BiPOm: Biological interlocked Process Ontology for metabolism

How to infer molecule knowledge from biological process?

Vincent Henry[1*] Fatiha Saïs[2] Elodie Marchadier[3] Juliette Dibie[4] Anne Goelzer[1] and Vincent Fromion[1]

[1]INRA, UR1404, MaIAGE, Université Paris Saclay, Jouy-en-Josas, France
[2] LRI, UMR 8623, CNRS, Université Paris-Sud, Université Paris Saclay, Orsay, France
[3] GQE- Le Moulon, INRA, Université Paris Sud, CNRS, AgroParisTech, Ferme du Moulon, Gif-sur-Yvette, France
[4] UMR MIA-Paris, AgroParisTech, INRA, Université Paris Saclay, 75005, Paris, France

ABSTRACT

In this paper, we introduce a new ontology BiPOm[1]. describing metabolic processes as interlocked subsystems while explicitly formalizing the active states of the involved molecules. We further showed that the annotation of molecules such as molecular types or molecular properties can be deduced using SWRL rules and automatic reasoning on instances of BiPOm. The information necessary to instantiate BiPOm can be extracted from existing databases or existing bio-ontologies. Altogether, this results in a paradigm shift where the anchorage of knowledge is rerouted from the molecule to the biological process.

1 INTRODUCTION

The progress in high-throughput biology are currently limited or largely unexploited by the difficulty to manage efficiently the biological knowledge in life science. To address this challenge, bio-ontologies have been developed and stored in public repositories (Whetzel et al., 2011; Smith et al., 2007). The Gene Ontology (TheGeneOntologyConsortium, 2004) is probably the best current ongoing initiative of the hierarchical organization of biological knowledge. Initially GO integrates knowledge in the domain of the molecular functions of gene products (GO-MF), cellular components (GO-CC), and biological processes (GO-BP). Recently, the GO-plus project (Hill et al., 2013) integrated other bio-ontologies that were dedicated to the description of other subcellular entities such as bio-chemicals (ChEBI; Hastings et al., 2013) or sequence features (SO; Eilbeck et al., 2005). The on-going "Logical Extension of the Gene Ontology" project (LEGO; TheGeneOntology-Consortium, 2017) is going one step further with the design of refined properties between GO-BP and GO-MF. When represented in OWL[2] (Web Ontology Language), ontology may benefit from the reasoning power provided by the underlying logical semantics of their axioms (e.g. subsumption, disjunction, functionality of properties, cardinalities). Even if automatic reasoning can be used to ensure the consistency of the representation model of bio-ontologies and subsumption inferences, the possibilities of automatic reasoning are largely under-exploited in existing bio-ontologies. The main use of bio-ontologies is as a shared controlled vocabulary between researcher communities as bio-ontologies fix unambiguous definitions, synonyms and class annotations for biological knowledge. Therefore, bio-ontologies have been widely used to annotate and manage huge amounts of biological data from sparse databases and provided useful data annotations for bioinformatics algorithms, such as sequence similarities for new genome annotation (Balakrishnan et al., 2013). Actually, the current representation of the cell components is strongly linked to early developments of the molecular biology where the genes were the central object of biologists interests. The systematic and intensive efforts of the community led to huge progresses on the understanding of the cell functioning. However, these huge progresses are associated with the necessity to manage more and more new types of information. For instance, a protein can have multiple states and multiple functions. Current annotations index the protein to multiple functions independently of its potential different states (for example the post-translational modifications of gene-product).

Since the beginning of the 21st century, the systems biology community introduced a novel representation of the cell based on the concept of "systems" of engineering science (Kitano, 2001) where each piece of the system is described and fully characterized. In systems biology, the cell is considered as a system composed of interlocked subsystems having their own dynamics of operations. In this approach, a subsystem is a biological process (and its related biological subprocesses). A process is defined as (see Fig.1):

a) **Elementary** (such as biochemical reaction in biology) by its inputs and outputs (usually molecule states) and eventually by a specific mediator (such as an active enzyme or a macromolecule) and/or by an activity;

b) **Aggregated** (such as pathway) by its subprocesses.

The systemic representation of the cell has been shown to be highly efficient to manage the complexity of the cell (Kildegaard et al., 2013). In systems biology, the representation of the cell is thus process-centered and not gene- or molecule-centered. Therefore, the properties of the molecule (i.e. role and related function) do not depend anymore on the molecule existence or structure anymore (Balakrishnan et al., 2013) but is now conditioned by the biological process to which it belongs.

