter[1]
[1]Institute of Biochemistry, Charité Universitätsmedizin Berlin, Seestr.73, 13347 Berlin, Germany
[2]Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S5B6
[3]Department of Computer Science, Humboldt-Universität Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

ABSTRACT

We present a streamlined data model for representation of biochemical processes which consistently adopts a perspective on these processes as molecular events.

Our model references a small number of established foundational relations predominantly from RO and employs BFO as upper ontology. It addresses some of the limitations in terms of interoperability, semantic compatibility and expressivity encountered in other approaches to modeling biochemical processes.

Using a role-based approach we demonstrate how from this perspective various metabolic and transport processes can be consistently represented across different levels of granularity and how relations between processes like sequence of events can be inferred.

1 INTRODUCTION

Computational approaches to study biochemistry require machine accessible representations of biochemical knowledge. While various schemes have been specified for the representation of biochemical processes, their underlying conceptualizations differ with regard to biochemical scope, molecular detail, and provision of meta-data and adopt formal syntax and semantics to varying degrees. BioPAX (Demir *et al.,* 2010) provides a basic ontology to exchange data on biochemical pathways and their interactions, with an emphasis that these represent bulk phenomena, as opposed to single molecular events. Although the BioPAX ontology is specified using the Web Ontology Language (OWL, Hitzler *et al.,* 2009), the axioms are mostly there to constrain the types of relations allowed, as opposed to a more expressive description of pathways and the molecular participants found therein. Towards addressing these limitations, an OWL-based representation was put forward to describe types of biochemical pathways and reactions in terms of the molecular participants, their parts and the roles that they play (Dumontier, 2008). Here, we extend on that preliminary work with a basic ontology of biochemical processes consisting of one or more biochemical reactions, and specifying roles that molecular entities play therein. Accompanied by relevant rules specified using the Semantic Web Rule Language (SWRL, Horrocks *et a.l.,* 2004), our approach facilitates biologically-relevant inferences.

*To whom correspondence should be addressed: christian.boelling@charite.de

$`courier` typeface denotes OWL-classes, *`italics`* denote OWL-properties

2 RESULTS

The perspective taken on biochemical processes in this work is that individual molecular entities, i.e. single molecules, interact with each other in various processes through which molecular structures of various complexity are formed and dynamic biochemical and physiological phenomena on the macroscopic scale are produced. In our OWL2 representation, individuals of classes describing biochemical processes represent singular molecular events, i.e. directed transitions of a chemical system from an initial to a terminal state involving individual molecules.

2.1 OWL-constructs for role-based representation of biochemical processes

As a matter of convenience, our representation uses the class and property distinctions identified by the Basic Formal Ontology (BFO, Grenon *et al.,* 2004) and the OBO Relation Ontology (RO, Smith *et al.,* 2005). Molecules are types of `bfo:object`, roles are types of `bfo:role` and biochemical processes are types of `bfo:process`.

We developed a basic ontology of roles that chemical participants hold in the context biochemical processes (Fig. 1). The role ontology includes a role for catalysts (`catalyst_role`), reactants (`reactant_role`), substrates (`substrate_role`), products (`product_role`), effectors such as activators (`activator_role, enzymatic_activator_role`) and inhibitors (`inhibitor_role, enzymatic_inhibitor_role`). Consistent with IUPAC and IUBMB terminology (IUPAC, 2011) reactants are participants that are present at the onset and products are participants that are present at the end of the process. Substrates are reactants that are converted to products by the activity of one or more enzymes. Enzymes are catalysts of mostly protein nature. Effectors are chemical entities that affect the functionality of enzymes with respect to the rate of reaction.

We further developed a simple ontology of biochemical processes which distinguishes between elementary reactions and overall reactions (Fig. 1). Elementary reactions pertain to fine-grained mechanistic aspects of biochemical proc-

esses and include association, dissociation and conversion events. Overall reactions comprise of single- and multi-enzyme reactions and net reactions catalyzed by structurally independent enzymes which as such reflect traditional biochemical pathways. The roles of chemical entities may be described in the context of the biochemical reactions in which they are realized. Stoichiometry may also be specified as cardinality restriction on the *realizes* property between the process and the role. For example, hexokinase-like reactions, i.e. the conversion of glucose (glc) and ATP to glucose-6-phosphate (g6p) and ADP are defined as:

```
(realizes exactly 1 (product_role
  and (has_bearer some adp)))
 and (realizes exactly 1 (product_role
  and (has_bearer some g6p)))
 and (realizes exactly 1 (reactant_role
  and (has_bearer some atp)))
 and (realizes exactly 1 (reactant_role
  and (has_bearer some glc)))
```

Due to their status in BFO as specifically dependent continuants, these roles are borne only by single molecules, thus reaction stoichiometry is duly reflected in our representation.

