La modélisation systémique de la cellule constitue-t-elle une base utile de la représentation des liens entre les entités de la cellule à travers des « graphes » ?

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

Gif sur Yvette, October 2019

Context

A lot of progress in the biological field in three 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

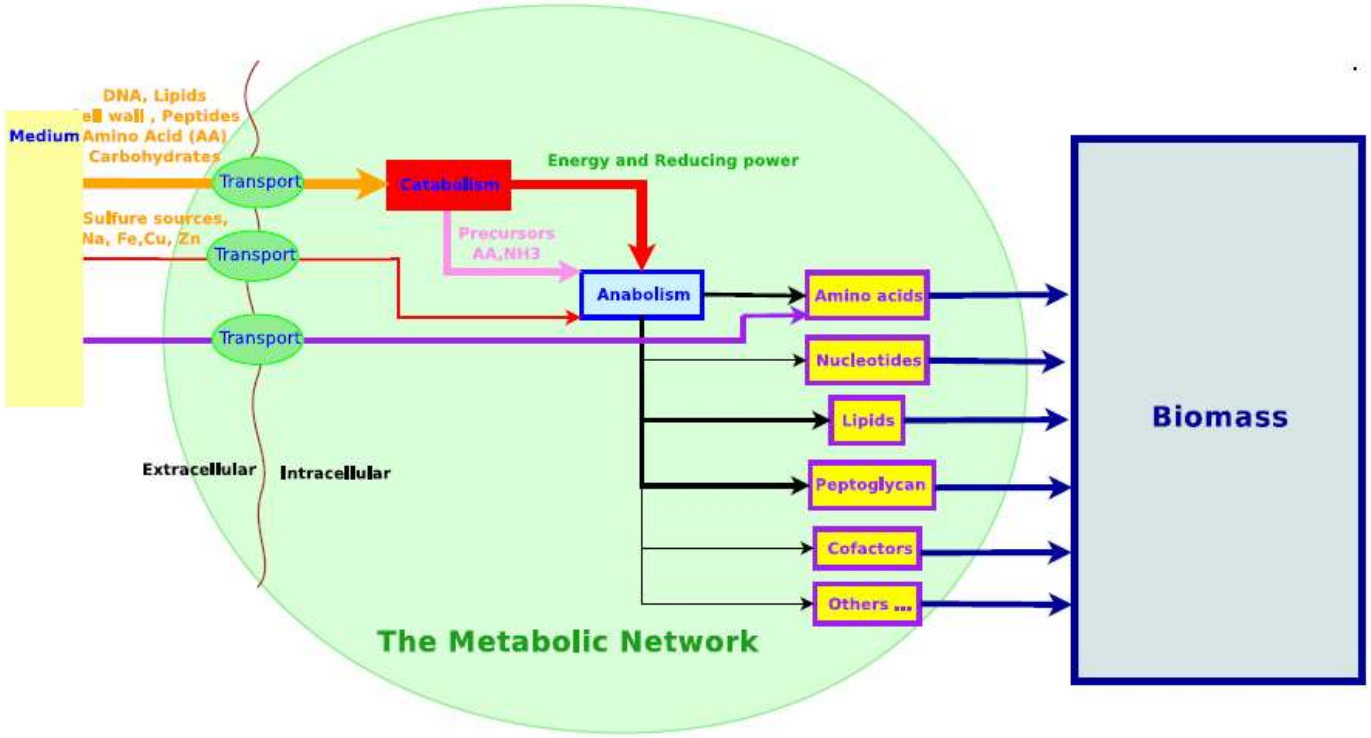
Different representations of the same object, for tackling the same issues and problems

The metabolic function

as

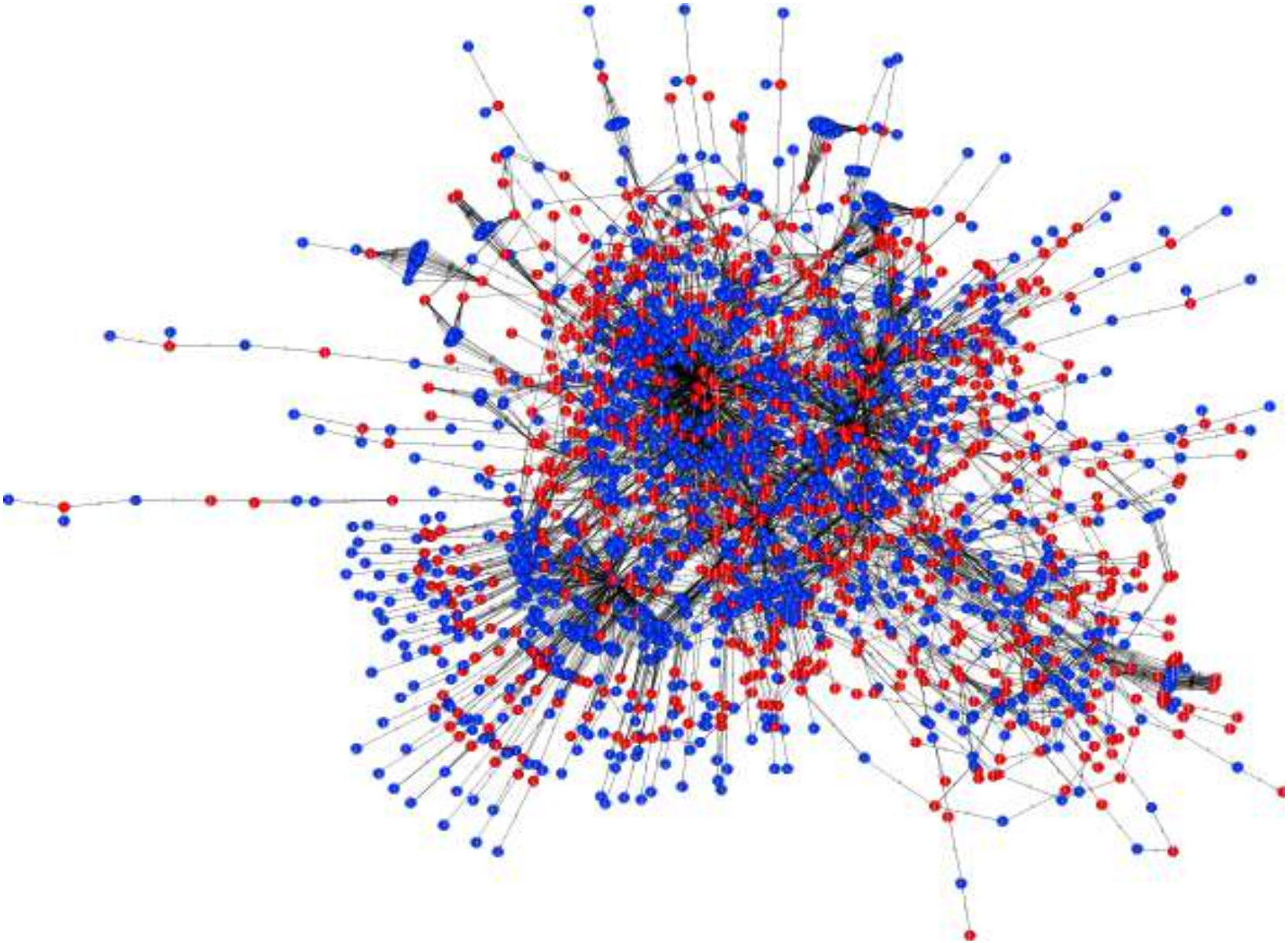
an illustrative example

The metabolic function: a quick presentation

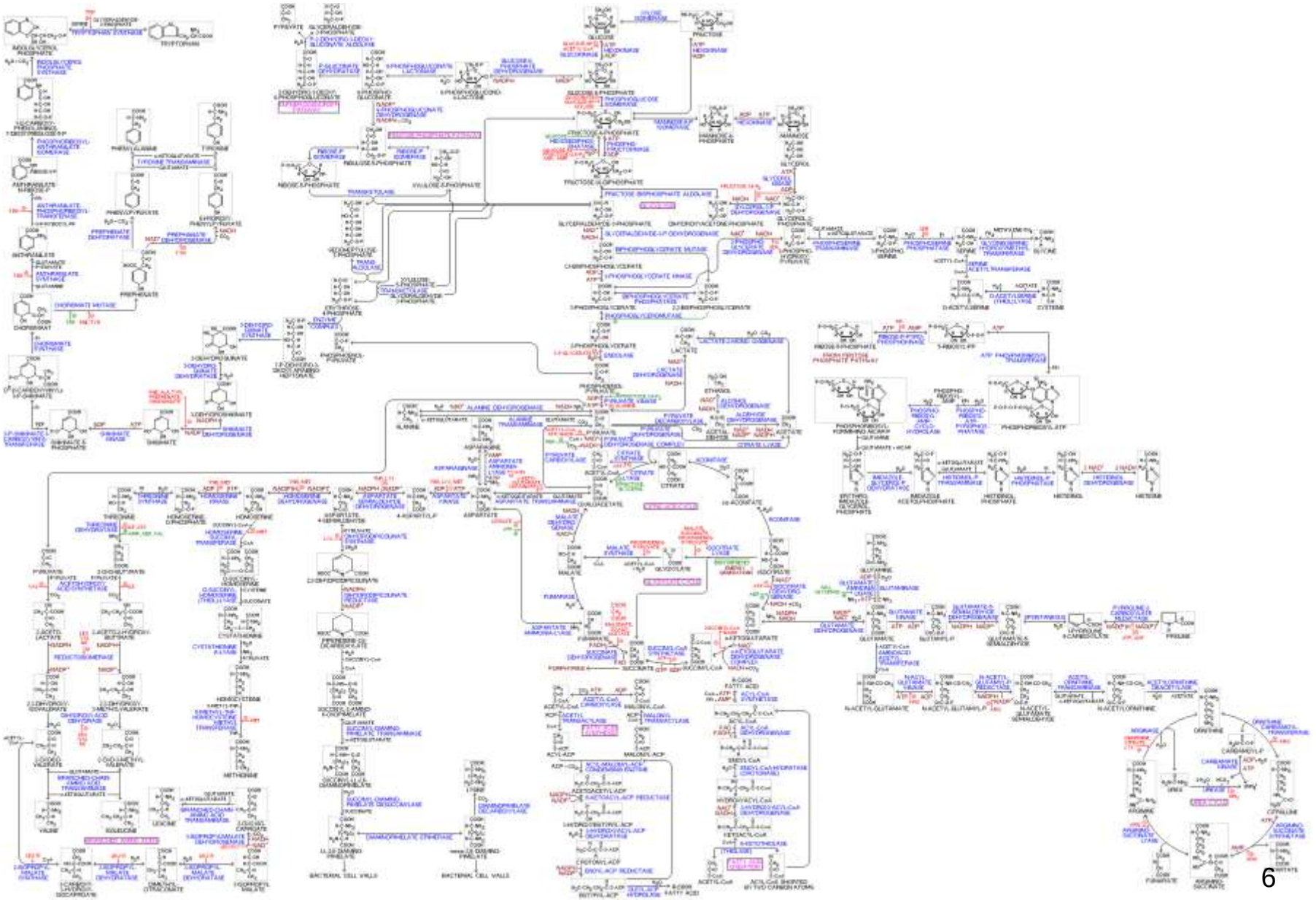


➡ The aim of the metabolic function is to produce the flux of metabolites required for the growth and for the survival of bacteria

