Representation of biological data, information and knowledge: opportunities offered by systemic biology

V. Fromion
vincent.fromion@inra.fr
INRA, MaIAGE, Université Paris-Saclay, Jouy-en-josas

DATAIA-JST International Symposium on Data Science and AI

Paris, July 11, 2018

A lot of progress in the biological field in 3 complementary aspects

- Enormous progress made by biologists in the understanding of the functioning of living systems (in particular at infra-cell scale)
- New observation technologies combined with a continuous increase in measurement quality and a drastic reduction in costs, i.e. the data Deluge !
- Great progress in the multi-scale integration of living systems, in particular for the bacterial cell through the use of the systems biology approach, i.e. Kitano's approach

The presentation focuses on issues raised by knowledge, information, data and model management attached to a bacterial cell (infra and cell scale)

DATA deluge in the biological field …

Omics technologies

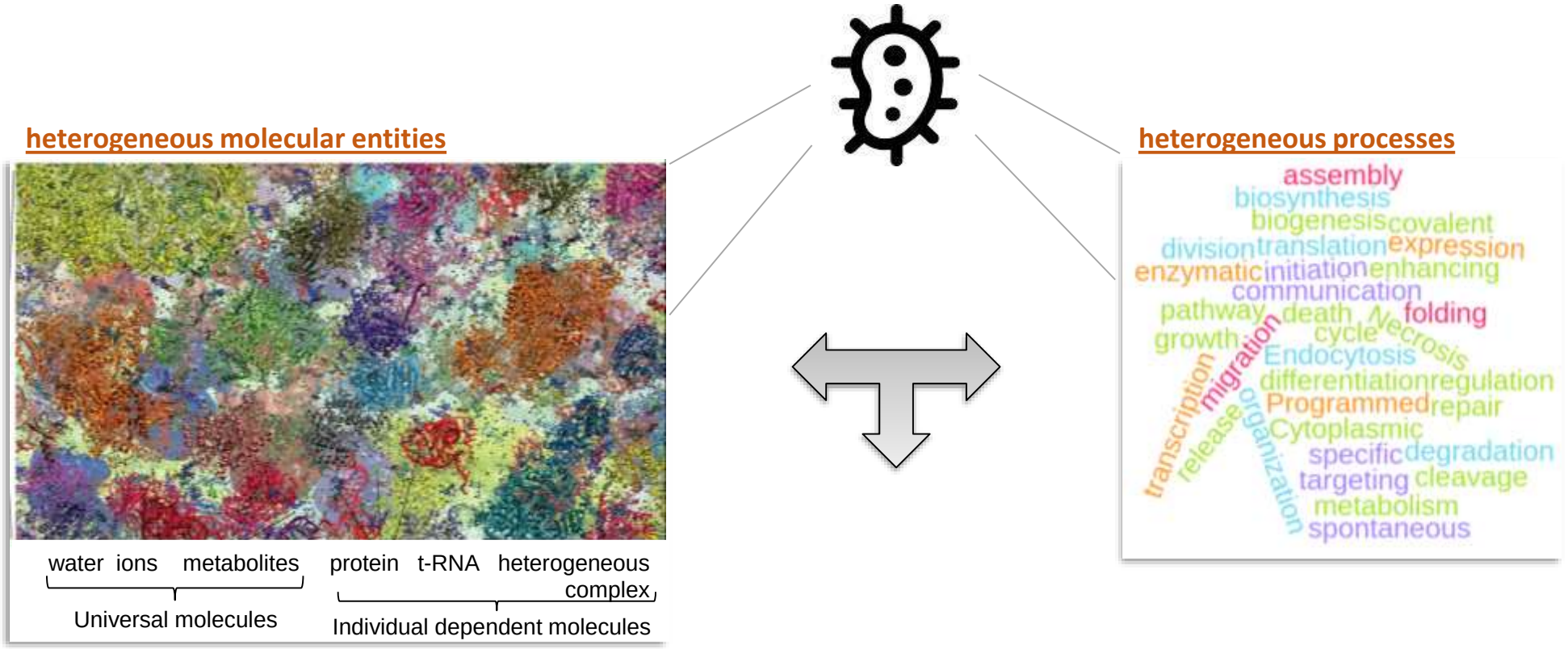
Biological data production

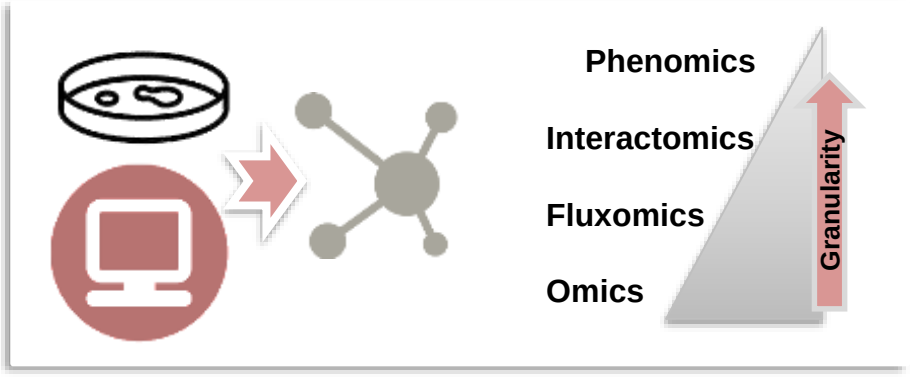
Needs for data and knowledge management

Bio-ontologies: controlled vocabularies for

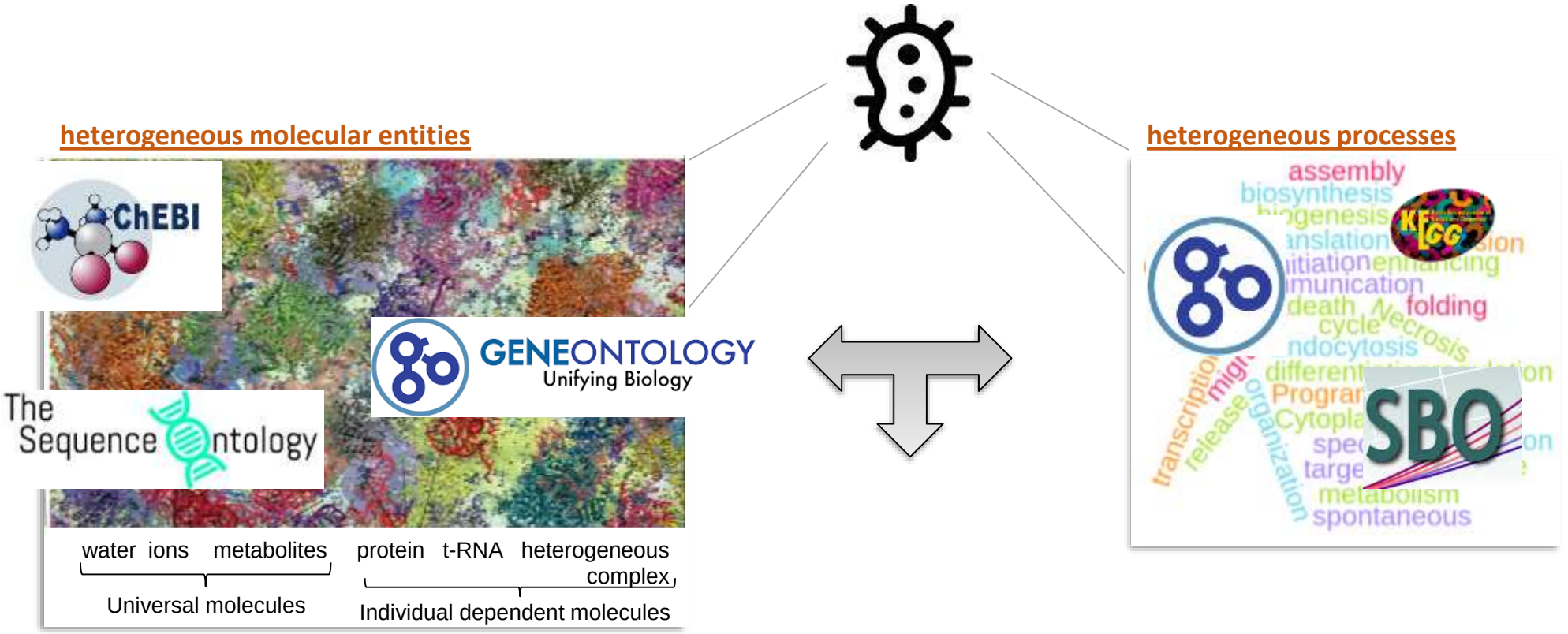
- Knowledge representation in Biology
- Structuration
- Indexation/annotation
- Data sharing
- Search retrieval