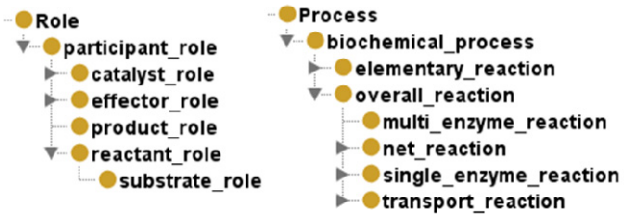


Fig. 1. Taxonomy of the ontology of molecular roles and biochemical processes.

2.2 Representation of biochemical processes at various levels of granularity

Both elementary reactions and overall reactions can be described in terms of its reactants and products, i.e. in terms of the molecular roles being realized in a biochemical process. For example, the association of ATP and the hexokinase enzyme is an elementary reaction of the hexokinase reaction, while the overall phosphorylation of glucose with ATP involves ATP, glucose, ADP and glucose-6-phosphate (Fig. 2). In the case of hexokinase, we observe that it plays the role of a reactant in the elementary reactions which are part of the hexokinase reaction and glycolysis, while it plays the enzyme role in those more "macro" reactions. Additional detail, such as the participation of catalysts or cofactors can be represented with the corresponding role classes, making clear the nature of their participation.

2.3 Relations between processes: process parts and sequence

The relation of more complex processes to their constituent process parts can be represented by *part_of* and

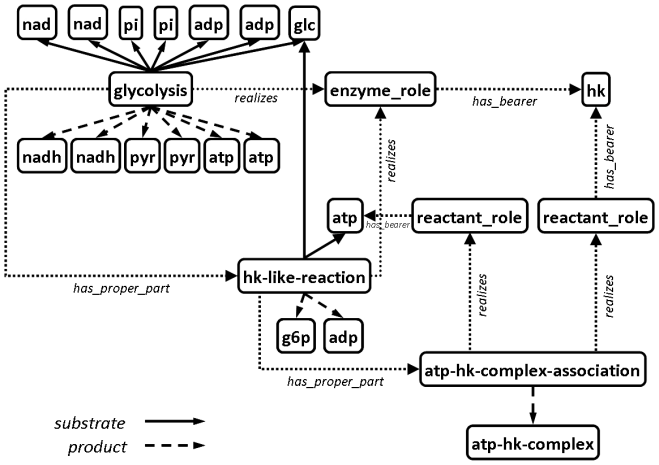


Fig. 2. Biochemical processes at different levels of granularity using reactant and product roles. Boxes denote named OWL-individuals. Dotted arrows denote object properties as labelled. Solid and dashed arrows denote substrate and product roles, resp. which connect processes in which they are realized with molecules by which they are borne.

preceded_by relations from RO as outlined in Dumontier 2008. In addition *directly_preceded_by*, as a sub-property of RO's *preceded_by* connects instances of processes which are coupled by joint participants which bear product roles in the preceding and reactant roles in the succeeding process (Fig. 3). In contrast to the *immediately_preceded_by* relation defined in RO, this property relates processes which are not necessarily temporally adjacent.

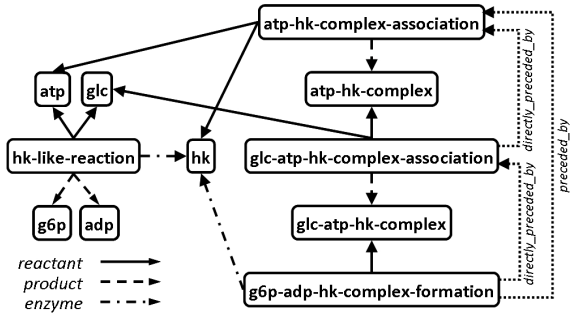


Fig. 3. Sequence of biochemical processes. Boxes denote named OWL-individuals. Dotted arrows denote object properties as labelled. Other arrows denote participant roles connecting processes with molecules as indicated.

2.4 Location of processes and representation of transport reactions

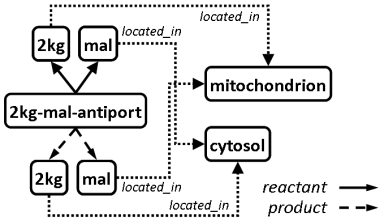
Location of molecules can be represented using the RO relations *located_in*. Location of processes, i.e. where they occur is specified using the *occurs_in* property.