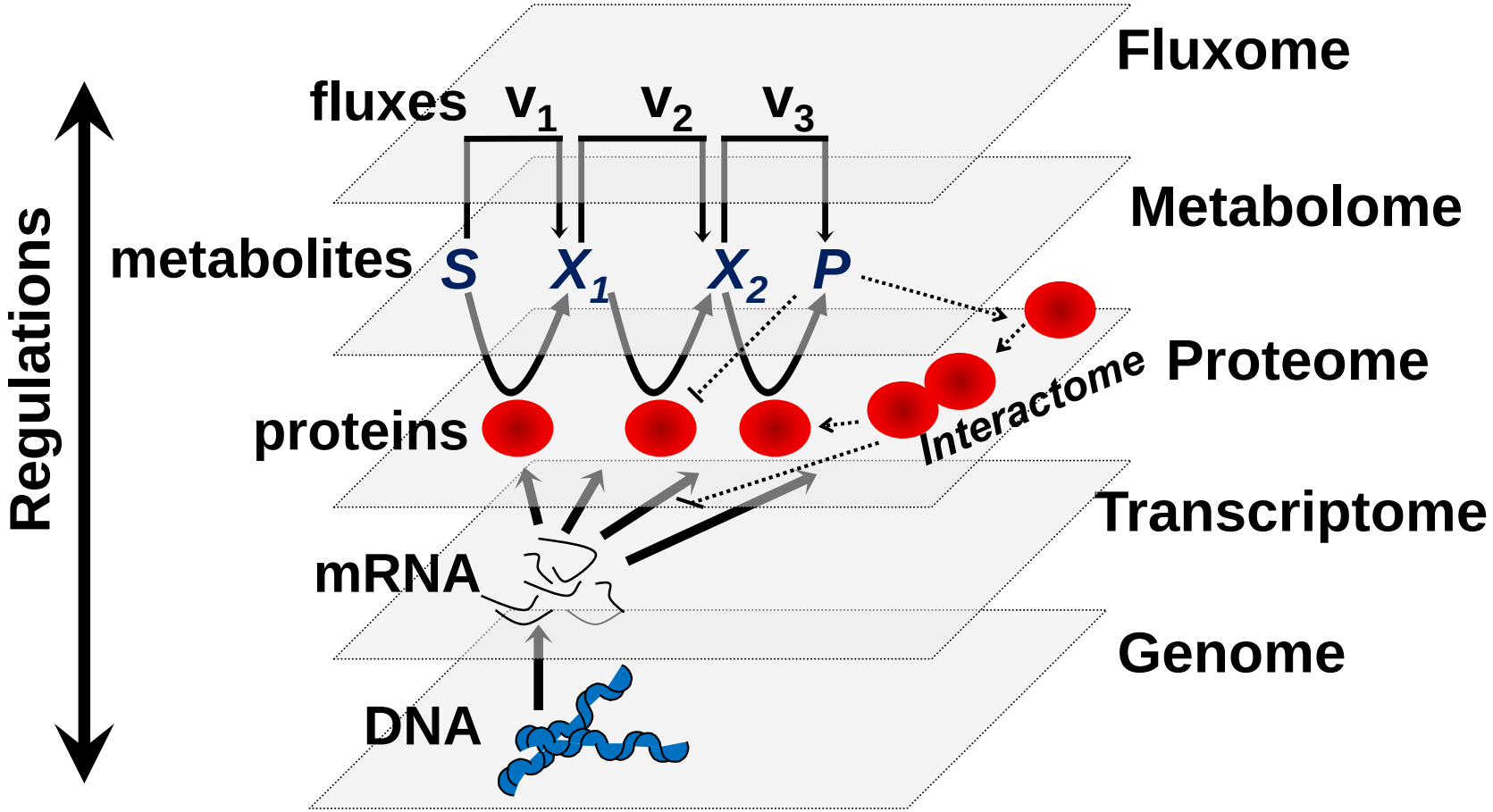
Metabolic network: reactions and metabolites



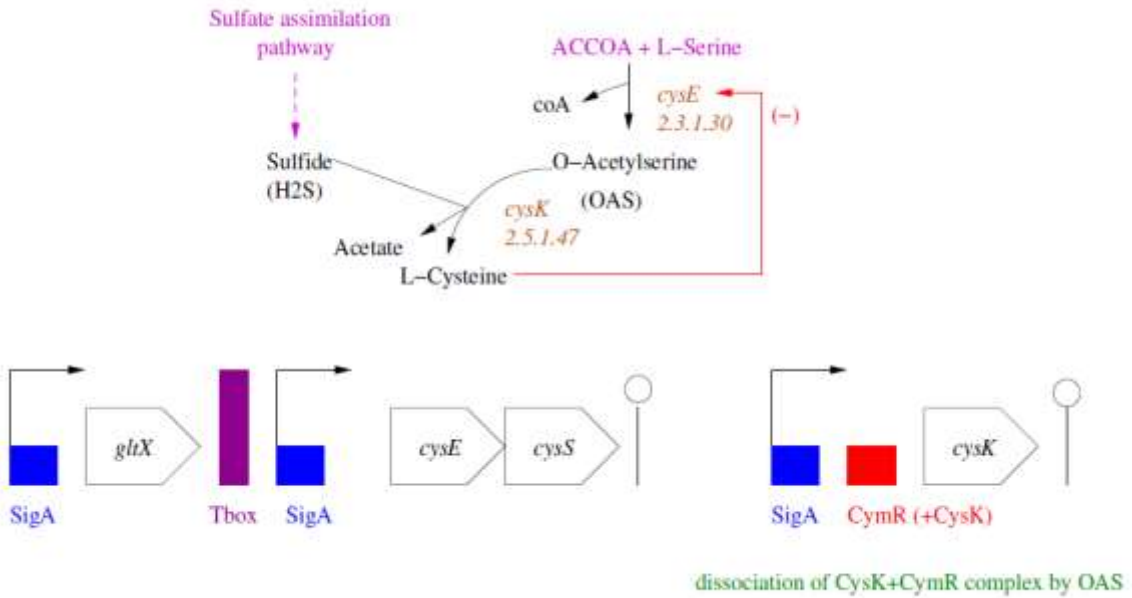
Metabolic network: reactions and metabolites



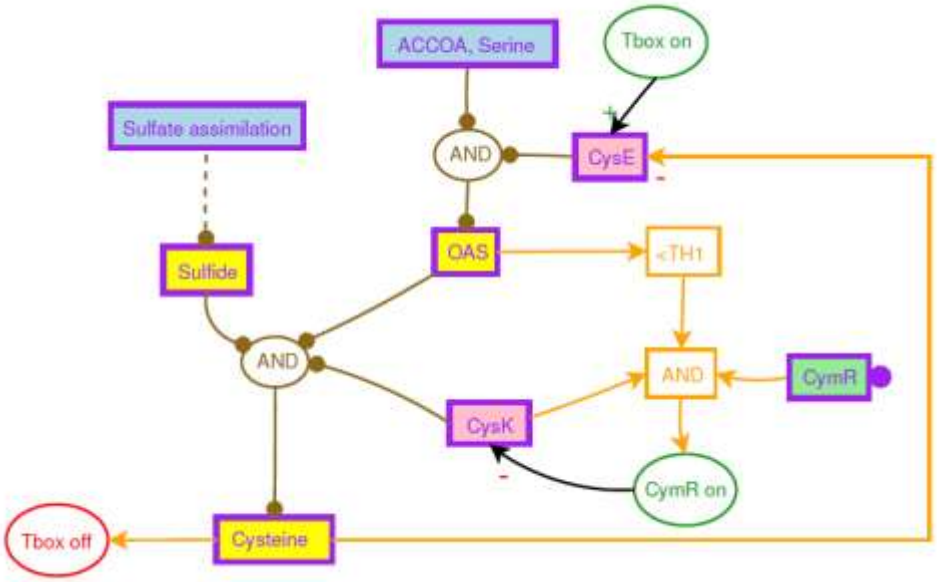
Metabolic network: different levels/time scales

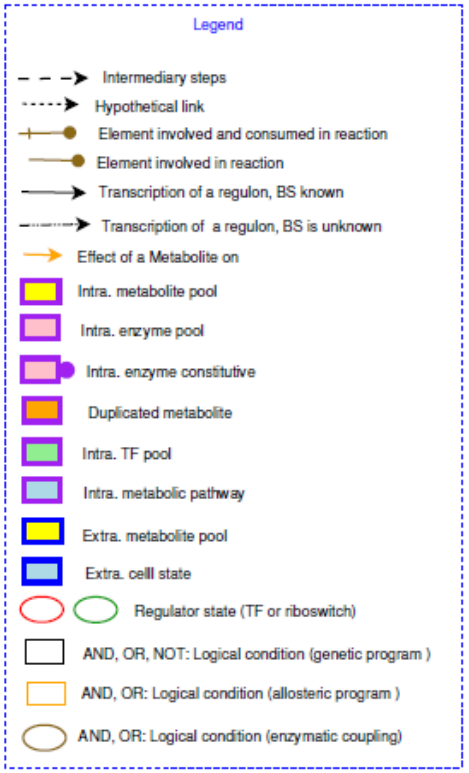


Metabolic network: reactions, metabolites and regulatory mechanisms

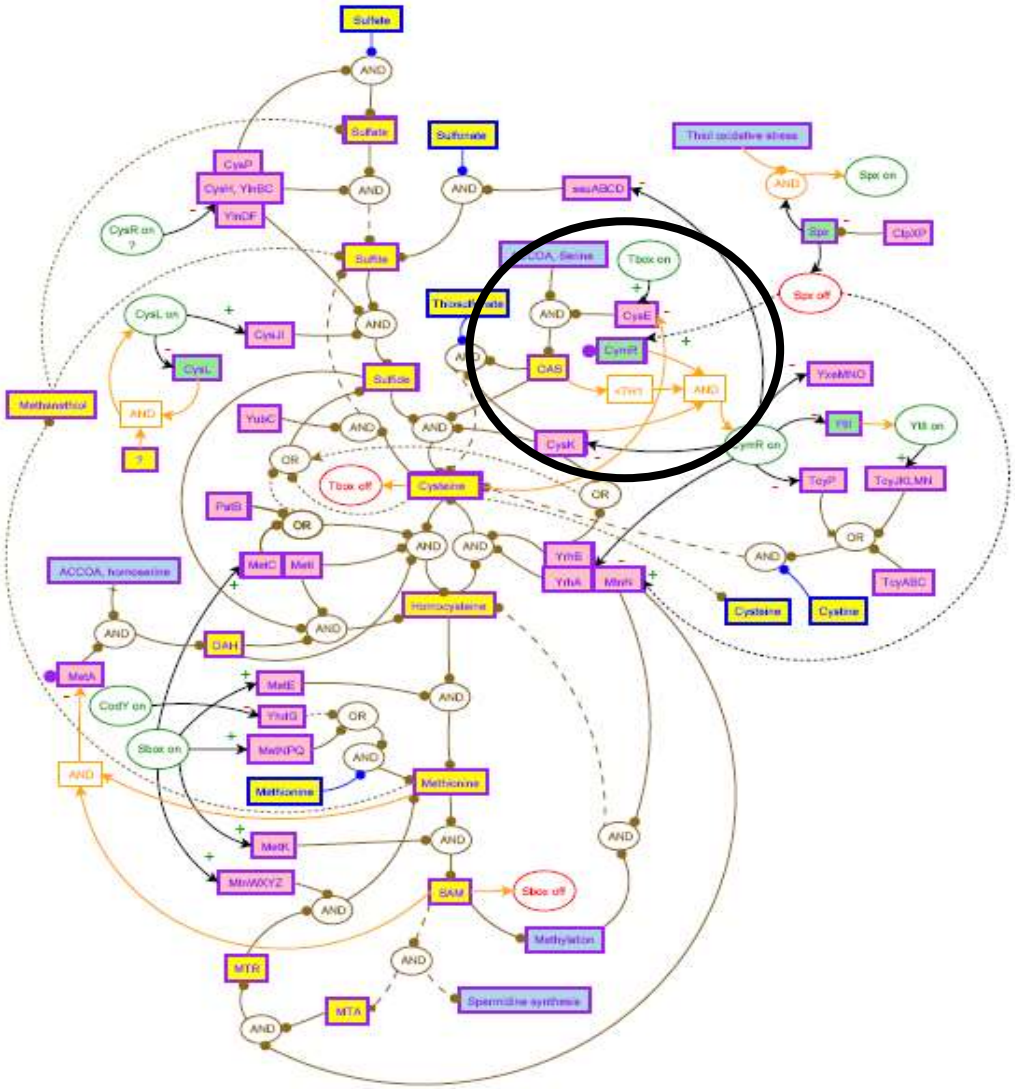


Metabolic network: reactions, metabolites and regulatory mechanisms





Metabolic network: reactions, metabolites and regulatory mechanisms



Metabolic and its regulatory network

Each metabolic pathway integrates

- the kinetic reactions and their known enzymatic regulations,
- the known transcriptional, translational and post-translational regulations and their metabolic effectors that have been experimentally validated,
- the organization of genes in operons,
- the “Boolean like conditions” of transcription and translation for each gene.