This paradigm-shift of the cell representation corresponds mainly to a change of point of view where the current annotation of the molecules can be reused and reroute to the biological processes.

As a proof-of-principle, we previously showed that the systemic

*To whom correspondence should be addressed: vincent.henry@inra.fr

[1] Available at `http://maiage.jouy.inra.fr/?q=en/biosys/ontology`

[2] `https://www.w3.org/TR/owl2-overview/`

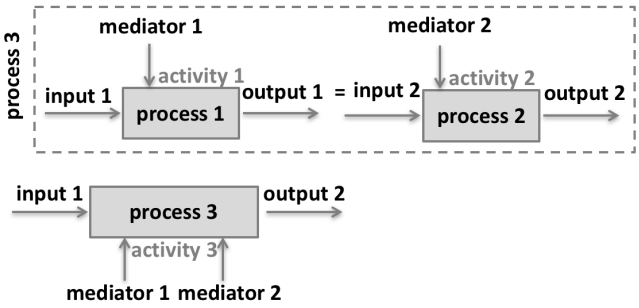

Fig. 1. Systemic representation of process: Process 1 and Process 2 are only represented with their systemic elementary definition. Process 3 is represented with its both systemic elementary and systemic aggregated definitions (dotted frame). These two kinds of representation correspond to two different scales: elementary and aggregated.

description of biological processes can be formalized as an ontological model[3](Henry et al., 2016). As a result, the fine description of more than 200 classes of processes and subprocesses for bacterial gene-expression can be related to a dozen classes of high-level process having a mathematical expression. Based on this previous work, we now show how BiPOm (Biological interlocked Process Ontology for metabolism), an ontology integrating only high level classes of metabolic process (described using the systemic approach) could 1) contain biological knowledge as instances and 2) use automatic reasoning through Semantic Web Rule Language (SWRL; Horrocks et al., 2004) in order to automatically infer, formalize and refine annotation of molecules. To do so, we introduce an ontological model carrying the main biological processes and molecular roles/functions at a high level of abstraction where the usual annotated resources are treated as instances. We apply the ontological model to describe a complex metabolic process, the *Arabidopsis thaliana*s "reductive pentose-phosphate cycle" (RPPC; also known as Calvin cycle), and illustrate how properties of the cell components participating to this metabolic pathway can be automatically inferred from the precise description of a process after logical reasoning.

2 ONTOLOGY OVERVIEW

This work aims at showing the substantial benefit of using an ontological model to describe molecular processes from a few knowledge on molecules. Our model was edited on Protégé (Musen and Team, 2015) and reasoning was performed using HermiT 1.3.8 (Glimm et al., 2014). BiPOm root is divided into three disjoint main classes: biological process, participant and activity (see Fig.2). All these classes contain few subclasses that corresponding to high-level classes imported from GO-BP, GO-MF, GO-CC, ChEBI and the Systems Biology Ontology (SBO; Courtot et al., 2011). In total BiPOm contains only 167 classes. Classes were formally defined with 9 "declared" properties (see Fig.2) and 3330 axioms. Moreover, 27 SWRL rules were defined for representing new knowledge and for supporting additional inferences (Krisnadhi et al., 2011).

[3] http://purl.bioontology.org/ontology/BIPON

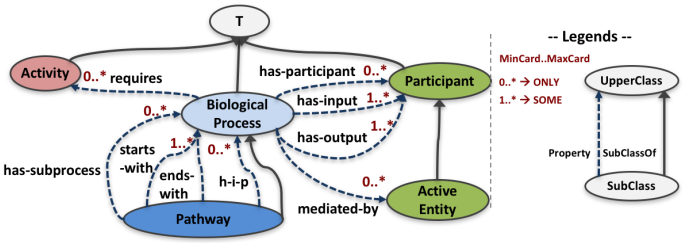

Fig. 2. Necessary and sufficient Classes and Object properties to formally define biological processes in BiPOm. (h-i-p: has-intermediary-process)

2.1 Process description

In systemics, a single process can be defined in two ways: (see Fig.1).

1. "**Systemic elementary process**" which has an input ("has_input" property) and an output ("has_output" property) at least one participant and may be (only) mediated by ("mediated_by" property) an active cell component and may require (only "requires" property) some molecular activity. It is reflected in our model by the "biological processes" class in BiPOm (see Fig. 2).
2. "**Aggregated process**" which starts with ("starts_with" property) and ends with ("ends_with" property) at least one biological process and may (only) have as subprocesses ("has_intermediary_process" property) some biological subprocesses. It is represented in our ontology by the "Pathway" class in BiPOm (see Fig. 2).