Cellular and molecular biology is a wide and heterogeneous field

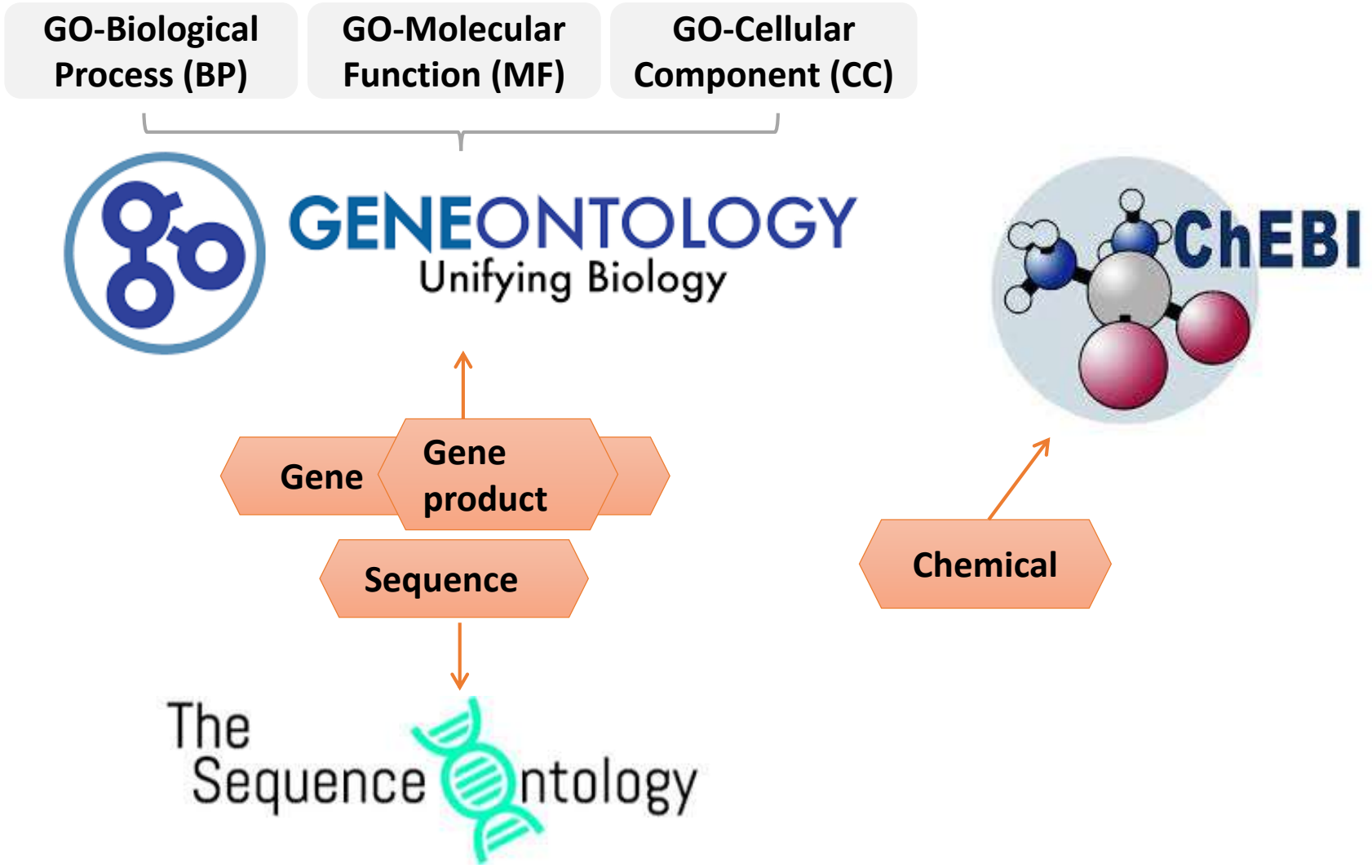




Bio-ontologies are useful tools to formalize biological knowledge representation…

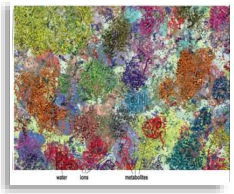


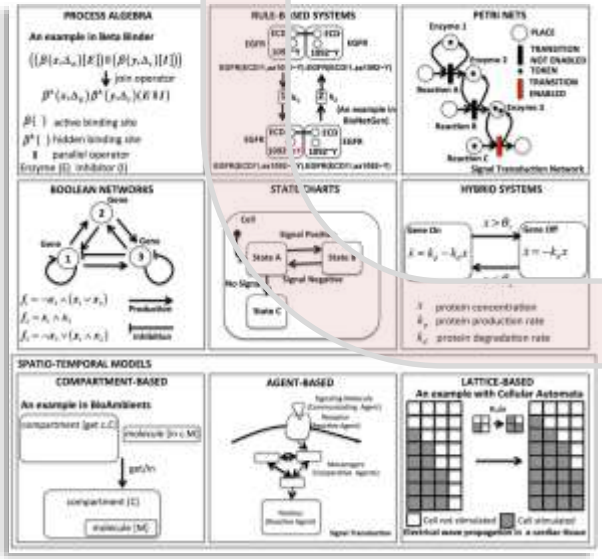
Bio-ontologies are useful tools to formalize biological knowledge representation…



- Annotations are gene-centered
  - genes and gene products have the same annotations
  - independent of the state of a molecule
- Annotations are "implicit" information