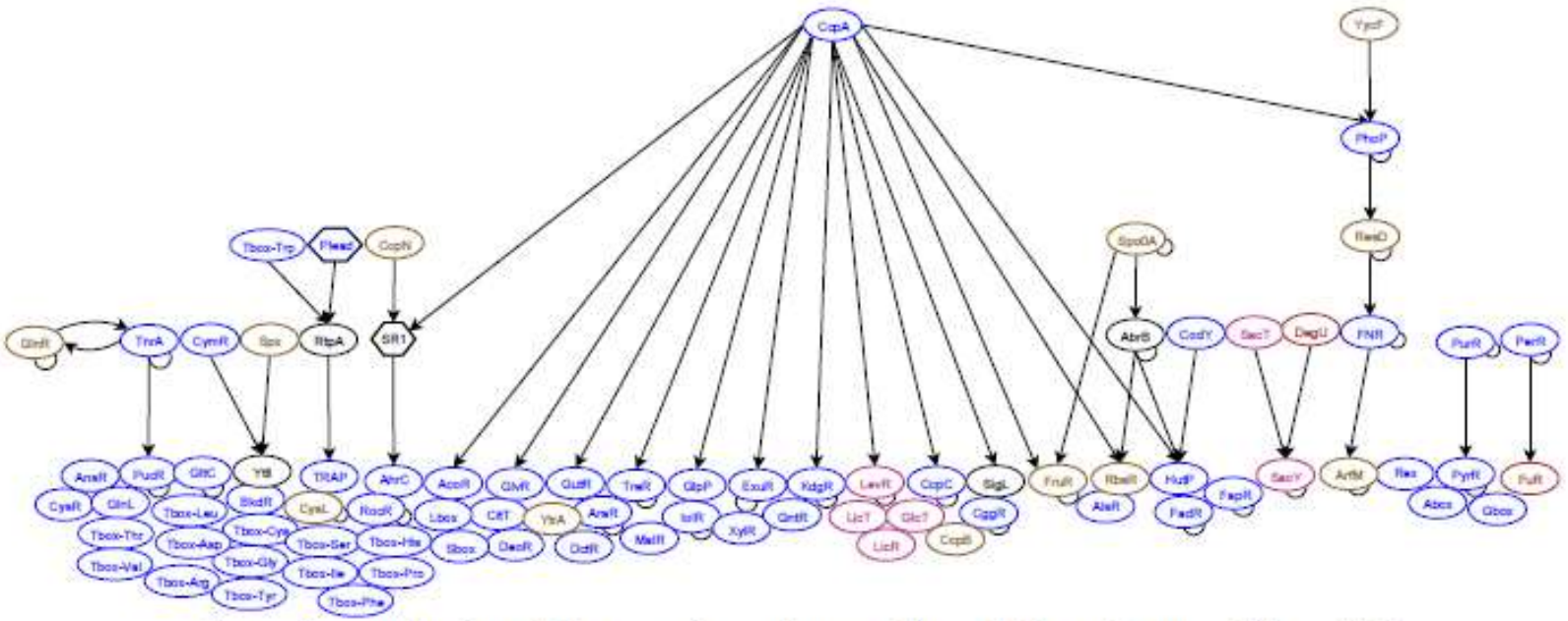
The model integrated (2008)

- 622 reactions,
- 587 genes and 67 transcription factors, 21 other genetic regulations,
- more than 400 references.

⇨ Goelzer et al., *BMC Systems Biology*, 2:20, 2008.

⇨ *The current model is now genome-scale and integrates more than 210 genetic regulations (the regulatory networks of various stresses (oxidative, heat, iron, etc.), the growth rate management loop, etc.*

Regulatory and metabolic networks are (strongly) connected



General organization of the genetic regulators of metabolic pathways of *B. subtilis*

Main points

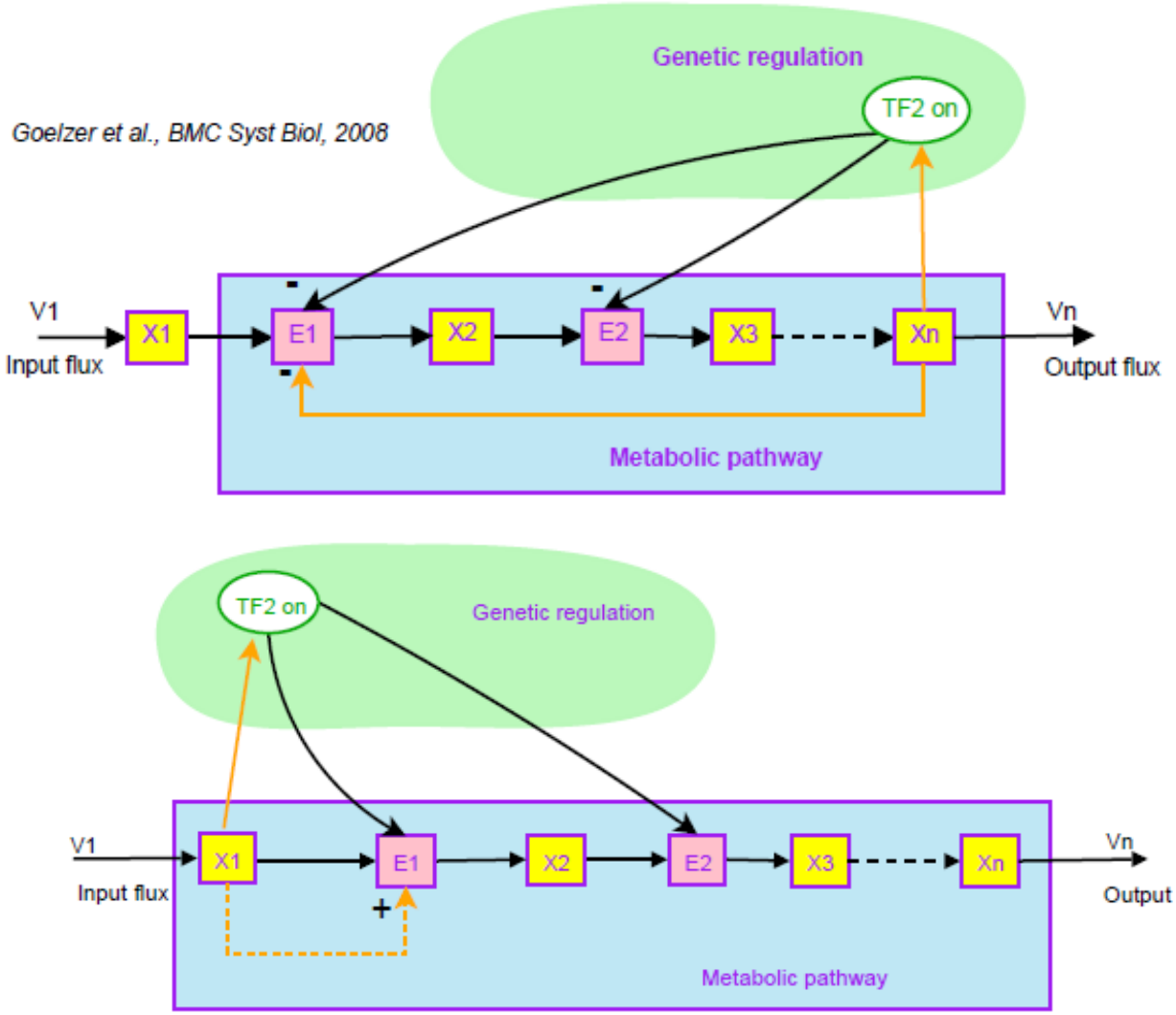
- ✓ A least 15 % of metabolites are involved in the control of 50 % of the metabolic enzymes,
- ✓ Few genetic regulations control other genetic regulations,
- ✓ Almost every genetic regulation concerns a ’simple’ metabolic pathway.

Regulatory and metabolic networks are (strongly) connected

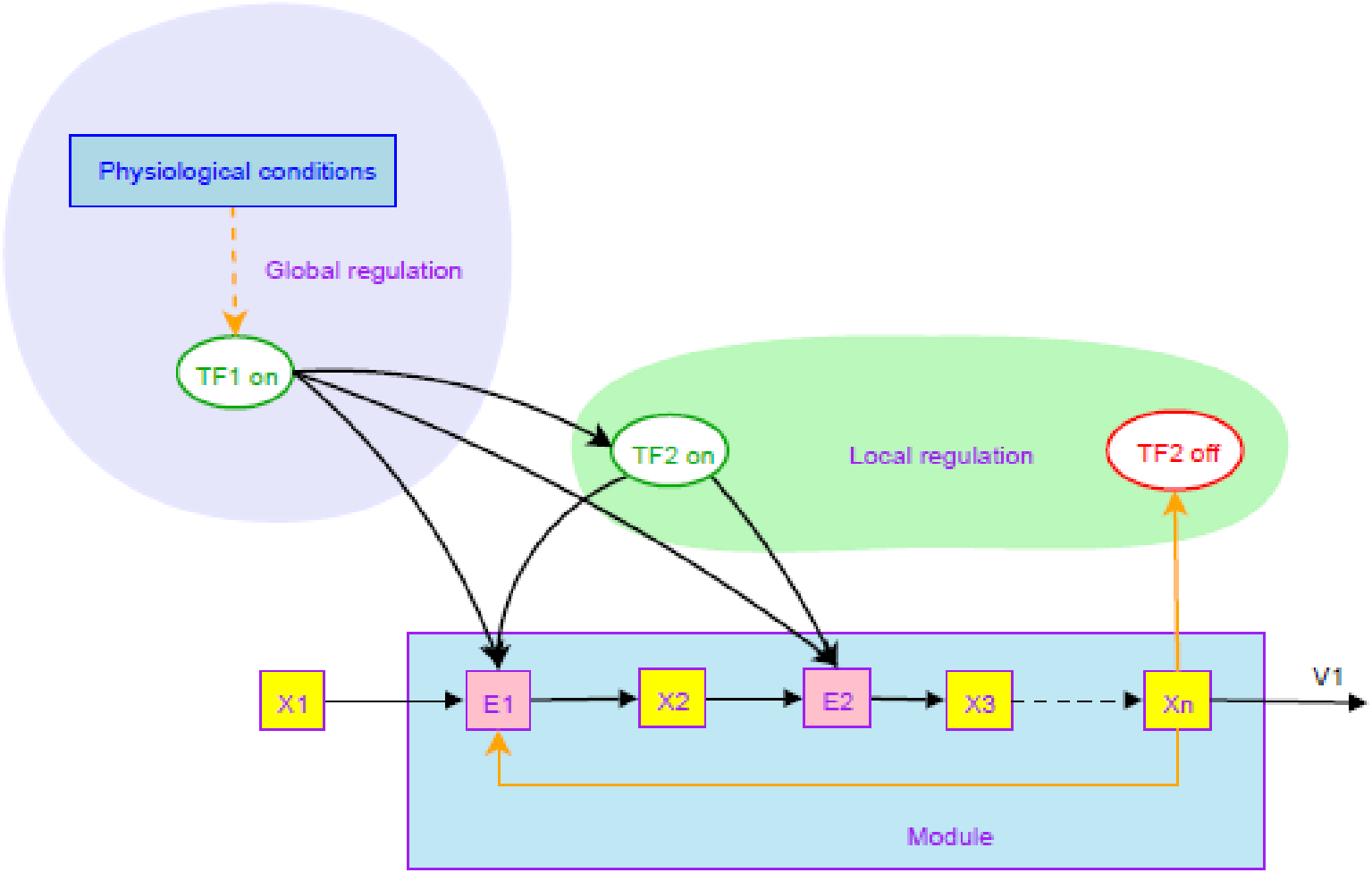
but are organized

(from a systemic point of view)

