



**HAL**  
open science

## Detection and correction of non-conformities and redundancies in complexes of molecules in BioPAX

Camille Juigné, Olivier Dameron, François Moreews, Florence Gondret,  
Emmanuelle Becker

### ► To cite this version:

Camille Juigné, Olivier Dameron, François Moreews, Florence Gondret, Emmanuelle Becker. Detection and correction of non-conformities and redundancies in complexes of molecules in BioPAX. Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), Jul 2022, Rennes, France. pp.1-25. hal-03752473

**HAL Id: hal-03752473**

**<https://hal.inrae.fr/hal-03752473v1>**

Submitted on 16 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Detection and correction of non-conformities and redundancies in complexes of molecules in BioPAX

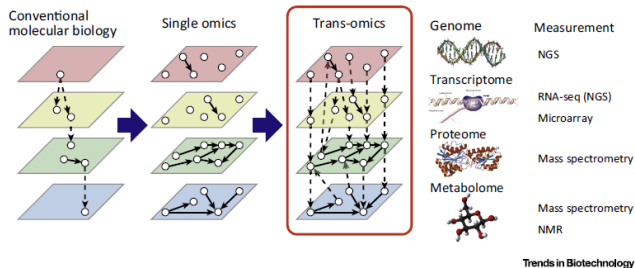
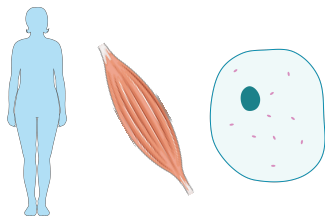
Camille JUIGNÉ<sup>1,2</sup> Olivier DAMERON<sup>1</sup> François MOREEWS<sup>1,2</sup>  
Florence GONDRET<sup>2</sup> Emmanuelle BECKER<sup>1</sup>

<sup>1</sup>Univ Rennes, Inria, CNRS, IRISA - UMR 6074, F-35000 Rennes, France

<sup>2</sup>PEGASE, INRAE, Institut Agro, 35590, Saint Gilles, France

July 5, 2022

# Biological context: understand how biological systems adapt to their environment

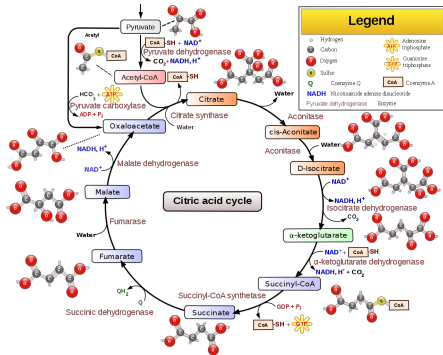


Understand the organization of biological pathways at different scales

# Biological context: metabolic pathways and complexes

## Biological pathway

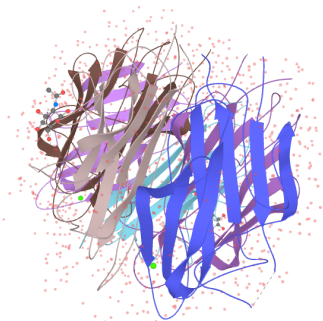
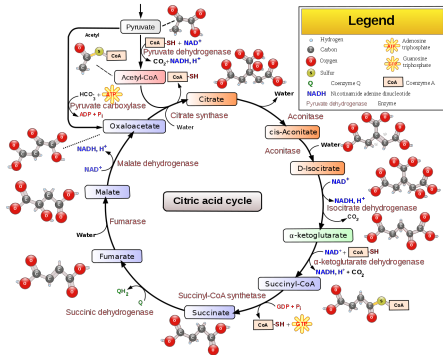
"a series of actions among molecules in a cell that leads to a certain product or a change in the cell" (NIH)



# Biological context: metabolic pathways and complexes

## Biological pathway

"a series of actions among molecules in a cell that leads to a certain product or a change in the cell" (NIH)