According to this formal definition, we designed several types of high-level and disjoint biological processes depending on the cardinality of their participants (see Fig. 2A):

1. "biological spontaneous process" or "biological mediated process";
2. "metabolic process" or "gene product modification process";
3. "non-covalent binding" or "dissociation".

Then these processes may be specified according to the role of their participant. For instance, an "enzymatic reaction" is a "biological mediated process" mediated_by one "enzyme", that "requires" "catalytic activity" and that "has_input" "substrate" and "has_output" "product" (see Fig. 3B); or an "activation" is a "post-translational protein modification" that "has_output" an "Active gene-product" (see Fig. 3C). Therefore, roles are included as "participant" subclasses. We have furthermore introduced operations allowing the combination of elementary processes. For instance, an "Enzymatic metabolic reaction" is a "metabolic process" and an "enzymatic reaction" (see Fig. 2B). In the same way, a "protein complex assembly" is a combination of "biological spontaneous process", a "non-covalent binding" and a "post-translational protein modification" (see Fig. 3C). Finally, two types of aggregated processes were designed to manage elementary processes. They were defined according to their subprocesses:

- "metabolic pathway" "has_subprocess" only "metabolic reaction" (see Fig. 3B);

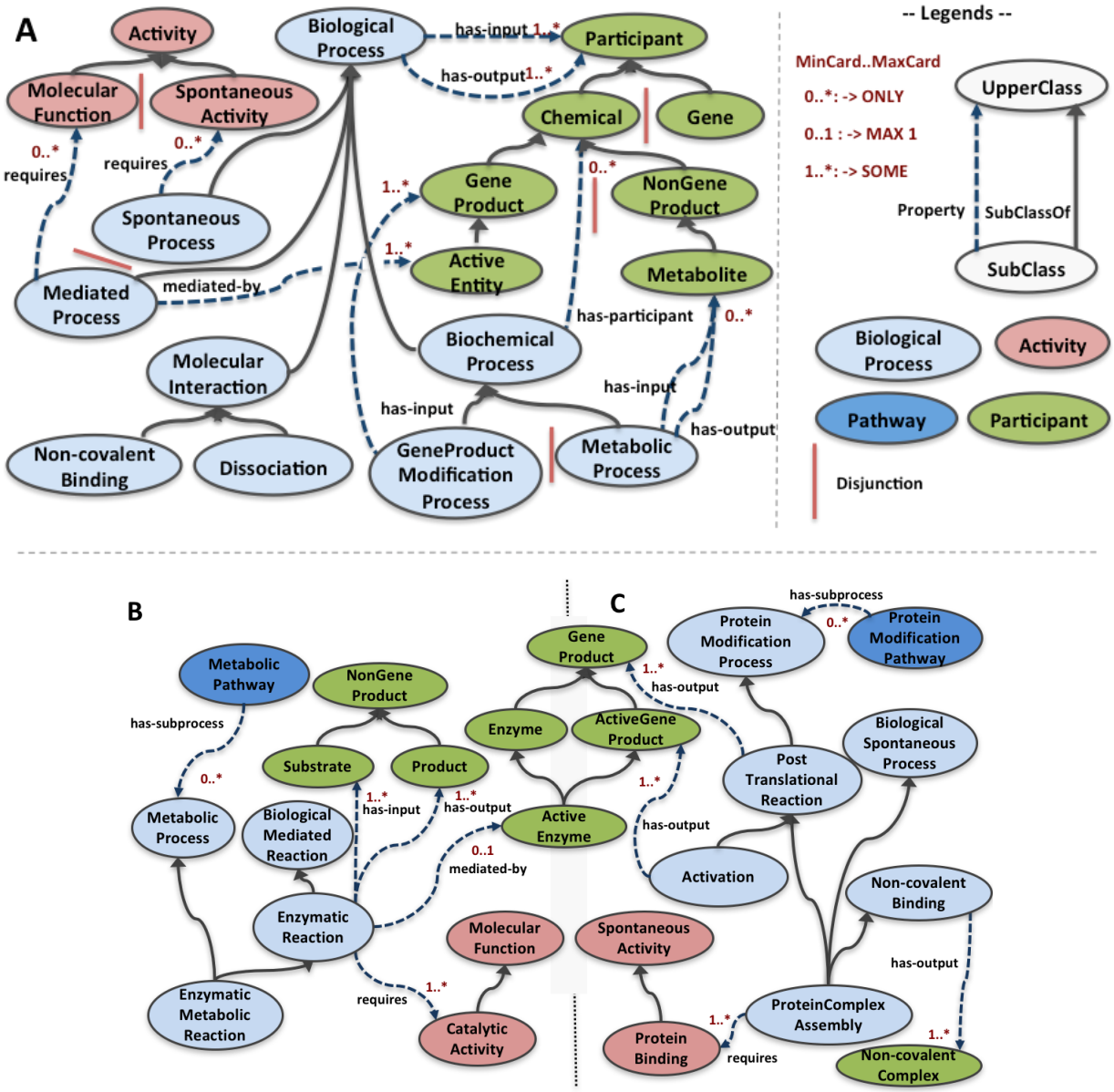

Fig. 3. BiPOm ontology: A) general presentation of BiPOm high level class and related properties. B/C) Contextualization of 3 of the 69 specified processes: B) "enzymatic metabolic reaction", C) "activation" and "protein complex assembly".

- "protein modification pathway" "has_subprocess" only "protein modification process" (see Fig. 3C).

Altogether our model contains 65 processes with a maximum depth of 7.

2.2 Participant

Participant is divided into two disjoint subclasses: "gene-product" and "non-gene product". "gene-product" refers to macromolecules or macromolecular complexes. The "gene product" initially depends on a gene (corresponding to macromolecules usually annotated by GO; TheGeneOntologyConsortium, 2004). On the other side, "non-gene product" refers to common biochemical that cannot be discriminated from a specie to another one and typically belong to ChEBI classes (Hill et al., 2013). Classes that refer to a participant role (e.g. enzyme, cofactor, metabolite) are subClassOf "Participant", "gene-product" or "non-gene product".