Two local regulatory motifs

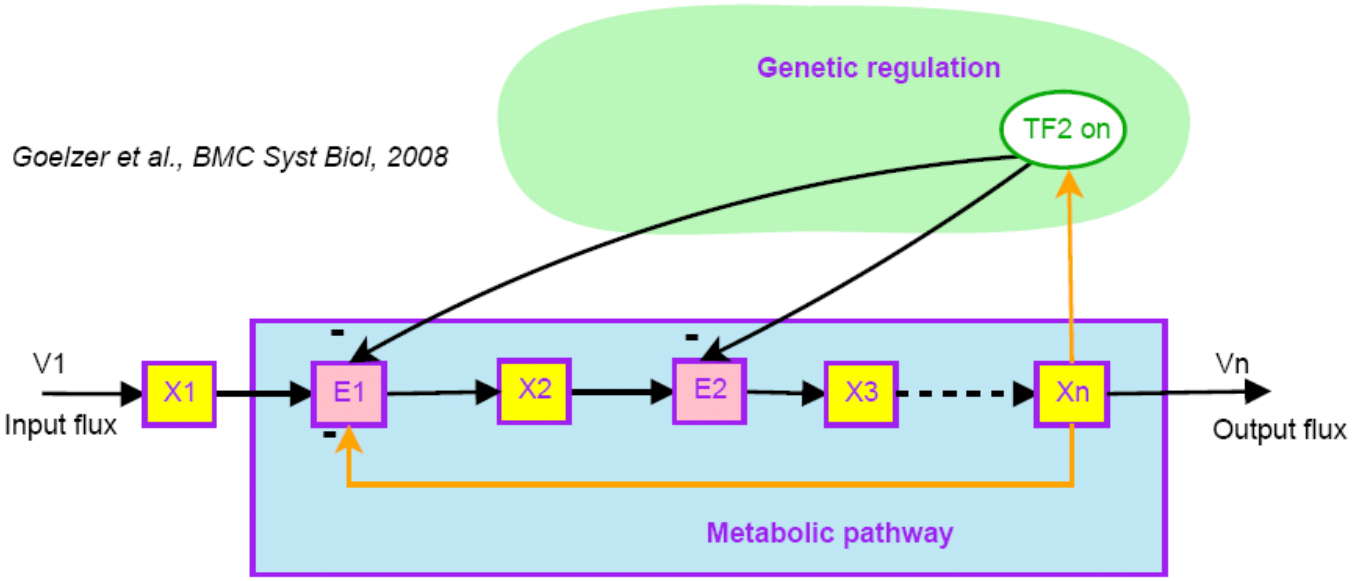


Global regulation/coordination



Local modules have very specific features
(which are easily derived …)

Properties of the end-product regulation motif



Two simple feedback loops

- regulation of enzyme activity by the end-product,
- regulation of enzyme production through a genetic regulation controlled by the end-product.

Module equilibrium during exponential growth phase

A model of the end product module is typically given by

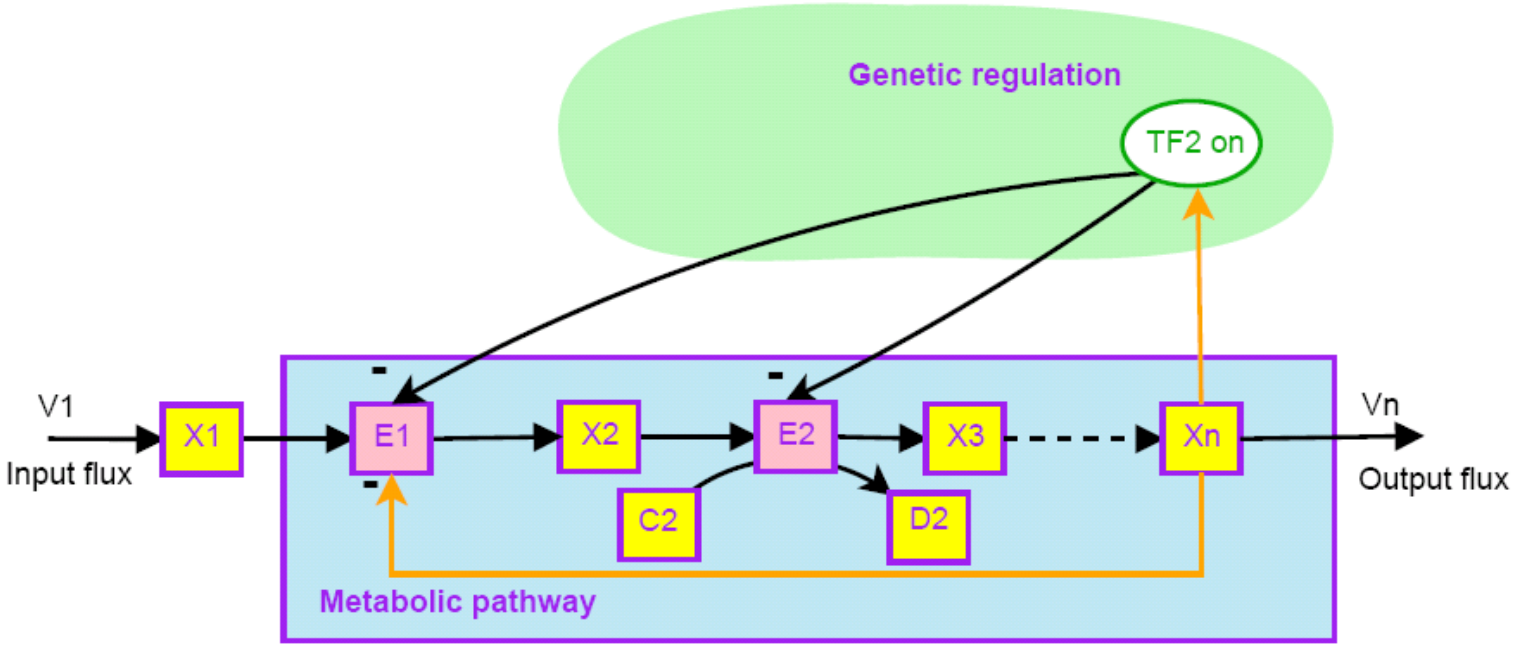
$$\begin{cases} \dot{x}_1 &= \nu_1 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 &= E_1 f_1(x_1, x_n) - E_2 f_2(x_2, x_3) \\ \vdots & \vdots \\ \dot{x}_n &= E_{n-1} f_{n-1}(x_{n-1}, x_n) - \nu_n \\ \dot{E}_1 &= g(x_n) - \mu E_1 \end{cases}$$

If we assume that intermediary enzymes do not saturate for the given input and output fluxes then the equilibrium regime of the I/O module is the unique solution of these equations:

$$\begin{cases} \overline{E}_1 = \dfrac{g(\overline{x}_n)}{\mu}, \\ \\ f_1(\overline{x}_1, \overline{x}_n) g(\overline{x}_n) = \mu \nu_n \end{cases}$$

- ✓ The end-product concentration is a decreasing function of the output flux,
- ✓ The flux through the pathway has an upper bound,
- ✓ The key assumptions are clearly the irreversibility of the first enzyme and the dilution effect.

Intermediary metabolites do not influence the I/O features



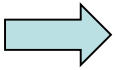
- The metabolic network is strongly connected through co-metabolites providing energy (ATP/ADP, NADPH/NADP), amino-groups (Glutamate/AKG, Glutamine/Glutamate).

- Under the condition that the first enzyme is irreversible, it is then easy to deduce that the I/O equilibrium regime does not depend on the effects of any co-metabolites necessary to produce the end-product if (and only if) they are not associated to the first enzyme and they do not lead to a saturating step.

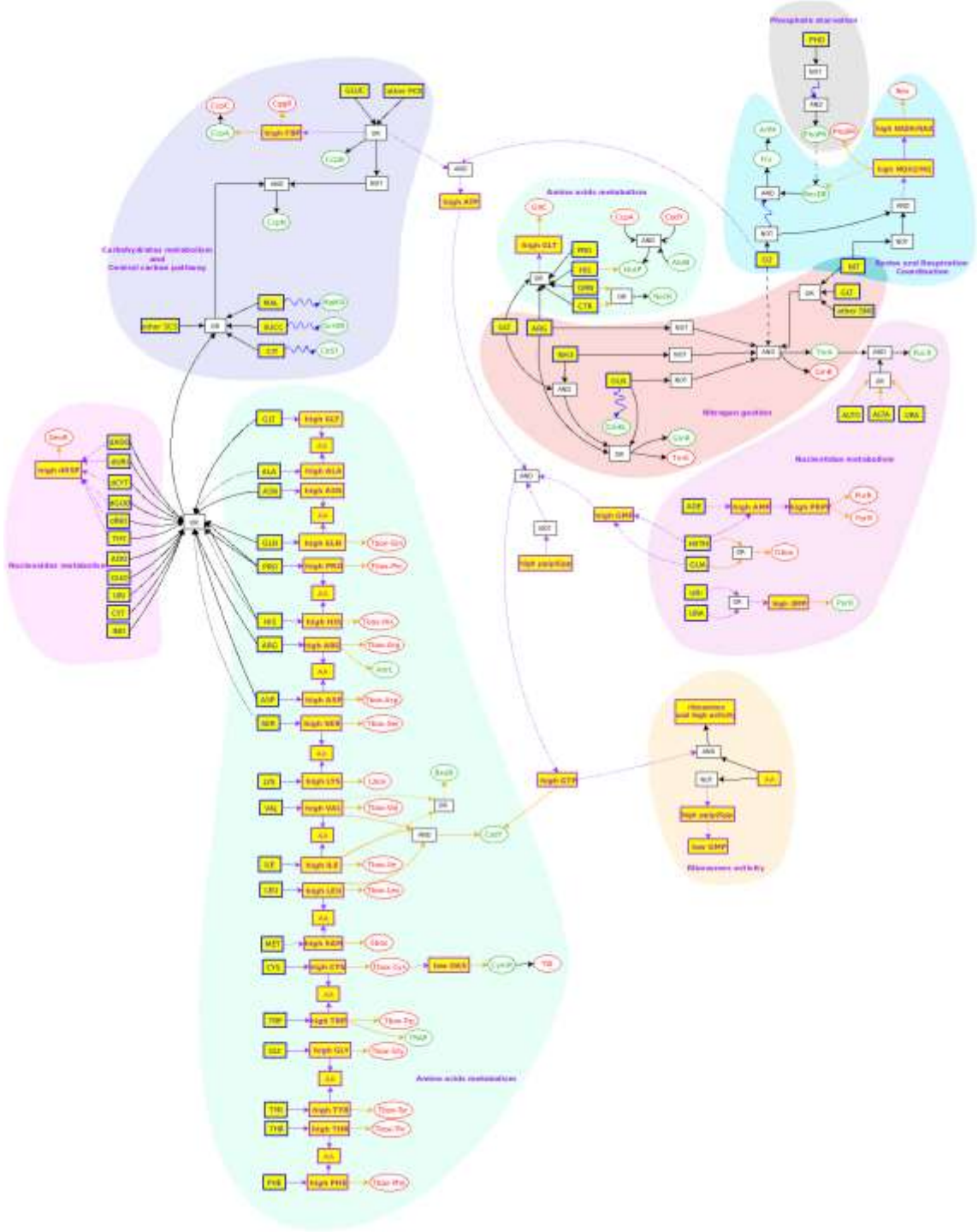
Modules have very specific (and nice) features

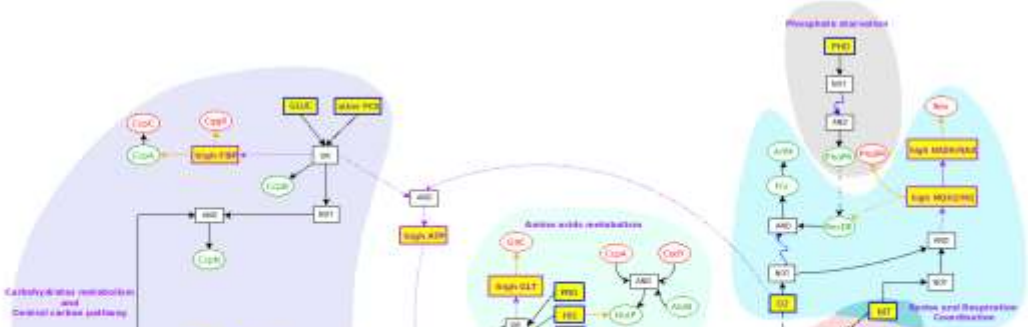
- From an analytical point of view
  - only a few reactions and metabolic pools determine the behavior of the steady-state,
  - simplification of the model analysis in the steady-state,
  - prediction of the behavior of enzyme and metabolite concentrations and of metabolic fluxes.