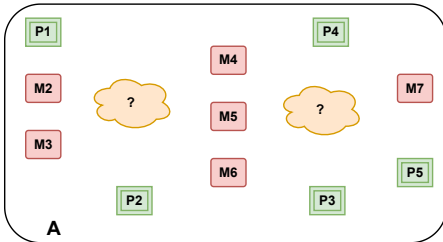
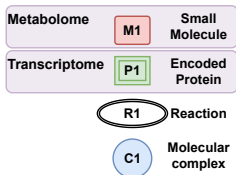
R-HSA-173584

## Complexes and interactions in biology

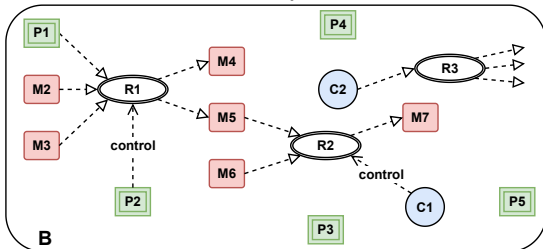
- Chemical assembly of **several molecules**
- Can either **participate in** or **control** interactions

# Biological context: complexes play a major role in pathways

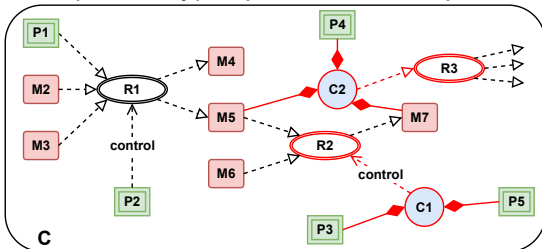
What we consider



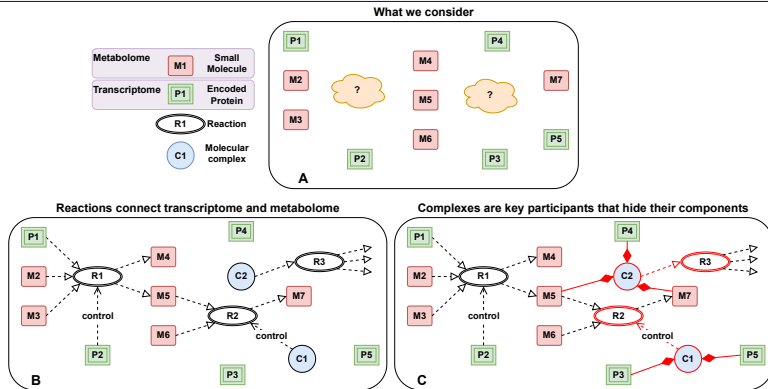
Reactions connect transcriptome and metabolome



Complexes are key participants that hide their components



# Impact on the graph topology



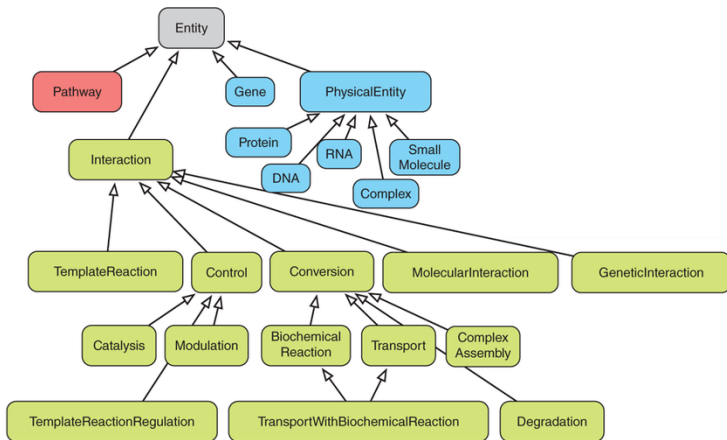
Taking into account (in)valid complexes is required for analyzing:

- Interactions in which a molecule participates
- Molecules participating in an interaction