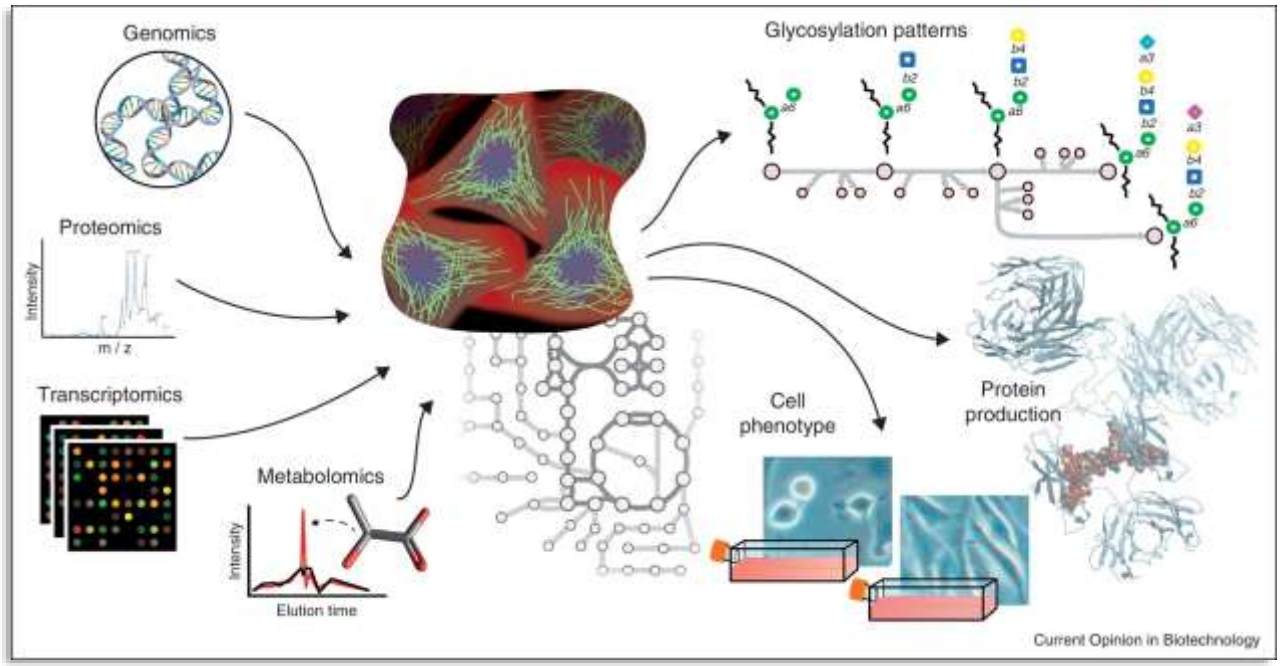
Systems Biology is a suitable framework to integrate heterogeneous entities at different scales…





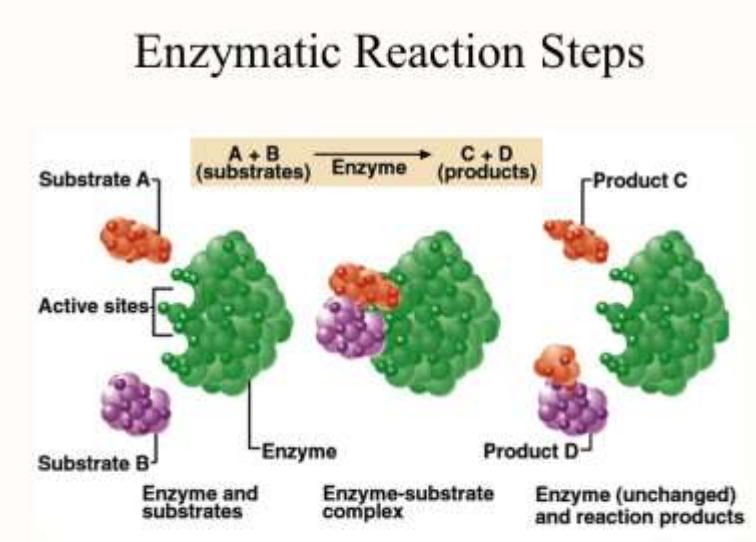
Mathematical models

Unformal biological data integration



… but even if a mathematical model is a formal object, it does not really manage the knowledge

Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge

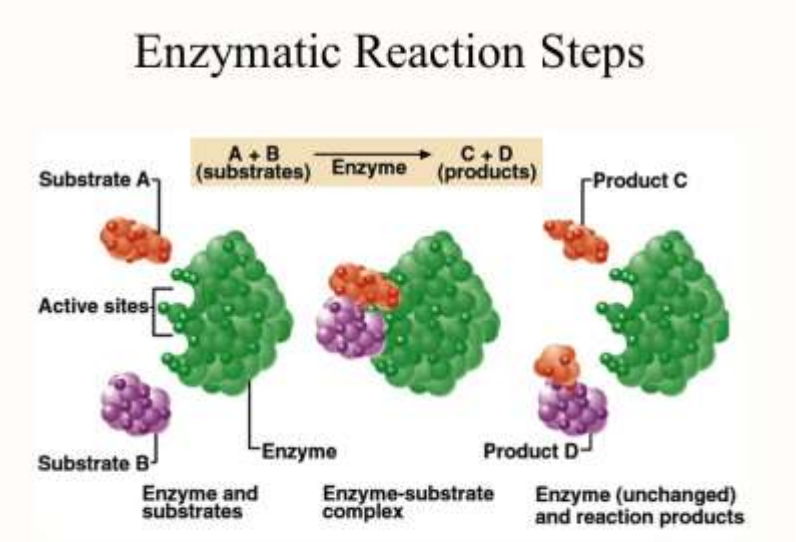


Biological representation of enzymatic reaction

$$V_0 = \frac{V_{\max}[S]}{(K_M + [S])}$$

Mathematical representation of enzymatic reaction

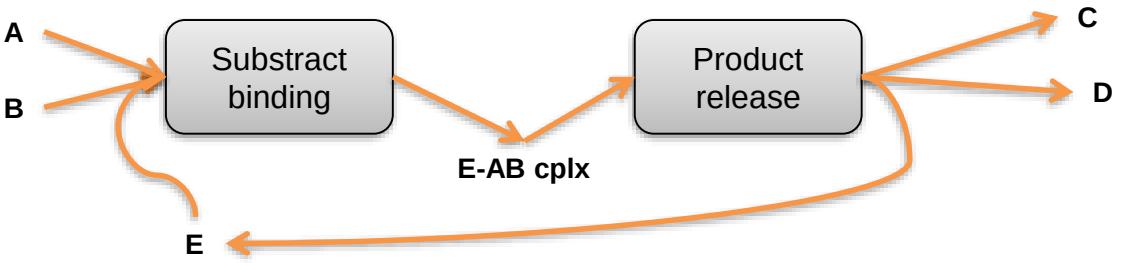
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge



Biological representation of enzymatic reaction

$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

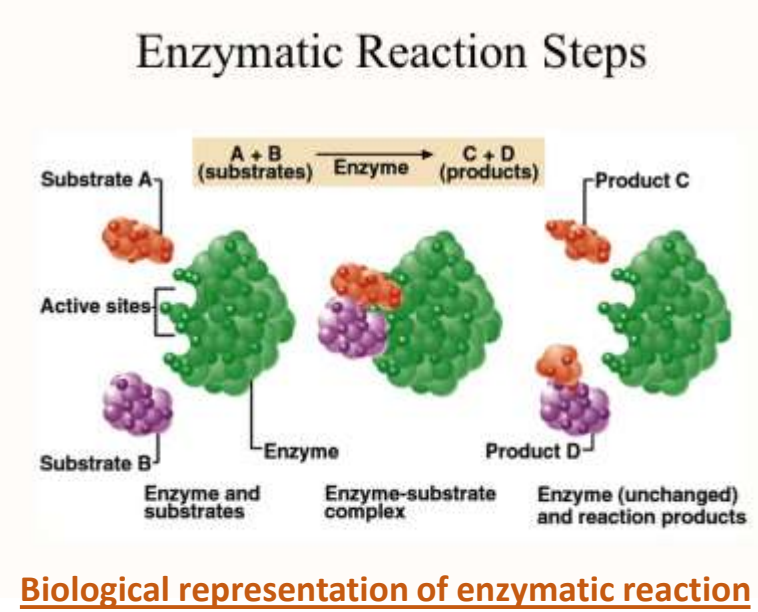
Mathematical representation of enzymatic reaction