- From a biological point of view
  - **key role of irreversible enzymes (often the first enzyme)**,
  - **key role of the growth rate through 'enzyme dilution',**
  - the metabolic network can be broken down into elementary modules,
  - coordinated by "global" (or pleiotropic) regulators.

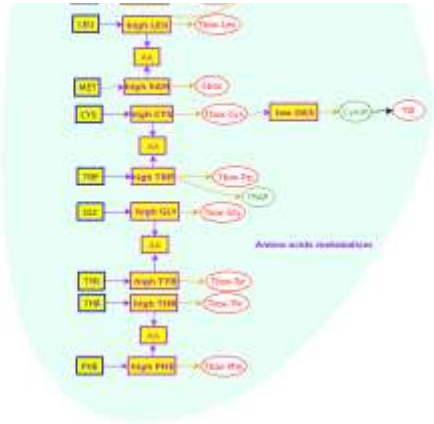


For proofs and details see e.g. Goelzer et al., *BMC Systems Biology* 2:20, 2008.





In steady-state regimes, the interactions between the various bacterial entities are then strongly reduced
(modularity)



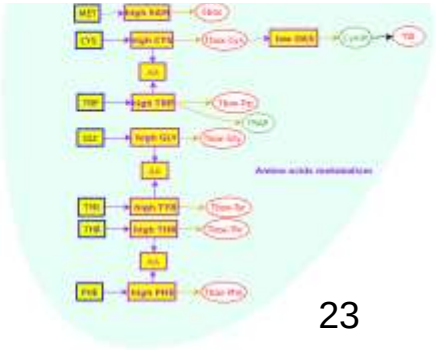
More about local modules

Towards a module ‘algebra’

Goelzer A. and Fromion V. Towards the modular decomposition of the metabolic network. in *System Theoretic Approach to Systems and Synthetic Biology*, Springer Verlag, (2013).

A MCA viewpoint on the local module properties

He F., Fromion V. and Westerhoff Hans V. (Im)Perfect robustness and adaptation of metabolic networks subject to metabolic and gene-expression regulation: marrying control engineering with metabolic control analysis, *BMC System Biology*, 2012



Modeling vs. data

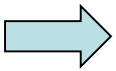
(what kind of interactions between the different entities?)

Glucose vs. malate

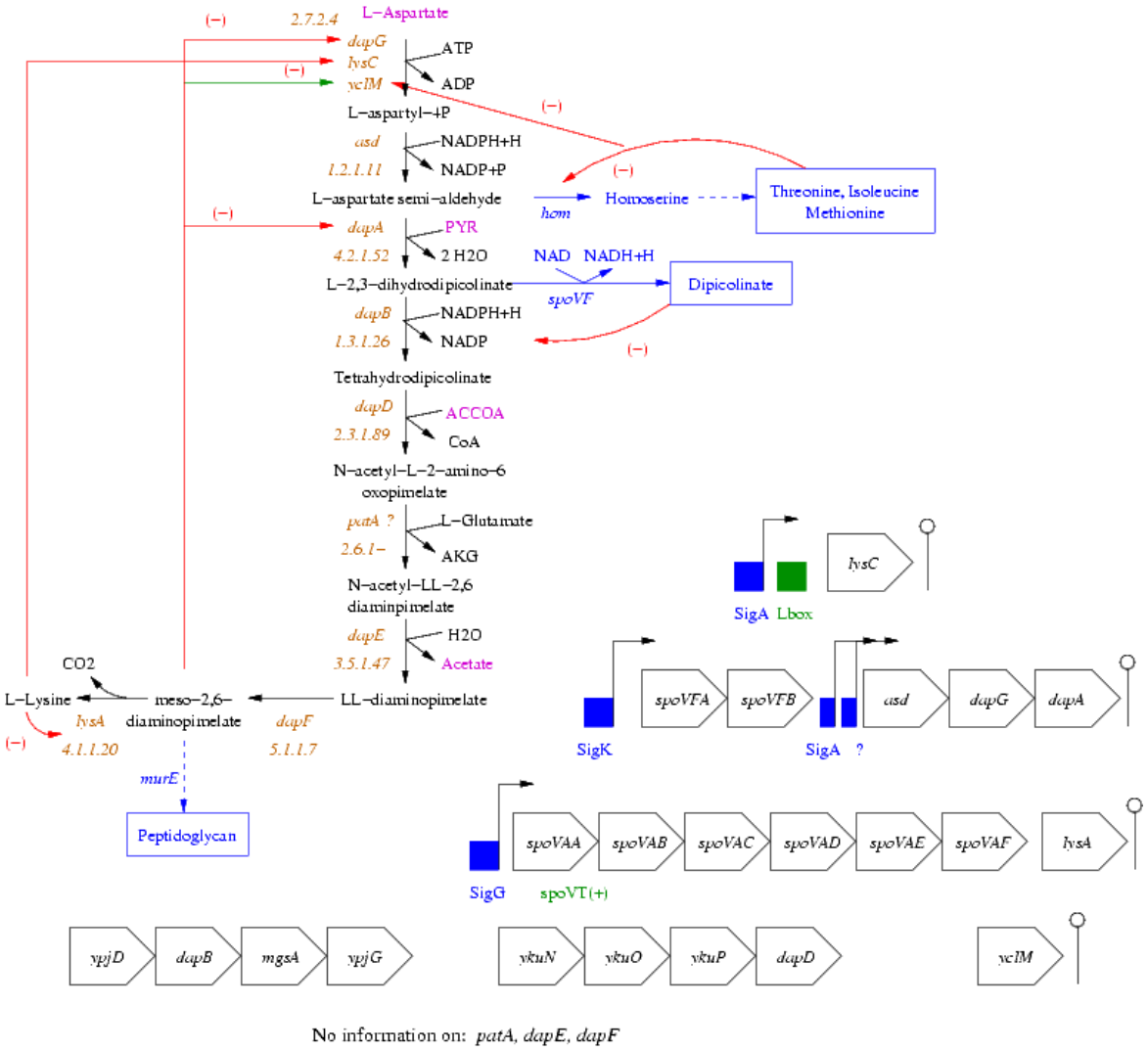
For both conditions, we have access through the 

project of

- Metabolite concentrations
- Transcriptomic data: tiling array
- Relative protein abundance
- Flux

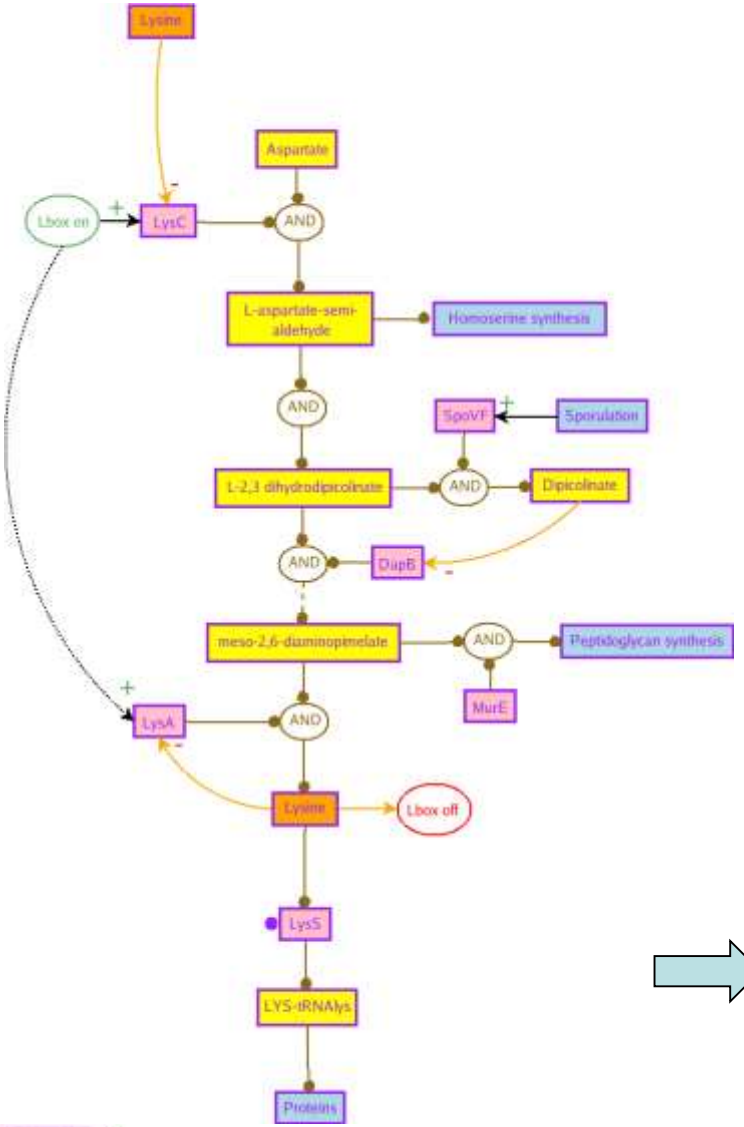


Classic differential analysis on transcriptomic data points out significant variations which are difficult to explain … as for example the significant repression of the first enzyme of the lysine synthesis pathway in the malate condition



Lysine synthesis pathway (extracted from Goelzer et al. BMC Syst Bio 2008)

The lysine pathway is an end-product module



- **LysC is irreversible due to the ATP hydrolysis,**
- **Inhibition of the first enzyme by the end product (lysine),**
- **Repression of the transcription of the first enzyme by the end-product through the L-box mechanism leading to a decreasing function of lysine.**

Qualitative prediction of the module behavior is then possible, in particular with respect to the aspartate concentration …

The observed variations deduced as a consequence of the end-product module properties

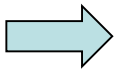
Module components	Malate	Glucose
Aspartate (µmol/gdwc)	10.47	1.42
Lysine (µmol/gdwc)	0.24	0.13
lysC (log)	12.3171	14.2756

- The significant variation of *lysC* mRNA is due to the strong variation of the aspartate node ….

- If the analysis is only considered at the transcriptomic level, the messenger of lysC is isolated… (it is controlled by a L-box)

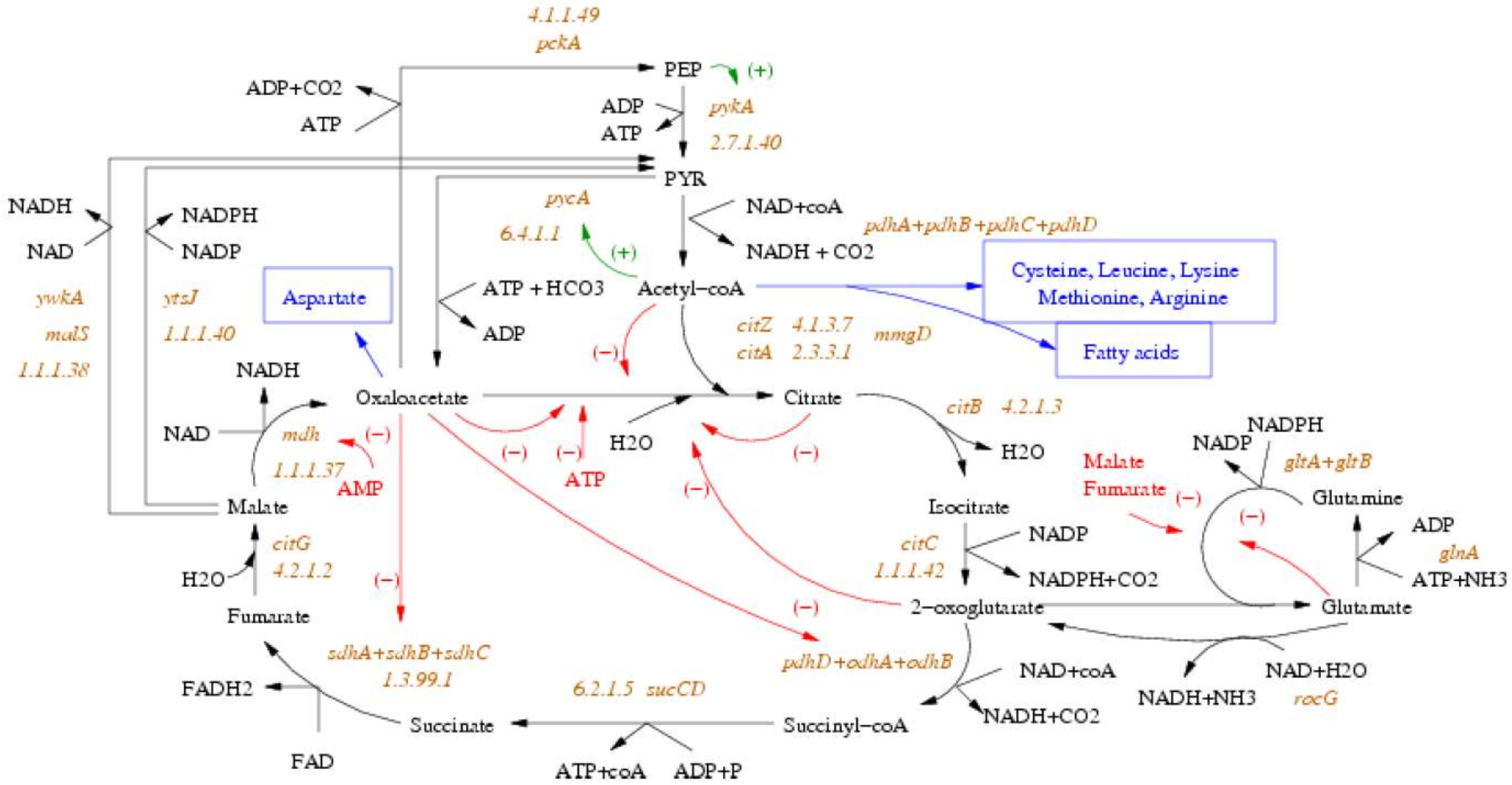
The observed variations deduced as a consequence of the end-product module properties