# Computational context: Biological Pathway Exchange format

## Database of biological pathways in BioPAX

- **Reactome, KEGG, PathwayCommons...**
- Well established **ontology** to represent pathways at molecular and cellular levels
- Represented in **graphs** (RDF and OWL)
- Can be queried with **SPARQL**



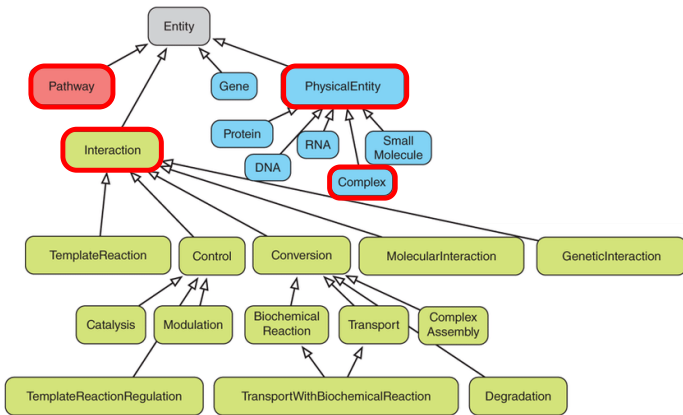
Demir et al. (2010)



# Computational context: Biological Pathway Exchange format

## Database of biological pathways in BioPAX

- Reactome, KEGG, PathwayCommons...
- Well established **ontology** to represent pathways at molecular and cellular levels
- Represented in **graphs** (RDF and OWL)
- Can be queried with **SPARQL**



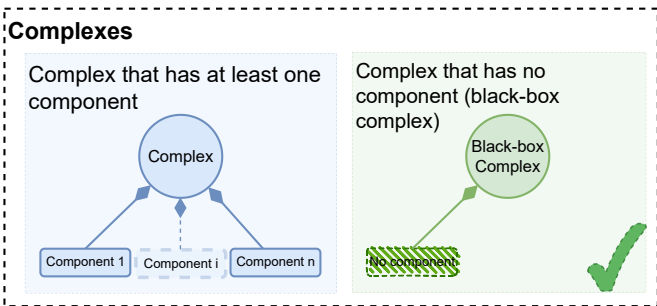
Demir et al. (2010)

In the Reactome database complexes participate to **32%** of the interactions

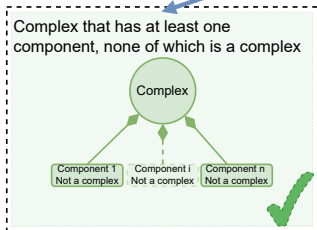
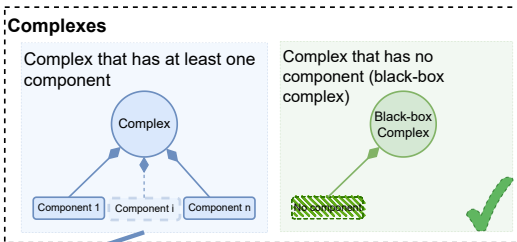
# Issue: Complexes composed of other complexes

A complex cannot be composed of other complexes

The components of a complex cannot have a component

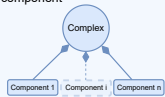


# Issue: Complexes composed of other complexes

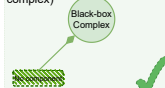


## Complexes

Complex that has at least one component



Complex that has no component (black-box complex)

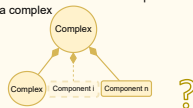


Complex that has at least one component, none of which is a complex



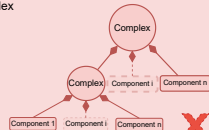
Complex that has at least one component that is a complex

Complex that has at least one component that is a complex

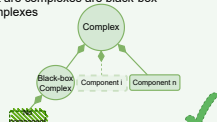


Complex that has at least one component that is another invalid complex

Complex that has at least one component that is a (valid) complex but not a black-box complex