2.3 Activity

The class "activity" is divided into two disjoint subclasses: "molecular function" including few GO-MF classes such as "catalytic activity" or "chaperoning activity" and "spontaneous ability" including few GO-MF "binding" classes.

2.4 Rule design

We designed 9 properties that have to be declared by the user (see Fig.2). We also define 47 other properties or rules based on the previous declared ones to automatically assert new type or property of biological interest. Some new properties are defined as inverse properties (e.g. the "input_of" property is defined as the InverseOf property of "has_input"; the "mediates" property is defined as the InverseOf property of "mediated_by") and the SWRL rules are defined in order to automatically associate the new properties to instances. New inferred knowledge may concern the type (i.e. metabolite, enzyme), the molecular composition (i.e. "has_molecular_part"), the molecular interaction ("interact_with"), the contribution (i.e. "contributes_to") or the functionality (i.e. "has_function") of participants. Let us give three examples of defined SWRL rules in close relationship in BiPOm:

• If a "gene-product" gp mediates a "mediated reaction" r that requires a "molecular function" f, then gp has_function f. This can be expressed in a SWRL rule R_1 as follows:

$$\begin{aligned} R_1: \quad & \text{GeneProduct}(gp) \wedge \text{MediatedReaction}(r) \cdots \\ & \wedge \text{MolecularFunction}(f) \wedge \text{mediates}(gp, r) \wedge \text{requires}(r, f) \\ & \Rightarrow \text{has\_function}(gp, f) \end{aligned}$$

• If a "participant" p_0 output_of a "protein complex assembly" $proc_a$ then p_0 has_molecular_part "participant" p_i that is input_of $proc_a$ and p_i that are simple proteins are typed as "Protein Complex Subunit". While a protein complex assembly may be mediated by an ATP-dependent chaperone, we assume that p_0 must be different from the ATP residue: ADP and phosphate. This can be expressed in a SWRL rule R_2 as follows:

$$\begin{aligned} R_2: & \text{ProteinComplexAssembly}(proc_a) \wedge \text{has\_output}(proc_a, p_0) \cdots \\ & \wedge \text{DifferentFrom}(p_0, ADP) \wedge \text{DifferentFrom}(p_0, \text{phosphate}) \cdots \\ & \wedge \text{has\_input}(proc_a, p_i) \wedge \text{SimpleProtein}(p_i) \\ & \Rightarrow \text{has\_molecular\_part}(p_o, p_i) \wedge \text{ProteinComplexSubunit}(p_i) \end{aligned}$$