Systemic representation of the enzymatic reaction

has_input/output
has_subprocess
has_parameter

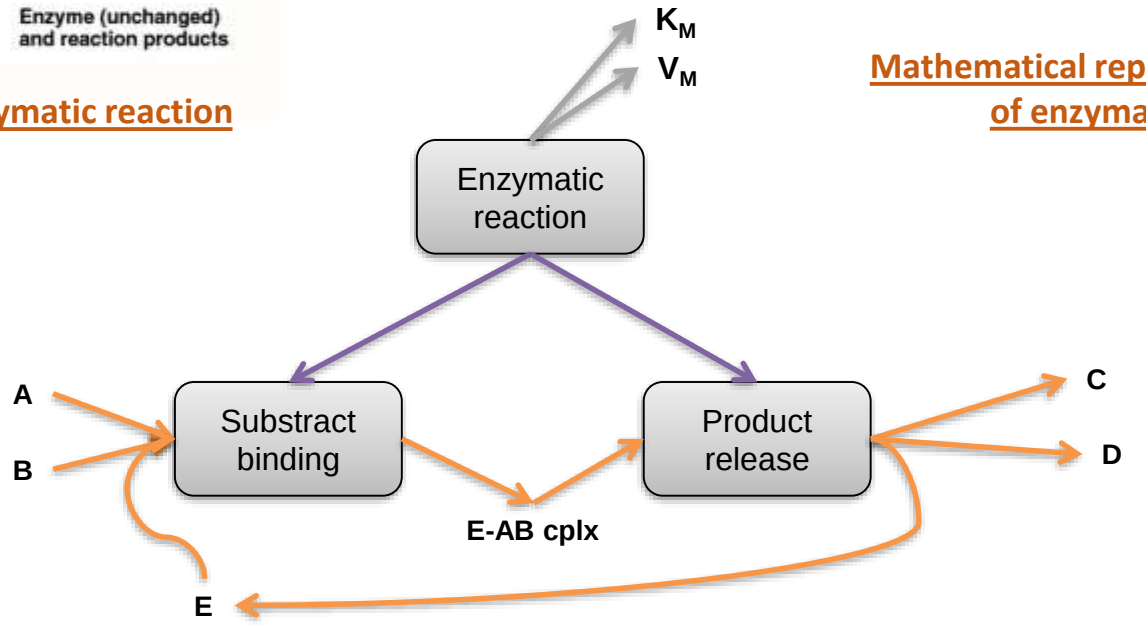
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge



Biological representation of enzymatic reaction

$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

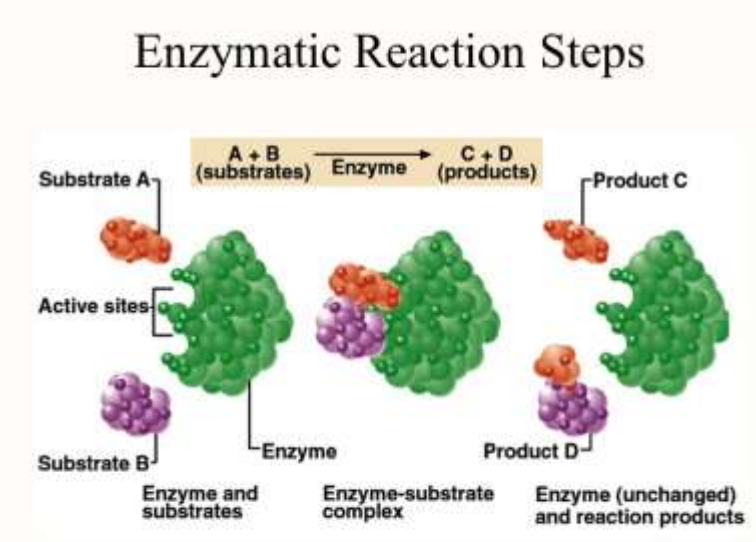
Mathematical representation of enzymatic reaction



Multi-scale systemic representation of the enzymatic reaction

has_input/output
has_subprocess
has_parameter

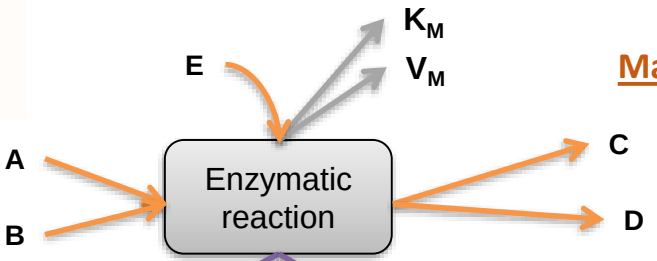
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge

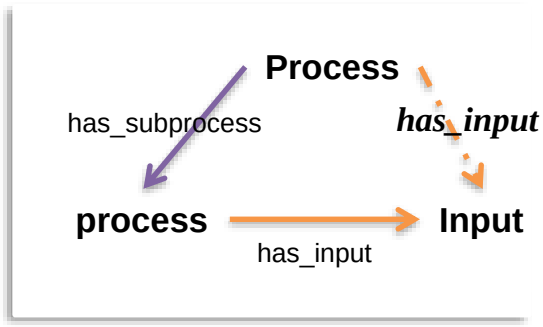


Biological representation of enzymatic reaction

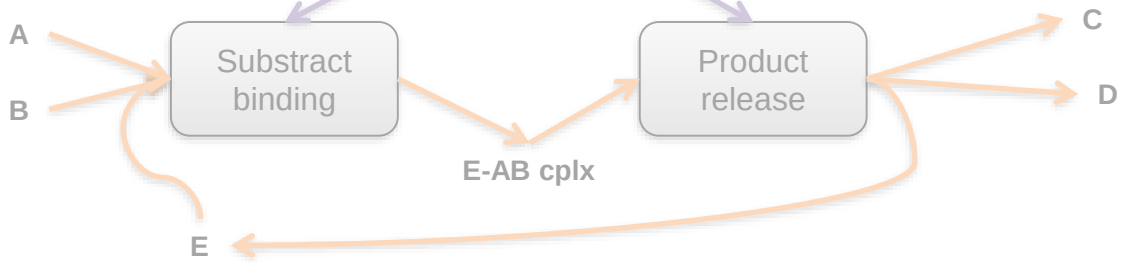
$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