Transport processes are represented also in terms of their reactants and products, formalizing the transported entities

as individual instances of the corresponding chemical species connected to instances of reactant role and product role via the *bears* property and to instances of the corresponding locations via *located_in* (Fig. 4). For example, the antiport of 2-ketoglutarate (2kg) and malate (mal) across the mitochondrial membrane is defined as:

```
(realizes exactly 1 (product_role
  and (has_bearer some (2kg
   and (located_in some mitochondrion)))))
 and (realizes exactly 1 (product_role
  and (has_bearer some (mal
   and (located_in some cytosol)))))
 and (realizes exactly 1 (reactant_role
  and (has_bearer some (2kg
   and (located_in some cytosol)))))
 and (realizes exactly 1 (reactant_role
  and (has_bearer some (mal
   and (located_in some mitochondrion)))))
```

Fig. 4. Representation of transport reactions. Boxes denote named OWL-individuals. Dotted arrows denote object properties as labelled. Other arrows denote participant roles connecting processes and molecules as indicated.



2.5 Inference of process and entity characteristics

Reasoning over the OWL representation of biochemical processes as described above enables the following:

Classification of processes: Processes can be classified according to specialization of roles and chemicals. For instance, a process involving a chemical as a reactant would subsume a process involving that chemical as a substrate. Given an ontology of chemicals (e.g. ChEBI), similar classification of processes are enabled.

Location of molecules: This can, for participants of localized processes, be inferred from the location of the process using SWRL: `biochemical_process(?p), occurs_in(?p,?l), has_participant(?p,?o) -> located_in(?o,?l)` (the `has_participant` property can be inferred for any bearer of any role realized in a reaction).

Sequence of processes: within the same location this can be deduced by invoking the SWRL-rule: `product_role(?r1), reactant_role(?r2), has_bearer(?r1,?o), has_bearer(?r2,?o), realizes(?p1,?r1), realizes(?p2,?r2), occurs_in(?p1,?l), occurs_in(?p2,?l) -> directly_precedes(?p1,?p2)` and the transitivity of the `preceded_by` relation.

3 DISCUSSION

In this representation biochemical processes can be consistently described on different levels of granularity accounting for different roles of participating molecules on different levels. By including location and transport even complex biochemical processes can be represented using a small set of basic relations. This provides a stable platform for interoperability with ontological descriptions of related biological entities (e.g. molecules, tissues, taxa) which could also be used to represent and interrelate GO biological processes via their participants. Our representation applies a consistent perspective on biochemical processes as microscopic chemical events. This provides, together with the formal semantics of OWL2, a clear semantic basis to represent complex processes and complex structure-function relationships and to interpret their asserted and inferred properties in terms of biochemical entities. For example, substrate channeling can be represented through molecule instances which bear product and substrate roles for the preceding and succeeding reaction in a straightforward manner. Thus, our representation is suited to overcome some of the limitations regarding interoperability, semantic compatibility and expressivity that have been identified in other models (Dumontier, 2008) which makes it a promising base for representation and analysis of the biochemistry of organs like the liver, i.e. of complex systems, comprising interrelated processes on several scales.

ACKNOWLEDGEMENTS

CB and MW were supported by the German Federal Ministry of Education and Research (BMBF) within the Virtual Liver Network (grant numbers 0315756, 0315741).

REFERENCES

Demir, E. et al. (2010). The BioPAX community standard for pathway data sharing. Nat Biotechnol 28, 935-942.

Dumontier, M. (2008). Situational Modeling: Defining Molecular Roles in Biochemical Pathways and Reactions. Proceedings of the 5th OWLED Workshop on OWL: Experiences and Directions.

Grenon, P., Smith, B., Goldberg, L. (2004). Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Tech Inf 102, 20–38.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S. eds. (2009). OWL2 Web Ontology Language: Primer. Latest version available at http://www.w3.org/TR/owl2-primer/

Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M. (2004). SWRL: A semantic web rule language combining OWL and RuleML. Latest version available at http://www.w3.org/Submission/SWRL/

IUPAC (2011) IUPAC Compendium of Chemical Terminology - the Gold Book. Latest version available at http://goldbook.iupac.org/

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L., Rosse, C. (2005). Relations in biomedical ontologies. Genome Biol 6, R46.