• At least, if a p_0 mediates r that requires f then p_i contributes_to f. This can be expressed in a SWRL rule R_3 as follows:

$$\begin{aligned} R_3 \quad & : \text{has\_function}(r, f) \wedge \text{molecular\_part\_of}(p_0, p_i) \\ & \Rightarrow \text{contributes\_to}(p_i, f) \end{aligned}$$

Some properties were also defined using SWRL rules to provide information on the relative order of elementary processes (e.g. precedes and its inverse: preceded_by). They are able to order reactions in a pathway. Precedes and preceded_by are used to infer participant of pathway, excluding those that are produced and consumed by consecutive reactions. Let us detail these rules:

• If a pathway P has subprocesses "biological process" p_1 and p_2 and p_1 has_output a "participant" mol and p_2 has_input mol, then p_1 precedes p_2. A SWRL R_4 can be expressed as follows:

$$\begin{aligned} R_4: \quad & \text{has\_subprocess}(P, p_1) \wedge \text{has\_subprocess}(P, p_2) \cdots \\ & \wedge \text{has\_output}(p_1, mol) \wedge \text{has\_input}(p_2, mol) \cdots \\ & \wedge \text{DifferentFrom}(p_1, p_2) \\ & \Rightarrow \text{precedes}(p_1, p_2) \end{aligned}$$

• If $Pagg$ has_subprocess p_1 and p_2, and p_1 has_output a "participant" $molAgg$ and p_1 precedes p_2 and not precedes by p_2 and if p_2 has_input another "participant" $molIn$ different from $molAgg$, then $Pagg$ has_input $molAgg$. A SWRL R_5 can be expressed as follows:

$$\begin{aligned} R_5: \quad & \text{has\_subprocess}(Pagg, p_1) \wedge \text{has\_output}(p_1, molAgg) \cdots \\ & \wedge \text{has\_subprocess}(Pagg, p_2) \wedge \text{DifferentFrom}(p_1, p_2) \cdots \\ & \wedge \text{has\_input}(p_2, molAgg) \wedge \text{precedes}(p_1, p_2) \cdots \\ & \wedge \text{preceded\_by}(p_1, p_x) \wedge \text{DifferentFrom}(p_2, p_x) \cdots \\ & \wedge \text{has\_input}(p_2, molIn) \wedge \text{DifferentFrom}(molAgg, molIn) \\ & \Rightarrow \text{has\_input}(Pagg, molIn) \end{aligned}$$

As a last example, rules also allow the interaction of protein transient interaction with other proteins that mediate post translational reaction (such as interact with annotation in Uniprot):

• If a process $proc$ has input a protein $prot_1$ and is mediated by another protein $prot_2$, then $prot_1$ interacts with $prot_2$. A SWRL R_6 can be expressed as follows:

$$\begin{aligned} R_6: \quad & \text{protein}(prot_1) \wedge \text{has\_input}(proc, prot_1) \cdots \\ & \wedge \text{mediated\_by}(proc, prot_2) \wedge \text{DifferentFrom}(prot_1, prot_2) \\ & \Rightarrow \text{interacts\_with}(prot_1, prot_2) \end{aligned}$$

2.5 Minimum information for instantiation

Thanks to logical rules, only few incoming assertions are necessary to describe an ontological process and instantiate the ontological model. Briefly, instances of participants have to be type by gene-product or non-gene-product. instances of processes have to be typed by one of the 69 different processes with (a) the description of its inputs, outputs and mediators (if any) and requirements and/or (b) the description of its start, intermediary and end processes. This information can easily be structured in a data table and imported in the ontology from the cellfie plugin (Kola and Rector, 2007).

3 USE CASE (FIG. 4)

We considered the *Arabidopsis thaliana*s RPPC available[4] on Plant reactome (Naithani et al., 2017). This metabolic process is present in photosynthetic organisms, well described in the literature and representative of the complexity of metabolic processes. The RPPC is an essential cyclic process enabling the CO_2 fixation and composed of 10 chemical reactions. Each chemical reaction is catalyzed by an active enzyme or enzymatic complex. The process of enzyme activation involves many different post-translational modifications, chaperoning, and complexations. In particular, the Ribulose Bisphosphate CarbOxylase (RuBisCO) enzyme that catalyzes the first reaction is an enzymatic complex composed of 16 subunits encoded by 5 genes. The activation of RuBisCO is achieved through many steps (spontaneous and chaperone-dependent complexation, carbamylation, magnesium binding, etc.) (Andersson and Backlund, 2008). Other enzymes of the cycle are also controlled at the post-translational level by redox-reactions transfer of disulfide bonds (generic in biology). Traditionally, the different steps of enzyme or enzymatic complex activation are poorly described in bio-ontologies.