Mathematical representation of enzymatic reaction





Logical rules



Automatic aggregation of the multi-scale systemic representation of the enzymatic reaction

has_input/output/mediator

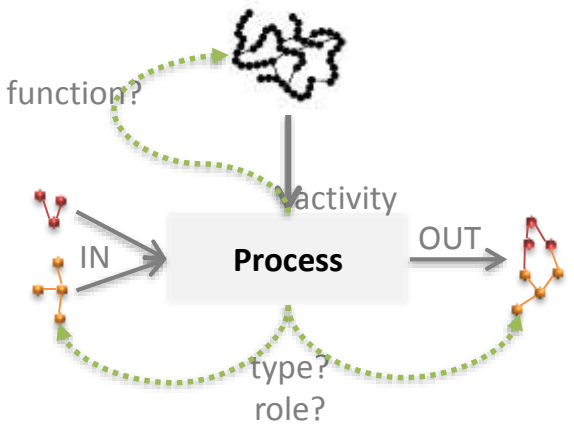
has_subprocess

has_parameter

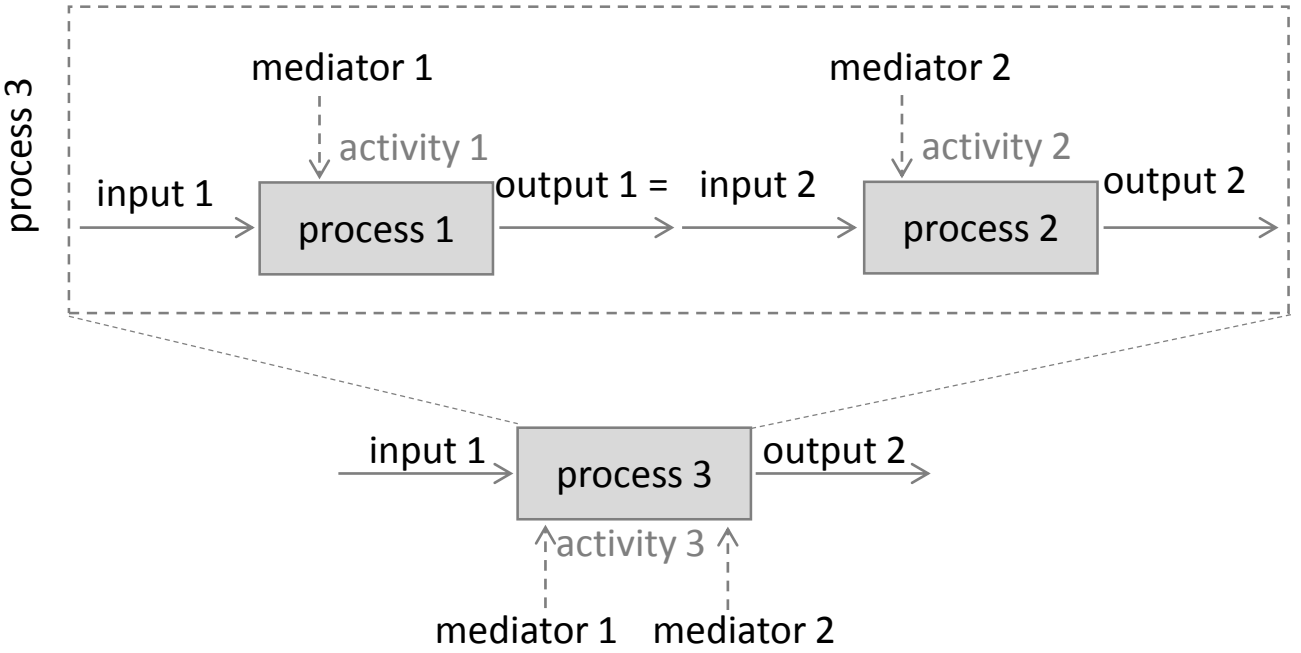
The main hypothesis of the approach

- In systemic approach, the representation is process-centered
- The information are supported by the process
- The molecule properties are conditioned by the biological process to which the molecule belongs

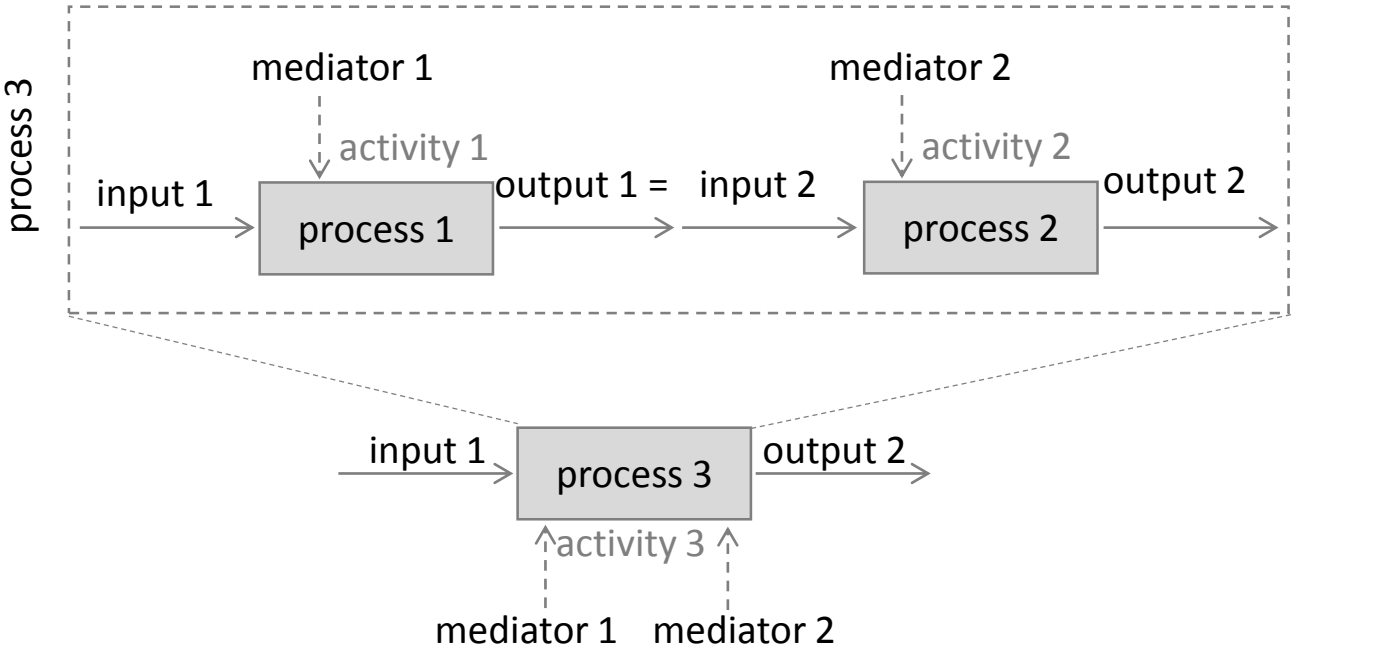
→ A fine description of biological processes as an instances should automatically conferred properties to its participants