Model analysis and (good) data explain an indirect effect…quite easily

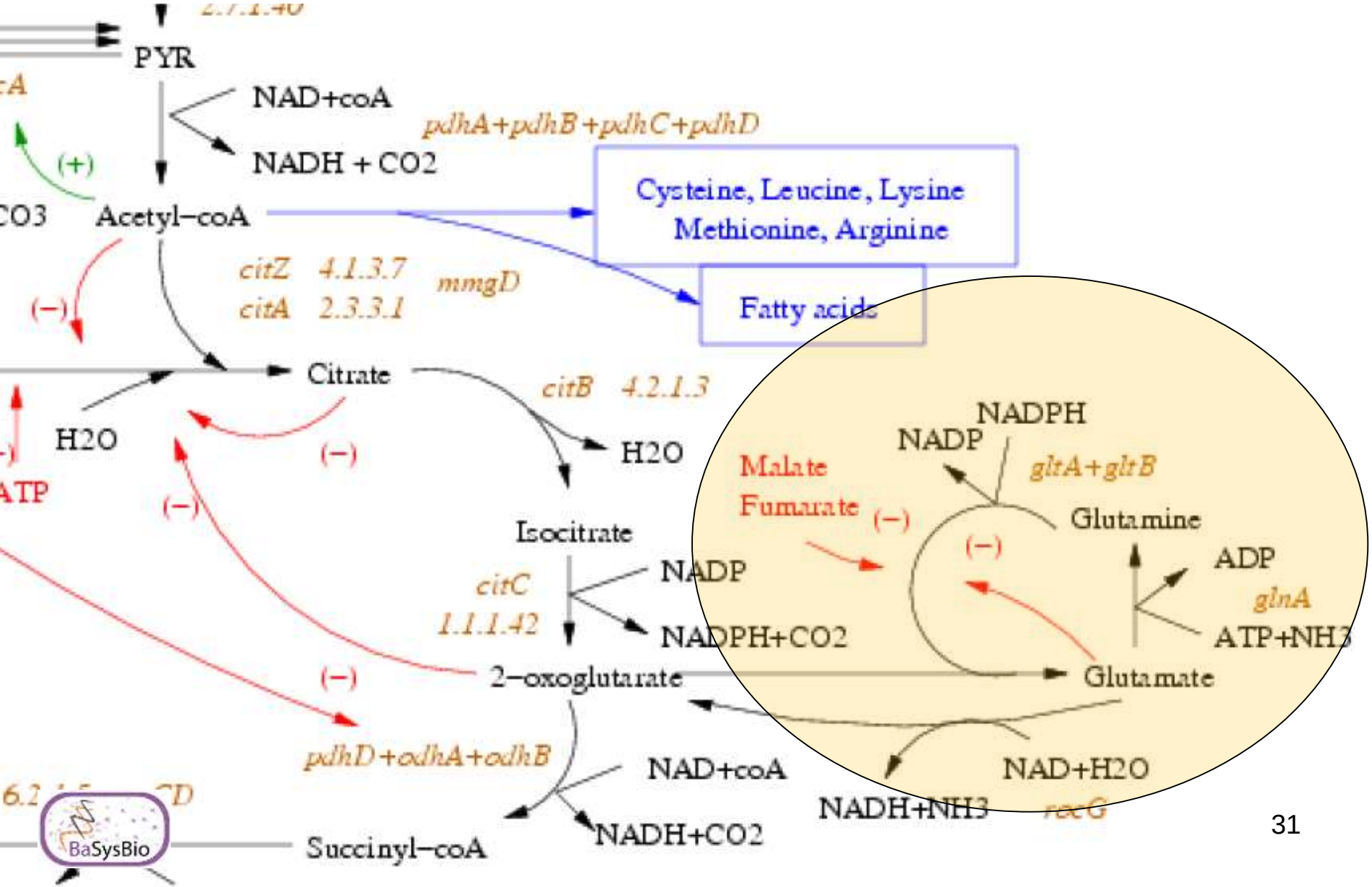


T
as

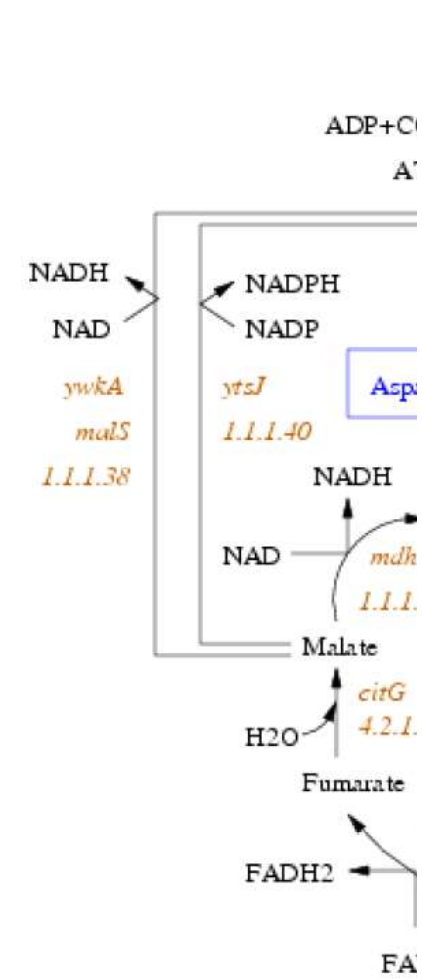
Variations of TCA metabolites and the glutamate/glutamine node

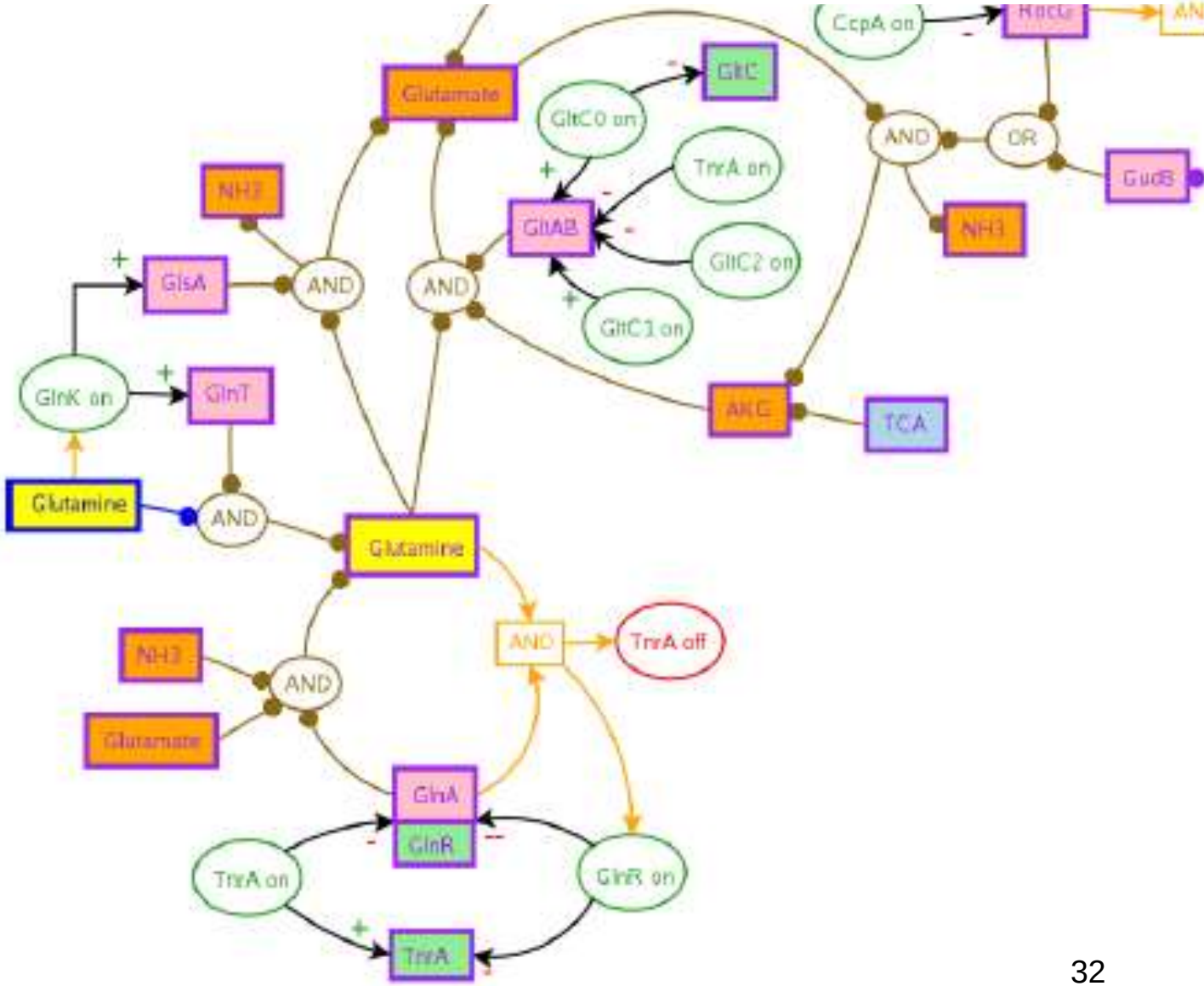


Variations of TCA metabolites and the glutamate/glutamine node



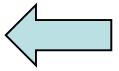
Variations of TCA metabolites and the glutamate/glutamine node





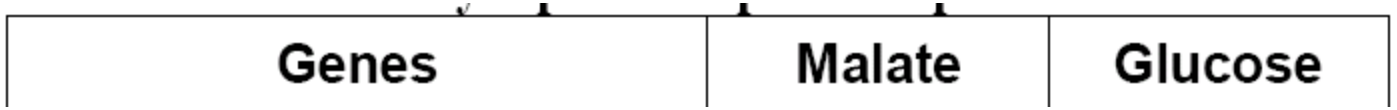
Variations of TCA metabolites and the glutamate/glutamine node

TCA components (µmol/gdwc)	**Malate**	**Glucose**
Malate	178.39	0.94
Citrate	6.34	1.61
Isocitrate	0.43	0.42
2-oxoglutarate	10.01	1.28
Succinate	125.12	9.53
Fumarate	0.57	0.45
Glutamine	59.67	6.30
Glutamate	86.17	65.70