[4] http://plantreactome.gramene.org/PathwayBrowser/#/R-ATH-1119519&SEL=R-ATH-5149661&PATH=R-ATH-2744345,R-ATH-2883407&DTAB=MT

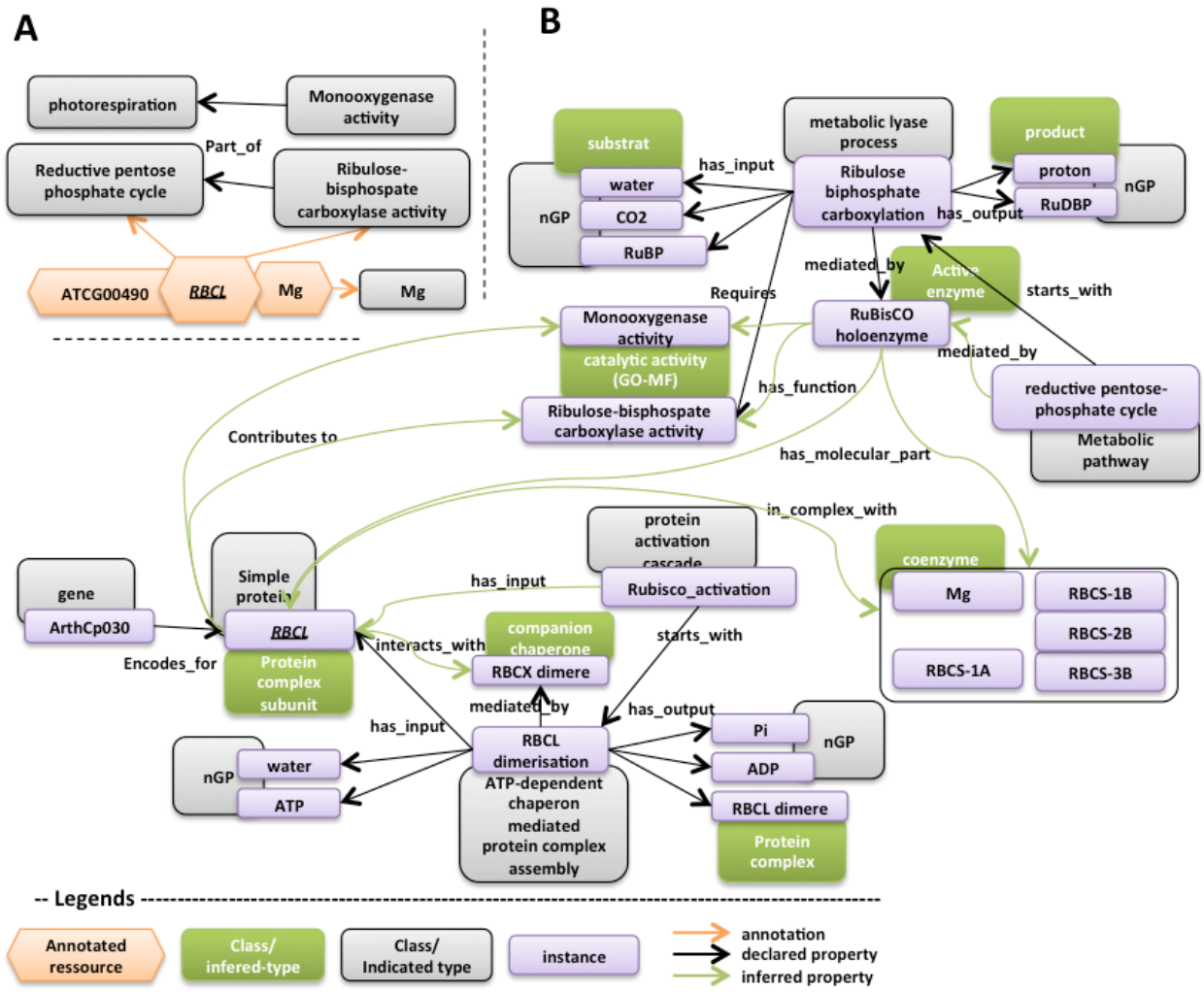

Fig. 4. Information on RuBisCO large subunit supported by ontological resources: (A) Annotation in Uniprot or AmiGO2, (B) formal relation using BiPOm. (RAF1: Rubisco accumulation factor 1; RBCL: Ribulose bisphosphate carboxylase large chain; RBCX; Chloroplastic Chaperonin-like RBCX protein. GP: gene-product; NGP: non-gene product)