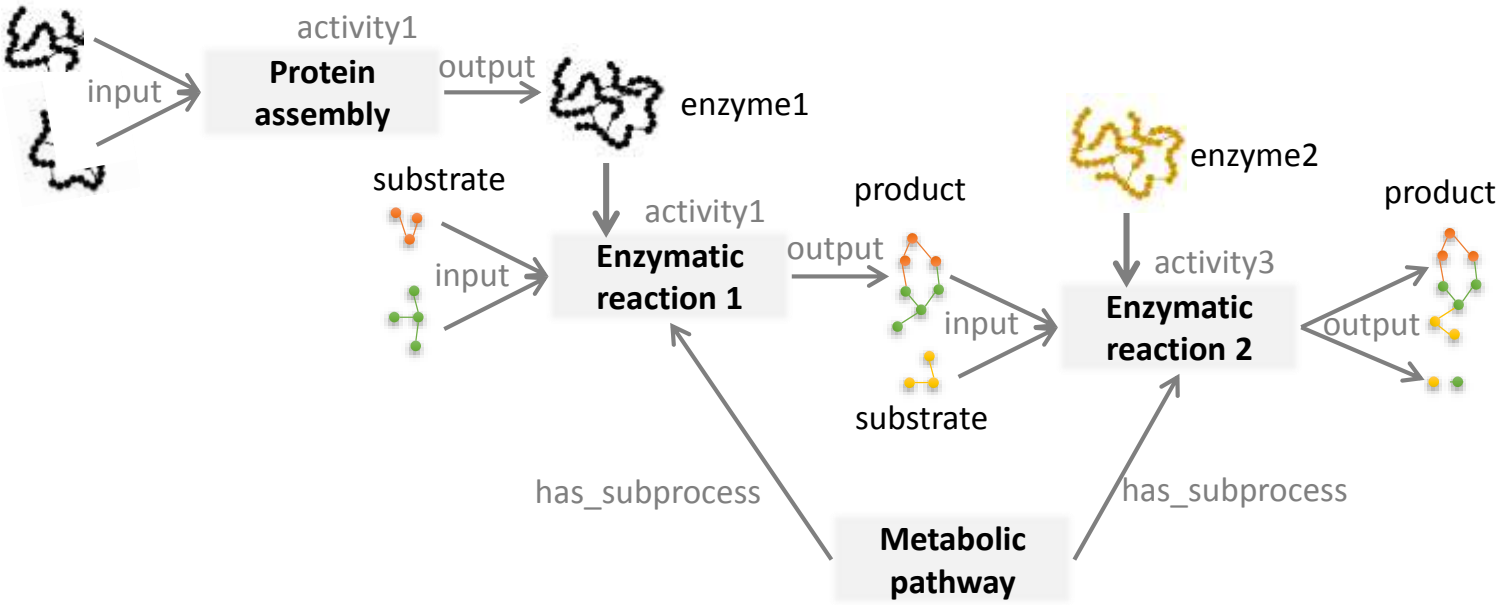


Systemic approach: a process-centered representation of systems



System biology: a process-centered representation of biology



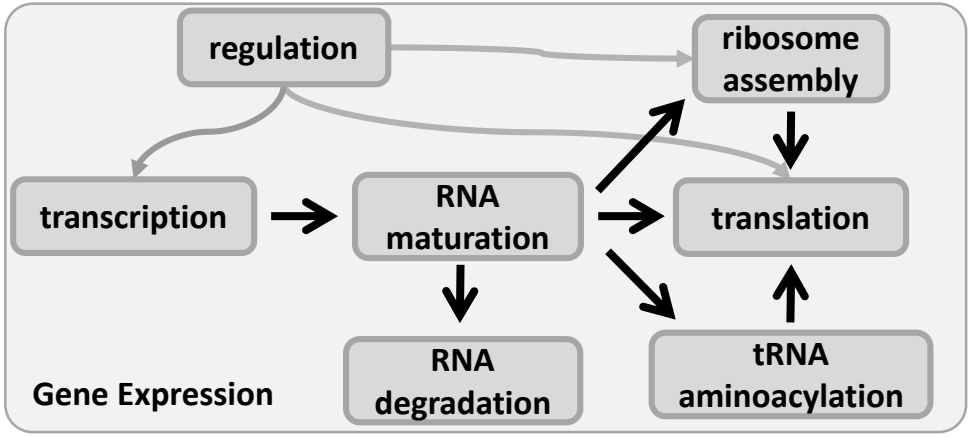


System biology: a process-centered representation of biology

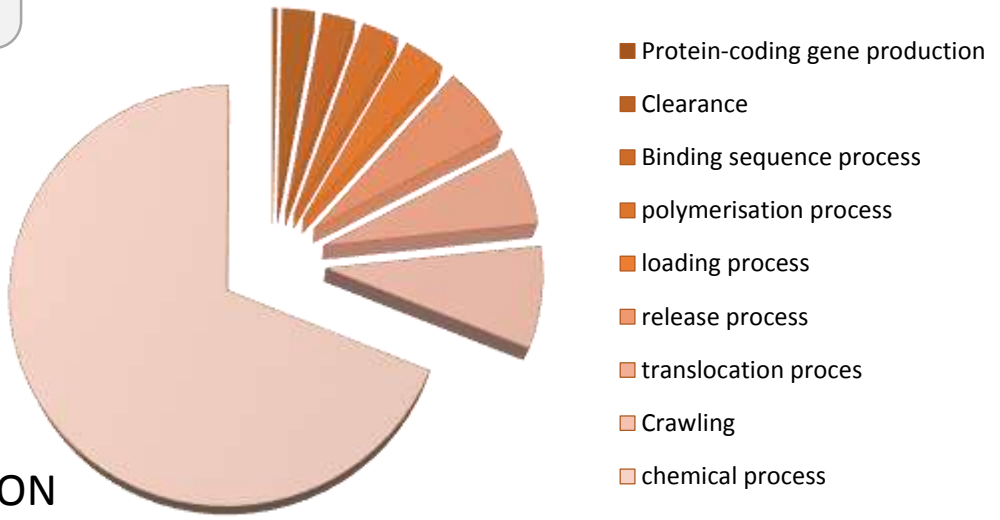
Bacterial interlocked Process ONtology (BiPON)

bioBiPON

> 300 biological processes and subprocesses with representative singletons as instances



SWRL rules &
automatic reasoning



modelBiPON

→ 9 abstract processes defined by mathematical expression

- **Complex and heterogeneous biological knowledge at the molecular scale**
  - **could be described using a systemic representation**
  - **could be automatically reclassify under a few more abstract processes and gain new properties**

System biology: a process-centered representation of biology

Bacterial interlocked Process ONtology (BiPON)

bioBiPON

Henry et al. Journal of Biomedical Semantics (2017) 8:53
DOI 10.1186/s13326-017-0165-6