BaSysBio

Variations of TCA metabolites and the glutamate/glutamine node



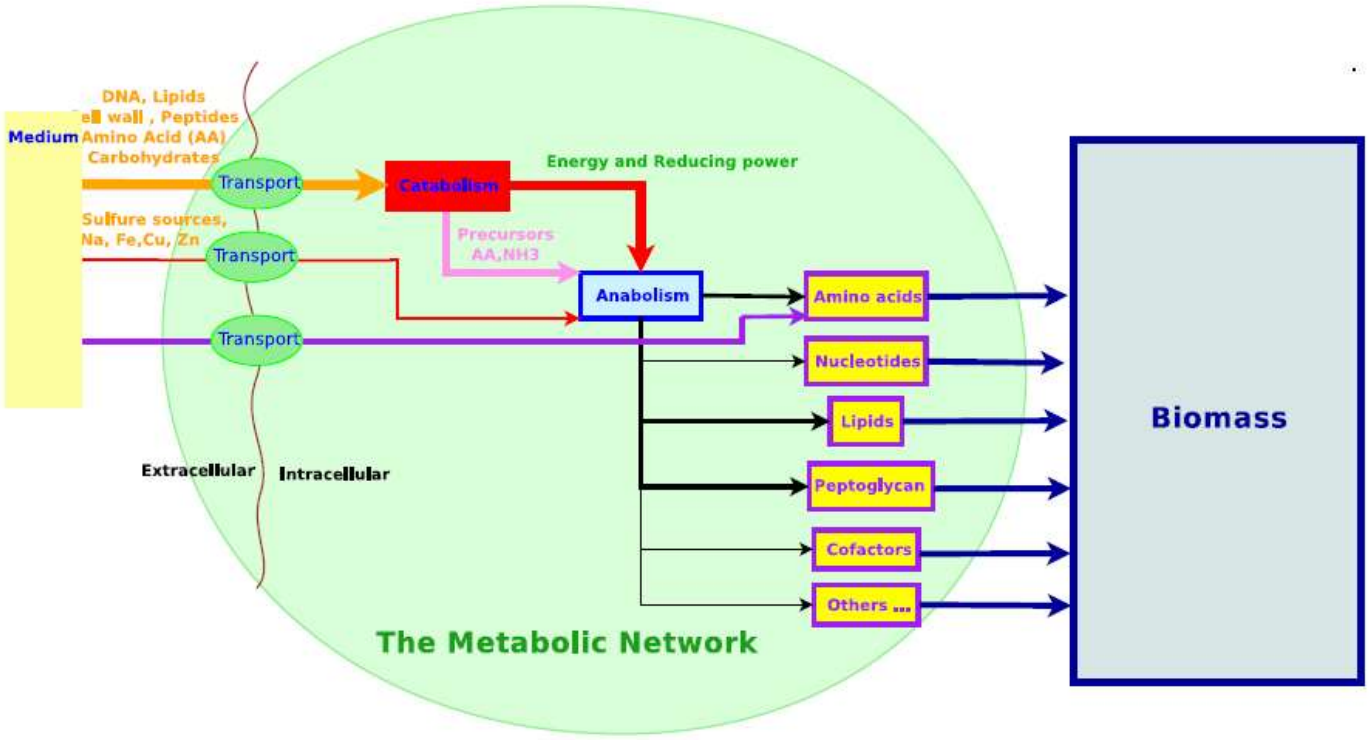
Model analysis (less obvious) and (very good) data explain and validate the complicated consequence of an effect of malate on the activity of the enzyme implies in the glutamine synthesis

To integrate all the levels of regulations,

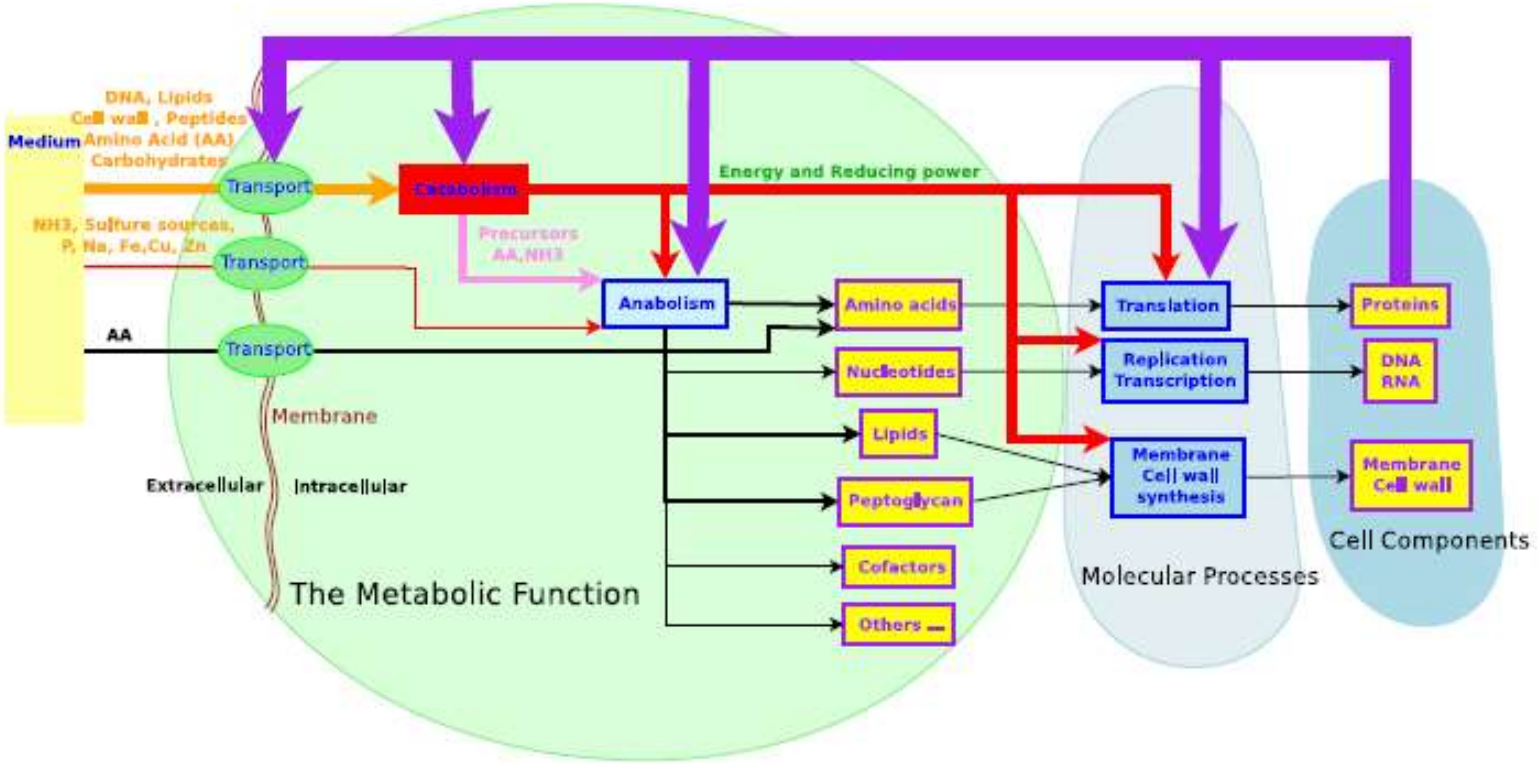
other levels of control

have to be considered …

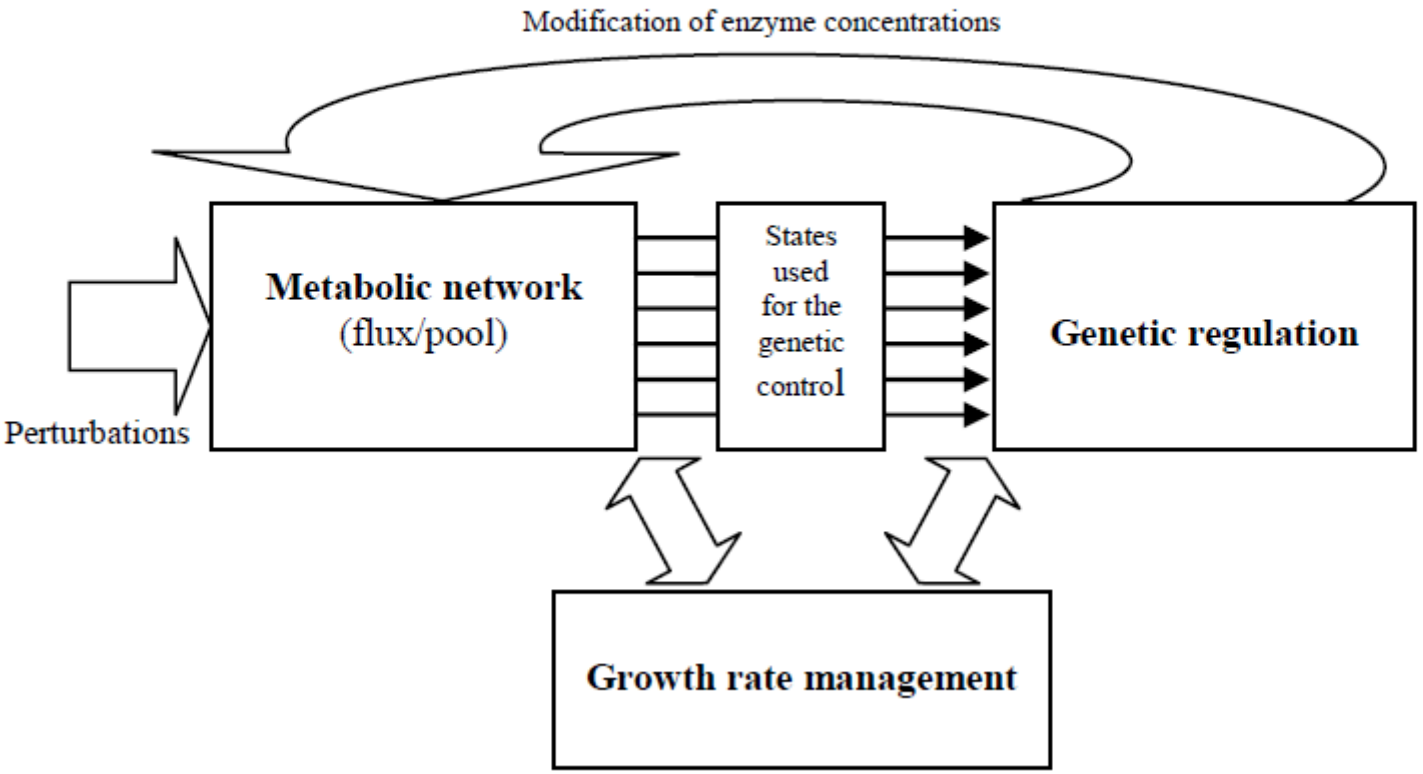
The metabolic function: a quick presentation



The metabolic function into “the whole cell”



Bacterial growth rate management loop

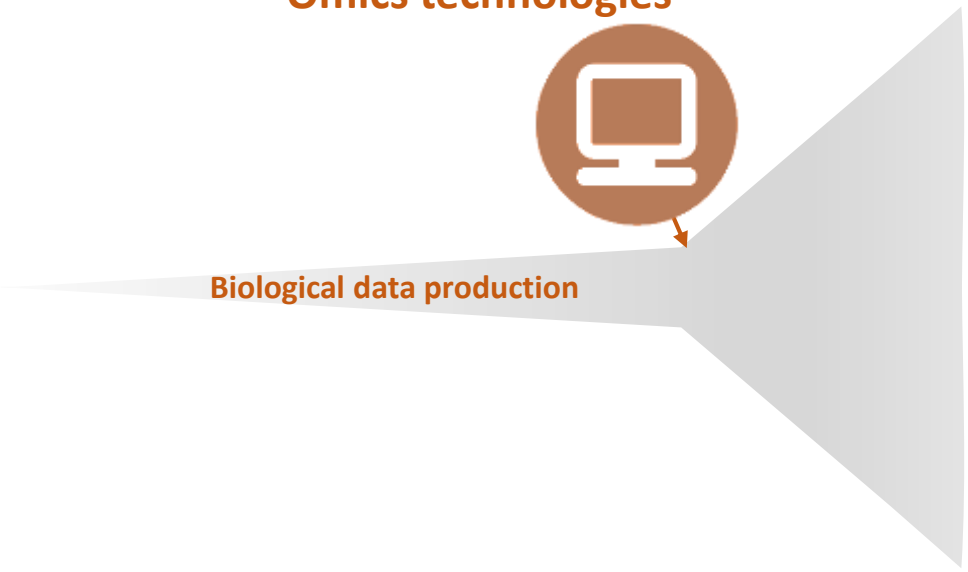


Goelzer, A., & Fromion, V. (2011). Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1810(10), 978-988.