Complex that has at least one component that is a complex, and all its components that are complexes are black-box complexes



We observed some invalid complexes in Reactome (not detected by the BioPAX validator)

# Objectives

---

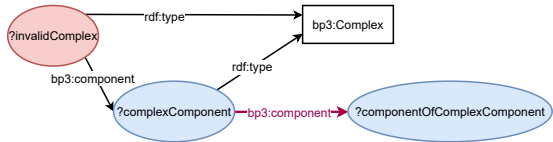
1. Identify invalid complexes
2. Fix invalid complexes
3. Evaluation of the benefits of the procedure

# Contrib 1: Identify and quantify invalid complexes

Invalid complexes are composed of  $\geq 1$  complex with components

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bp3: <http://www.biopax.org/release/biopax-level3.owl#>

SELECT DISTINCT ?invalidComplex
WHERE {
  ?invalidComplex rdf:type bp3:Complex .
  ?invalidComplex bp3:component ?complexComponent .
  ?complexComponent rdf:type bp3:Complex .
  ?complexComponent bp3:component ?componentOfComplexComponent .
}
```

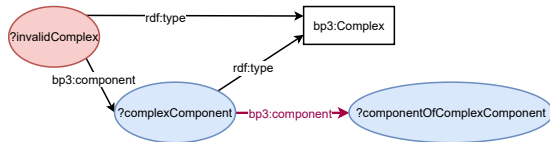


# Contrib 1: Identify and quantify invalid complexes

Invalid complexes are composed of  $\geq 1$  complex with components

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bp3: <http://www.biopax.org/release/biopax-level3.owl#>
```

```
SELECT DISTINCT ?invalidComplex
WHERE {
  ?invalidComplex rdf:type bp3:Complex .
  ?invalidComplex bp3:component ?complexComponent .
  ?complexComponent rdf:type bp3:Complex .
  ?complexComponent bp3:component ?componentOfComplexComponent .
}
```



Use case



reactome

(v79)

**Homo sapiens:** 39% complexes are invalid out of 14,840

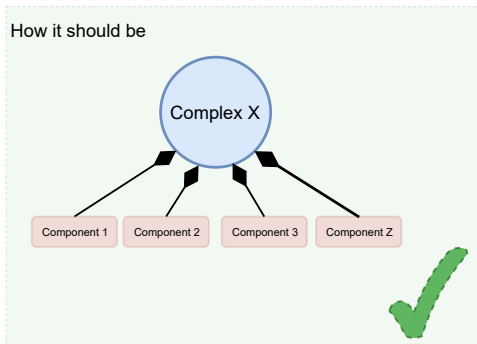
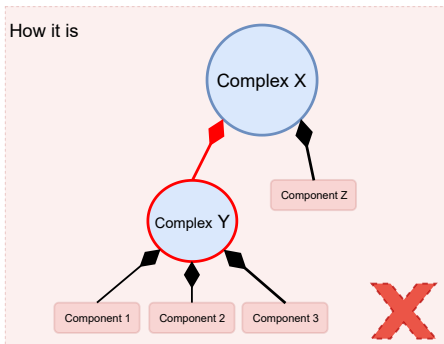
**Mus musculus:** 39% complexes are invalid out of 10,761

**Sus scrofa:** 40% complexes are invalid out of 7,769

Complexes represent a large fraction of biological entities

Invalid complexes are present in large quantities in the data sets of different organisms