Journal of Biomedical Semantics

RESEARCH Open Access

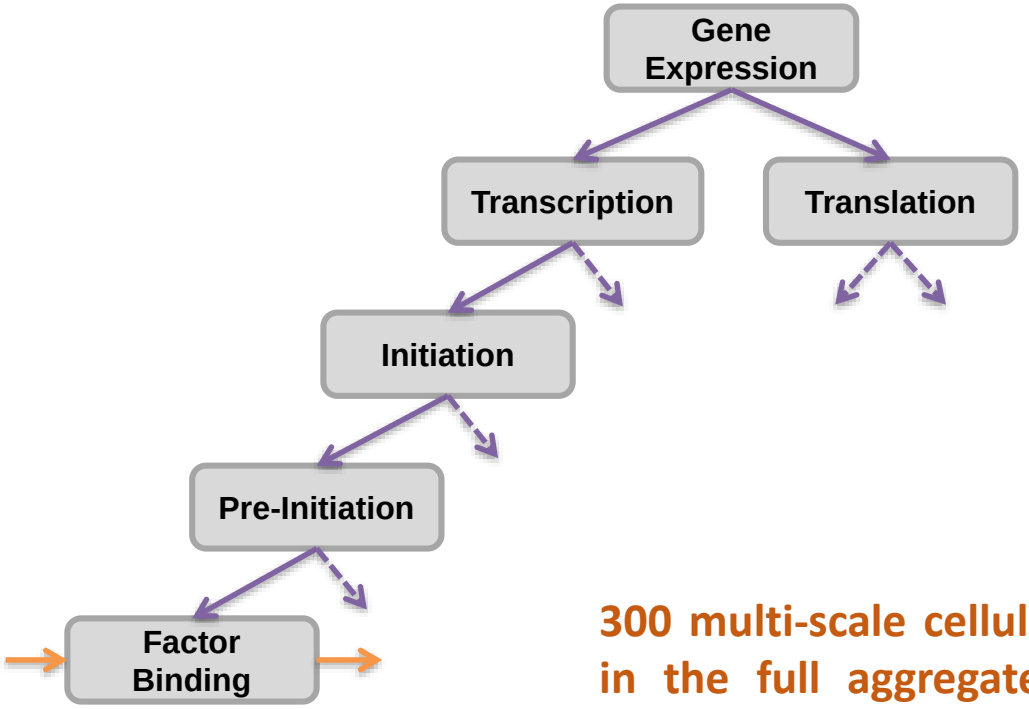
The bacterial interlocked process ONtology (BiPON): a systemic multi-scale unified representation of biological processes in prokaryotes

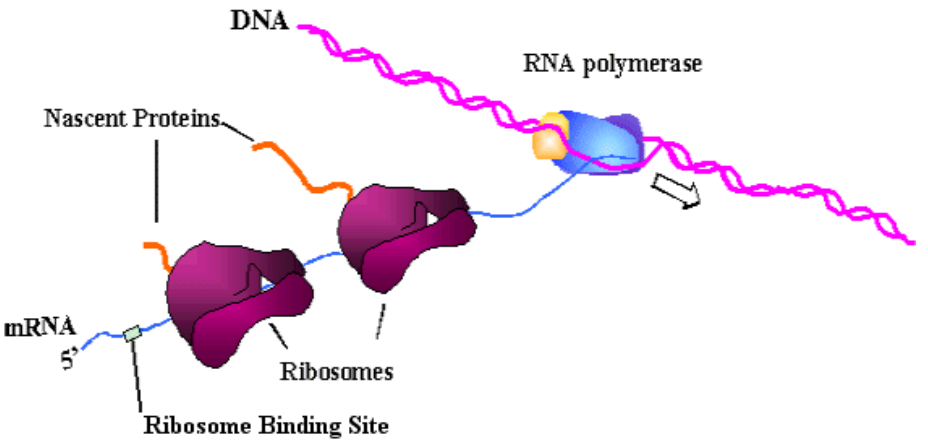
Vincent J. Henry [1,2*†], Anne Goelzer [2*†], Arnaud Ferré [1], Stephan Fischer [2], Marc Dinh [2], Valentin Loux [2], Christine Froidevaux [1] and Vincent Fromion [2]

→ 9 abstract processes defined by mathematical expression

- **Complex and heterogeneous biological knowledge at the molecular scale**
  - **could be described using a systemic representation**
  - **could be automatically reclassify under a few more abstract processes and gain new properties**

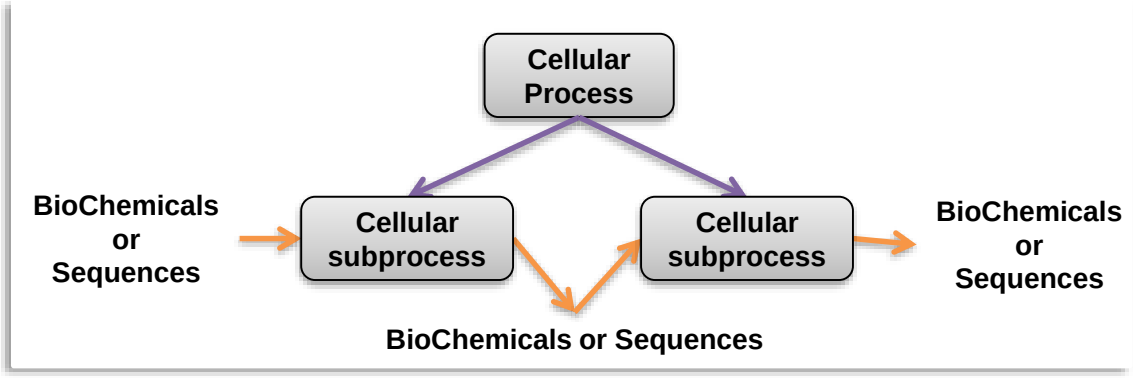
Modeling heterogeneous and multi-scale processes of bacterial gene expression

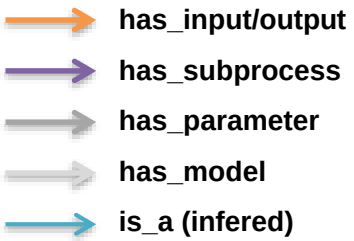




300 multi-scale cellular processes involve in the full aggregated Gene Expression process and its regulation processes

Biological knowledge representation

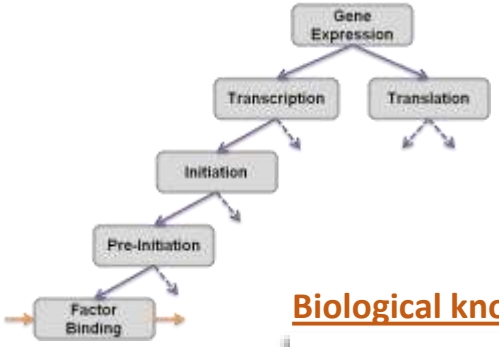




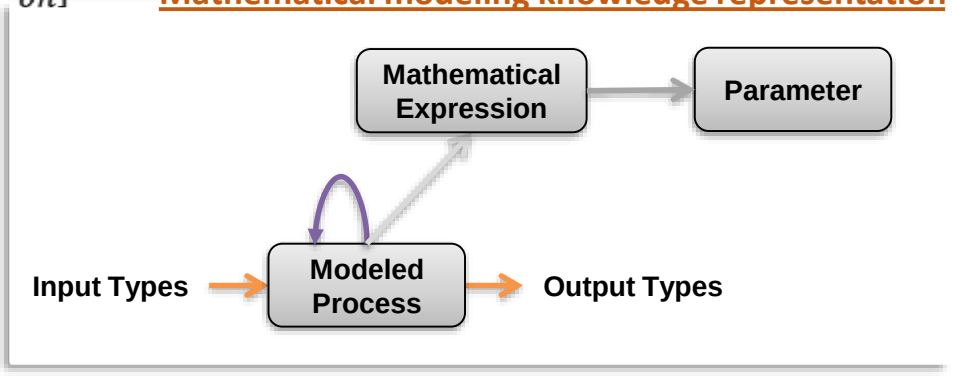
Modeling heterogeneous and multi-scale processes of bacterial gene expression

9 multi-scale Modeling Processes related to their mathematical expressions and parameters