In the standard gene-centered annotation, information is anchored manually to protein complex subunits. For instance, in Amigo2 the RuBisCO large chain (RBCL; O03042) is annotated 1) by the functions of the RuBisCO: Ribulose-bisphospate carboxylase activity (GO:001698) and Monooxygenase activity (GO:0004497) and 2) by the process involving RuBisCO: reductive pentose-phosphate cycle (GO:0019253) and Photorespiration (GO:0009853) (see Fig. 4A). In Uniprot, the RBCL knowledge is merged with the encoding gene knowledge. The information is completed in natural language describing Ribulose biphosphate carboxylation enzymatic reaction mediated by RuBisCO. Moreover, cofactors of RuBisCO such as Magnesium are also described in natural language. Cross-references are managed using a link to ChEBI for Magnesium (annotated by magnesium(2+): CHEBI: 18420). Due to the ambiguity between full RuBisCO complex and the related subunits and genes, subunits are annotated like the full RuBisCO complex. In BiPOm, we used the same knowledge while we finely described the elementary enzymatic processes of RPPC and the enzyme activation processes according to natural language found in Uniprot and related publications. It results in the description of 2 pathways (RPPC and RuBisCO activation pathway) and 82 biological reactions: 24 enzymatic metabolic reactions for the RPPC and 58 post-translational protein modifications involved in the RPPC-enzyme activation (e.g. RPPC starts with RuBP carboxylation that is mediated by RuBisCO holoenzyme and RuBisCO activation starts with RBCL dimerisation having RBCL as input and ends with RuBisCO-Mg complexation). RuBisCO activation reactions involve different states of RuBisCO: in complex with chaperones, uncarbamylated, carbamylated and associated with Mg (holoenzyme). Following automatic reasoning, RBCL and Mg are typed by “Protein Complex Subunit” and “coenzyme”, respectively (see Fig. 4B). RuBisCO holoenzyme is typed by “active entity”, “holoenzyme”, “lyase”, “oxidoreductase” and “protein complex”. Information on RuBisCO or its subunit RBCL

are disjoint (see Fig. 4B): while RuBisCO holoenzyme has the function, RBCL contributes to the function. Moreover, we obtained computationally interpretable information from natural language section of Uniprot, e.g. relationship between the inputs / outputs and the reactions are formally related with the "has_input" and "has_output" properties, the protein complex and their subunits or coenzyme are formally related with the "has_molecular_part" property or the protein complex component together are formally related with the "in complex_with" property.

4 CONCLUSION AND PERSPECTIVES

Here we introduced a new ontology BiPOm describing metabolic processes as interlocked subsystems. We explicitly handled the different states of molecules including the "active state" involved in a biological process. Using SWRL rules and automatic reasoning on instances of BiPOm, we inferred annotations of cellular entities such as molecular types or molecular properties. Few information is required to instantiate BiPOm and can be extracted from existing databases or bio-ontologies. Our approach is actually to take advantage of existing public repositories to finely describe biological processes and to extend the use of bio-ontologies from controlled vocabularies only to automatic reasoning instead. We assume that this paradigm shift where the anchorage of knowledge is rerouted from the molecule to the process could thus be benefit to the biological knowledge organization. The use case of *A. thaliana*s RPPC is typical of the complexity of metabolic processes and is thus highly informative of the added value of automatic reasoning on instances. We inferred new annotations on that cycle compared to the classical annotation stored in public repositories such as amigo2 or Uniprot. Due to its flexibility, our ontology could be straightly extended with the localization of molecules or with other biological processes such as the gene-expression (Henry et al., 2016). This requires the integration of new types of participants such as sequence patterns of bio-informative molecules (Eilbeck et al., 2005) and the integration of new molecular properties such as "polymerase", "transcription factor" or "termination factor" that characterize gene-expression processes. Eventually, each biological process can be related to its mathematical model having its parameters (Henry et al., 2016). New HTO provide quantitative multi-level information including notably the observation on the states of cellular entities. The limiting factor of biological approaches is now no more the data generation but the development of approaches allowing the extraction of information from these data. BiPOm provides a promising framework to address the current multi-level big data challenge in biology. As it provides a formal rational framework to relate multi-level HTO together by considering the cellular system (or the organism) as a whole and by making easier the reasoning on system components.