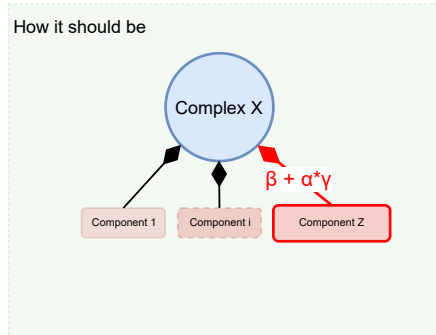
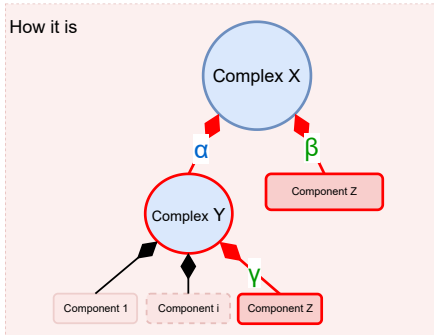
## Contrib 2: Fix the invalid complexes



Collapse as direct components all the (in)direct components that do not have component



## Contrib 2: Fix the invalid complexes

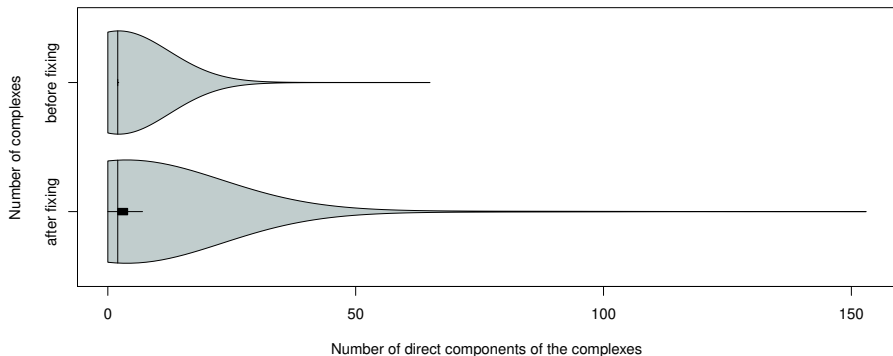


$$S(Z) = \sum_{p \in \text{parent nodes}}^P S_p(Z) * S(p)$$

Stoichiometry has to accommodate the fact that components can occur at several places

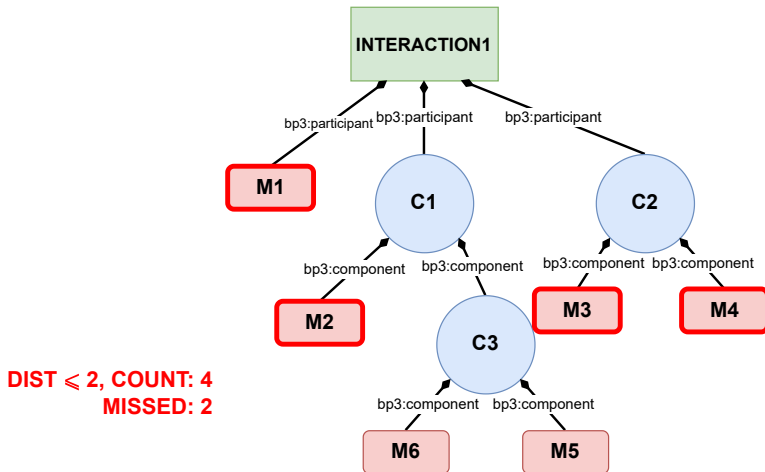
## Contrib 3: Homo sapiens Reactome use-case (repair)

All invalid complexes were fixed



Fixing invalid complexes increases the number of direct components

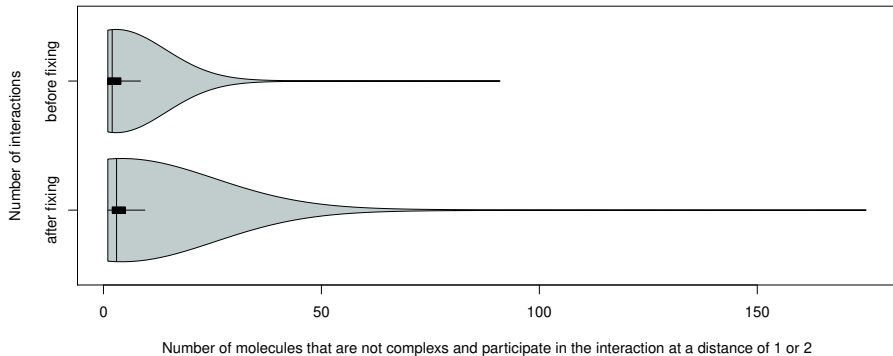
# Number of molecules that are not complexes and participate in interactions at a distance of 1 or 2