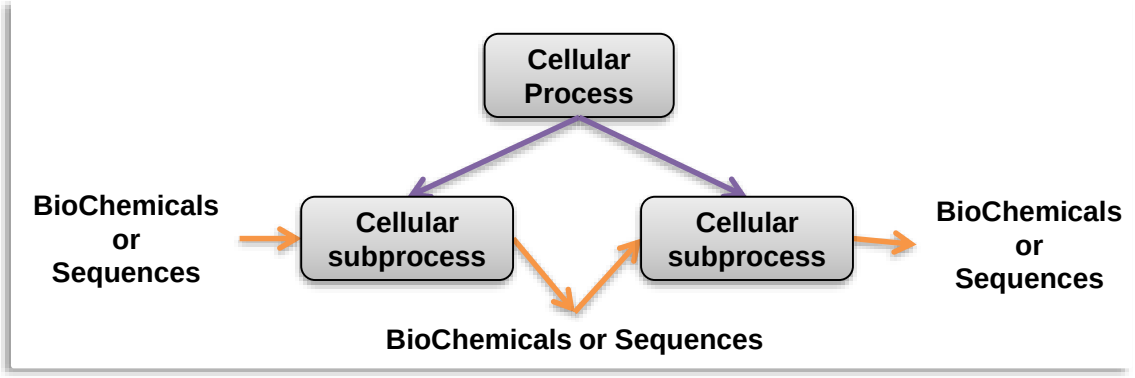
$$\frac{k_{1i}[P_{af}]}{k_{2i}+[P_{af}]} \times \frac{k_{1r}}{k_{1r}+[TF_{on}]}$$

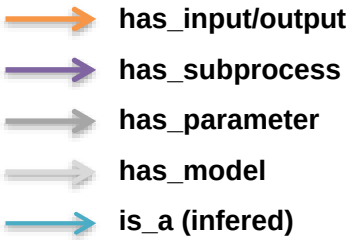


Mathematical modeling knowledge representation



Biological knowledge representation

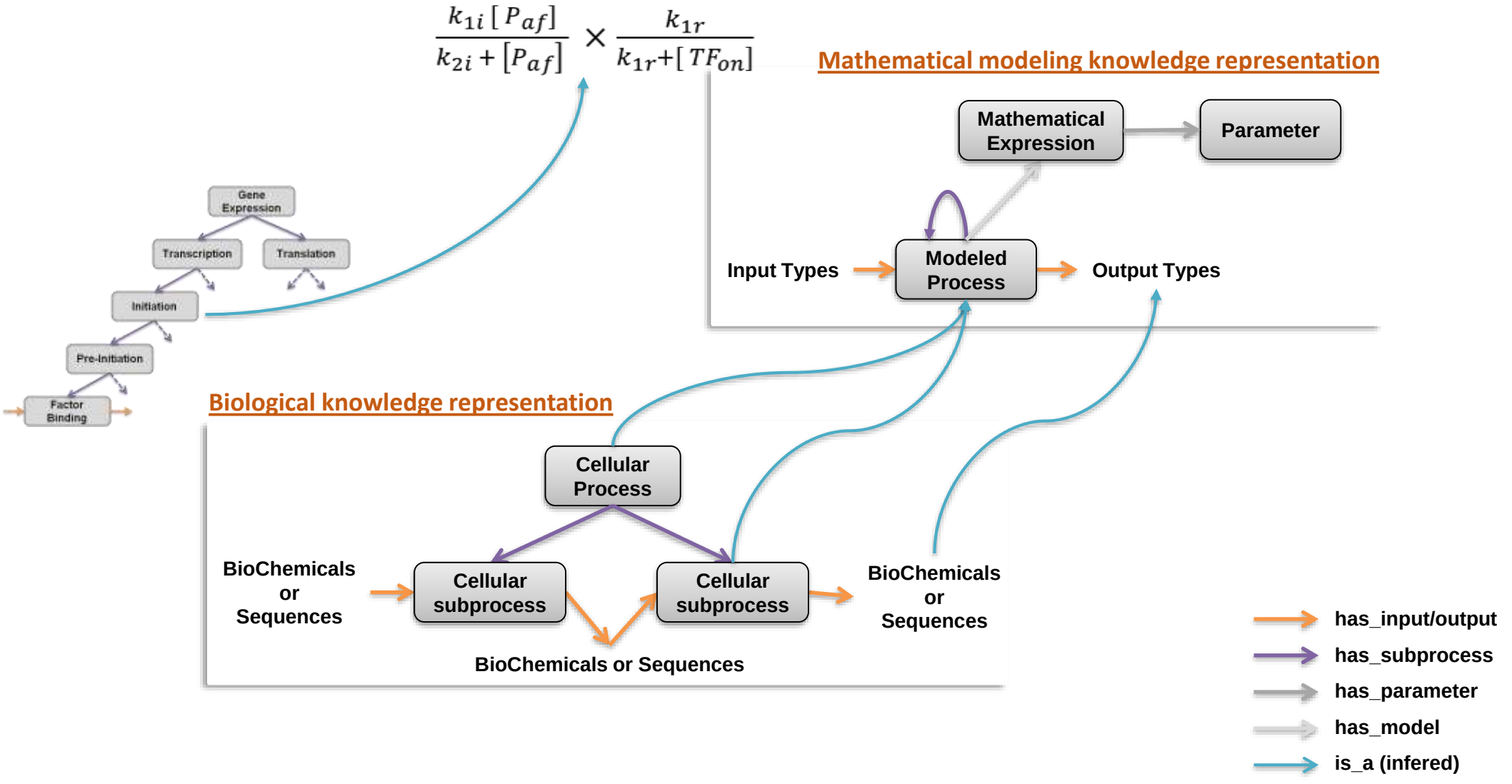




Modeling heterogeneous and multi-scale processes of bacterial gene expression

Our model:

- Could describe heterogeneous using a systemic multi-scale representation with a single pattern
- Could automatically relate Biological Process to Mathematical Models



Conclusion

Just a change of point of view :

- Processes are already described (GO-BP & GO-MF)
- Some are in relationship with chemical (GO-plus / LEGO)
- Public databases contain annotated data

Needs a “as fine as possible” description of biological processes and molecular states. This description is based on:

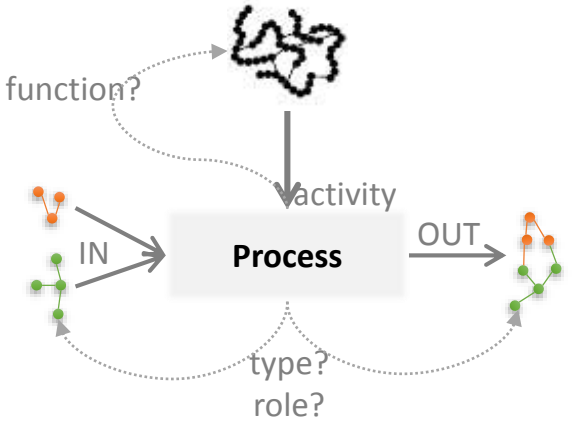
- description of different states of molecule (multimer, PTM,…)
- a systematic template (few properties define a process)
- genericity (adapted to all biochemical reactions)
- plasticity (flexibility of SWRL rules)

Thanks!

Vincent Henry, Arnaud Ferré, Anne Goelzer
Christine Froidevaux,

and also

Fatiha Sais, Elodie Marchadier, Juliette Dibie



vincent.fromion@inra.fr