ACKNOWLEDGMENTS

This work has been funded by the French Lidex-IMSV of the Université Paris Saclay.

REFERENCES

Andersson, I. and Backlund, A. (2008). Structure and function of rubisco. *Plant Physiology and Biochemistry*, 46(3):275 – 291.

Balakrishnan, R., Harris, M. A., Huntley, R., Van Auken, K., and Cherry, J. M. (2013). A guide to best practices for gene ontology (GO) manual annotation. *Database*.

Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., Dumontier, M., Finney, A., Golebiewski, M., Hastings, J., Hoops, S., Keating, S., Kell, D. B., Kerrien, S., Lawson, J., Lister, A., Lu, J., Machne, R., Mendes, P., Pocock, M., Rodriguez, N., Villeger, A., Wilkinson, D. J., Wimalaratne, S., Laibe, C., Hucka, M., and Le Novère, N. (2011). Controlled vocabularies and semantics in systems biology. *Molecular Systems Biology*, 7(1).

Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, 6(5):R44.

Glimm, B., Horrocks, I., Motik, B., Stoilos, G., and Wang, Z. (2014). HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning*, 53(3):245–269.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., and Steinbeck, C. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research*, 41(D1):D456.

Henry, V., Ferré, A., Froidevaux, C., Goelzer, A., Fromion, V., Boulakia, S. C., Dérozier, S., Dinh, M., Fiévet, G., Fischer, S., Gibrat, J., Loux, V., and Peres, S. (2016). Représentation systémique multi-échelle des processus biologiques de la bactérie. In *IC 2016 : 27ème Journées francophones d'Ingénierie des Connaissances (Proceedings of the 27th French Knowledge Engineering Conference), Montpellier, France, June 6-10, 2016.*, pages 97–102.

Hill, D. P., Adams, N., Bada, M., Batchelor, C., Berardini, T. Z., Dietze, H., Drabkin, H. J., Ennis, M., Foulger, R. E., Harris, M. A., Hastings, J., Kale, N. S., de Matos, P., Mungall, C. J., Owen, G., Roncaglia, P., Steinbeck, C., Turner, S., and Lomax, J. (2013). Dovetailing biology and chemistry: integrating the gene ontology with the ChEBI chemical ontology. *BMC Genomics*, 14:513–513. 1471-2164-14-513[PII].

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosofand, B., and Dean, M. (2004). SWRL: A semantic web rule language combining OWL and RuleML.

Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E., and Betenbaugh, M. J. (2013). The emerging cho systems biology era: harnessing the omics revolution for biotechnology. *Current Opinion in Biotechnology*, 24(6):1102 – 1107.

Kitano, H. (2001). *Foundations of Systems Biology*. MIT press, Cambridge.

Kola, J. S. and Rector, A. (2007). Importing spreadsheet data into Protégé: Spreadsheet plug-in. In *Proc. of the Intl. Protégé Conference*.

Krisnadhi, A., Maier, F., and Hitzler, P. (2011). OWL and rules. In Polleres, A., d'Amato, C., Arenas, M., Handschuh, S., Kroner, P., Ossowski, S., and Patel-Schneider, P., editors, *Reasoning Web. Semantic Technologies for the Web of Data: 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures*, pages 382–415. Springer Berlin Heidelberg.

Musen, M. A. and Team (2015). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, 1(4):4–12.

Naithani, S., Preece, J., D'Eustachio, P., Gupta, P., Amarasinghe, V., Dharmawardhana, P. D., Wu, G., Fabregat, A., Elser, J. L., Weiser, J., Keays, M., Fuentes, A. M.-P., Petryszak, R., Stein, L. D., Ware, D., and Jaiswal, P. (2017). Plant reactome: a resource for plant pathways and comparative analysis. *Nucleic Acids Res*, 45(Database issue):D1029–D1039.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S. (2007). The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255.

TheGeneOntologyConsortium (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Res*, 32:D258–D261.

TheGeneOntologyConsortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*, 45:D331–D338.

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., and Musen, M. A. (2011). Bioportal: enhanced functionality via new Web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Res*, 39:W541–W545.