2 molecules missed due to invalid complexes

# Impact on the graph topology

---



Taking into account invalid complexes has a strong impact on the interaction graph topology

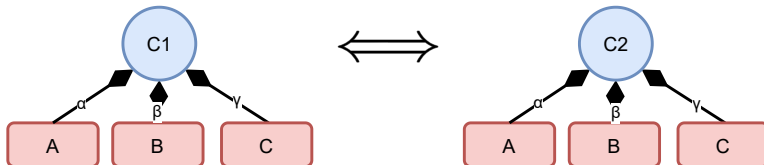
## Side effect: detection of artificial redundancy (Homo Sapiens)

---

What we call redundant complexes:

Complexes that share the **same components** with the **same stoichiometric coefficients** and have the **same cellular location**

Before fixing, we identified **241** of these complexes

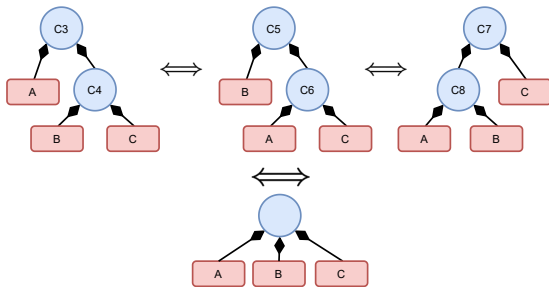


## Side effect: detection of artificial redundancy (Homo Sapiens)

What we call redundant complexes:

Complexes that share the **same components** with the **same stoichiometric coefficients** and have the **same cellular location**

Before fixing, we identified **241** of these complexes



Fixing invalid complexes allowed to identify **92** additional redundant complexes (+38%)

# Conclusion

---

- Semantically-rich queries for identifying and fixing invalid complexes that are **reproducible** on other databases
- Improves the conformity and the analysis of the graph by repairing the **topology**
- Will allow to **apply reasoning methods on better quality data**
- Will allow a better understanding of the **regulation of complex phenotypes**
- Side effect of **allowing the detection of complex redundancies**
- Essential methodology to **analyse and advance in the knowledge of biological processes**

# Acknowledgments

---



Team Croissance



camille.juigne@irisa.fr



# Appendix

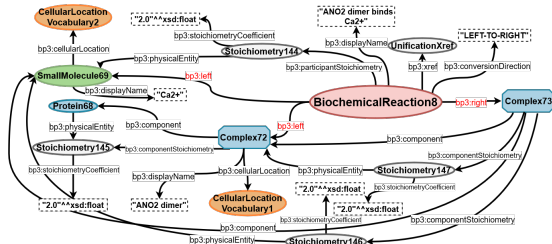
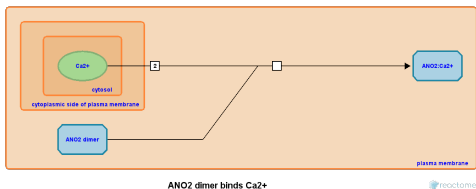
---

# Issue: Complexes composed of other complexes

Complex component: defines the subunits of a complex (BioPAX v3 spec)

"This property should not contain other complexes, i.e. it should always be a flat representation of the complex. [...]"

Exceptions are black-box complexes (i.e. complexes in which the component property is empty), which may be used as component's of other complexes because their parts are unknown."



We observed some invalid complexes in Reactome (not detected by the BioPAX validator)