Systems Biology

opens opportunities for the representation of biological data, information and knowledge

DATA deluge in the biological field …

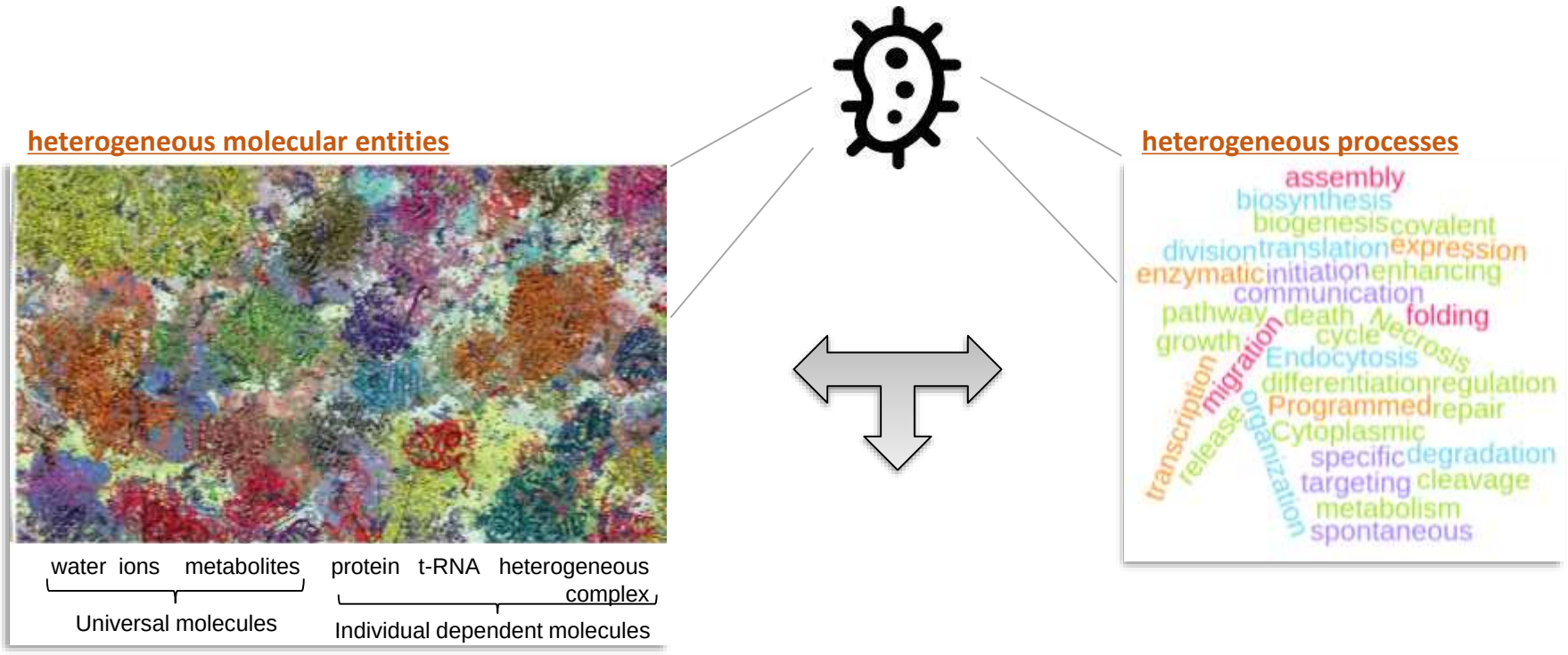


Needs for data and knowledge management

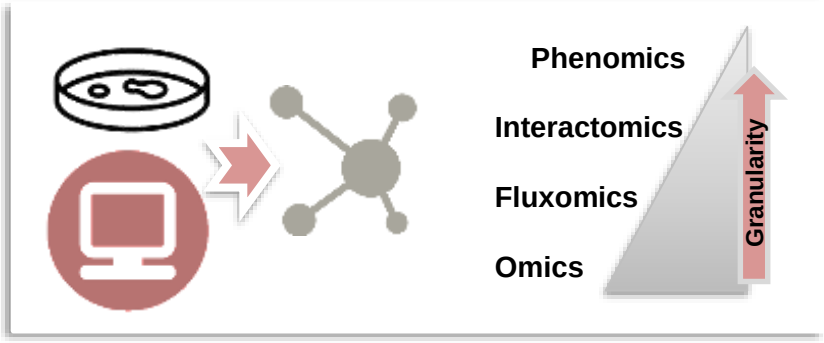
Bio-ontologies: controlled vocabularies for

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

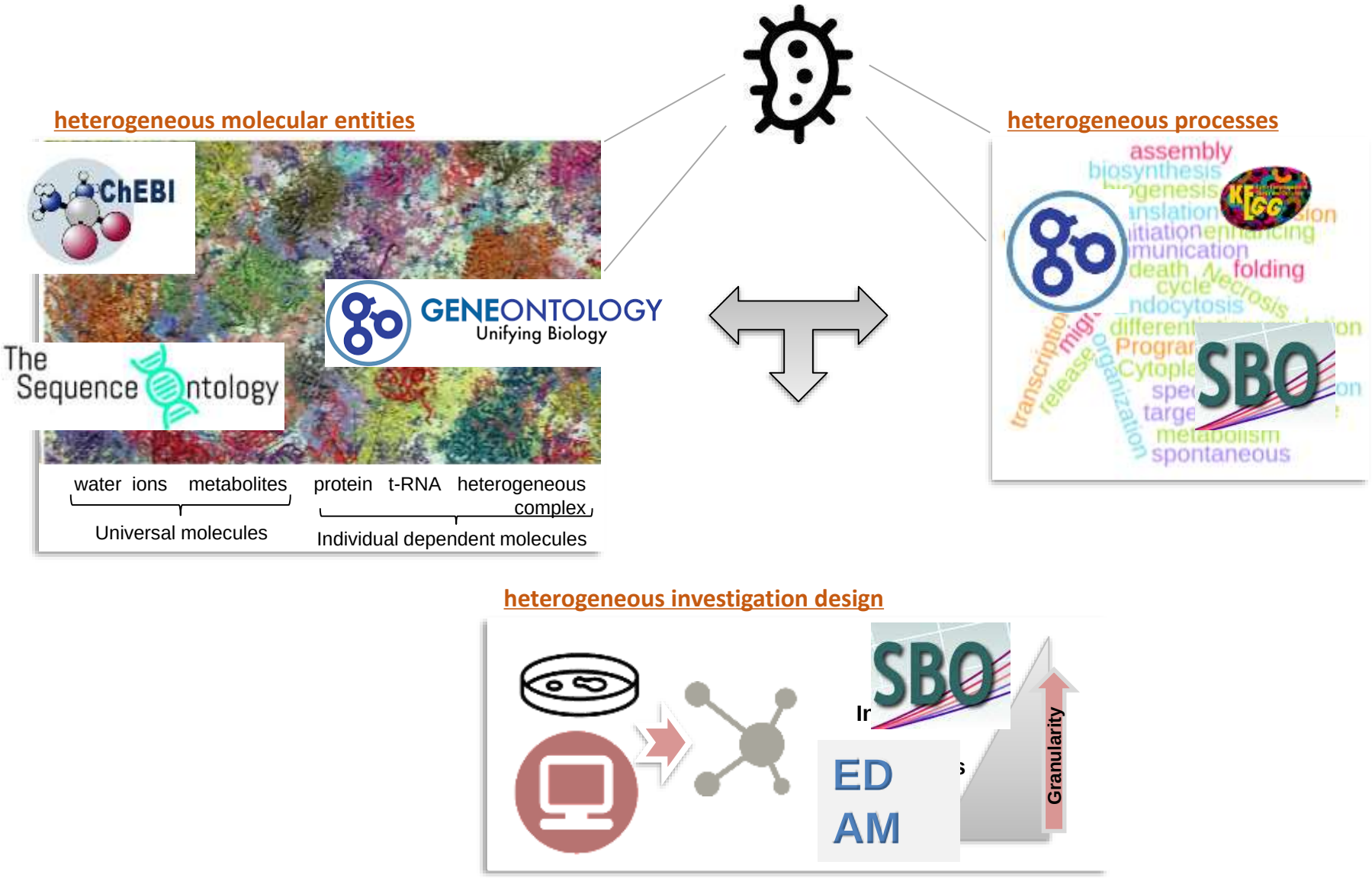
Cellular and molecular biology is a wide and heterogeneous field



heterogeneous investigation design



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



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

GO-Biological Process (BP) | GO-Molecular Function (MF) | GO-Cellular Component (CC)

GENEONTOLOGY
Unifying Biology

ChEBI

Gene | Gene product

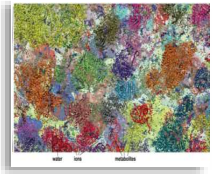
Sequence

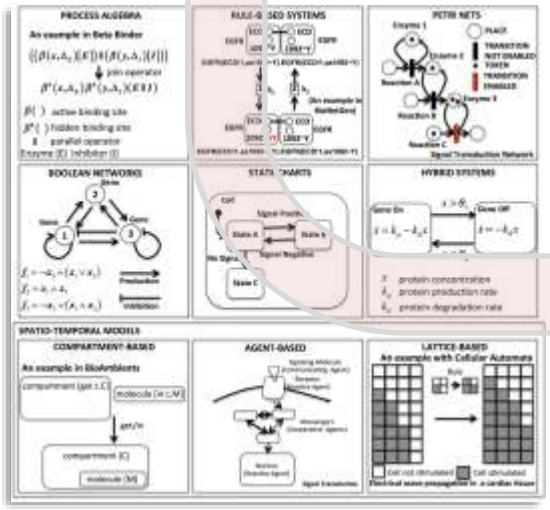
Chemical

The Sequence Ontology

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

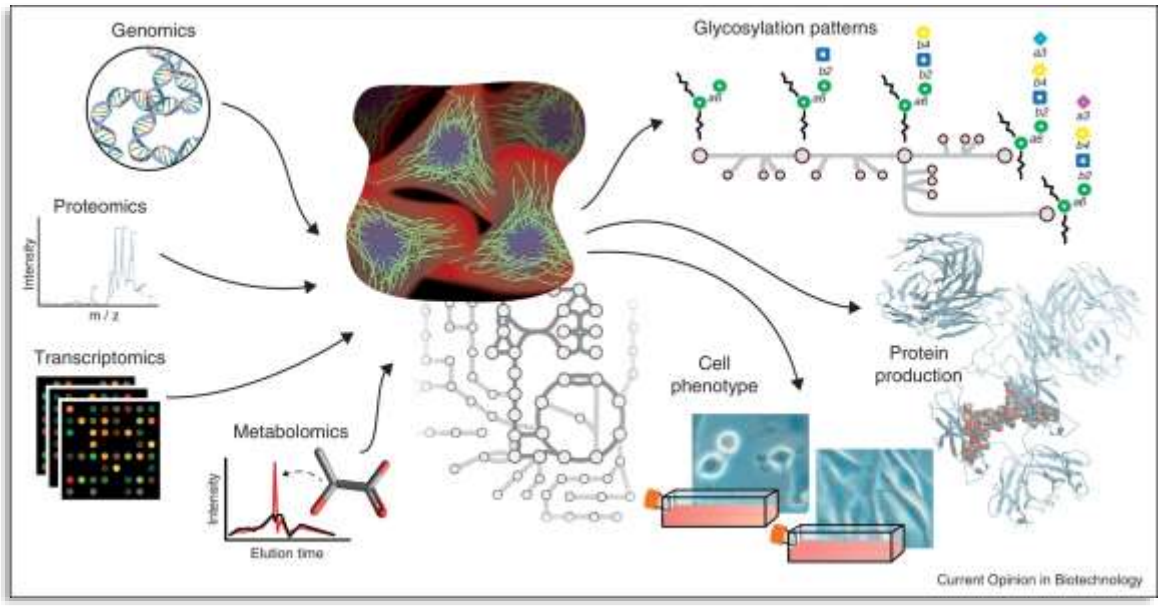
Systems Biology is a suitable framework to integrate heterogeneous entities at different scales…





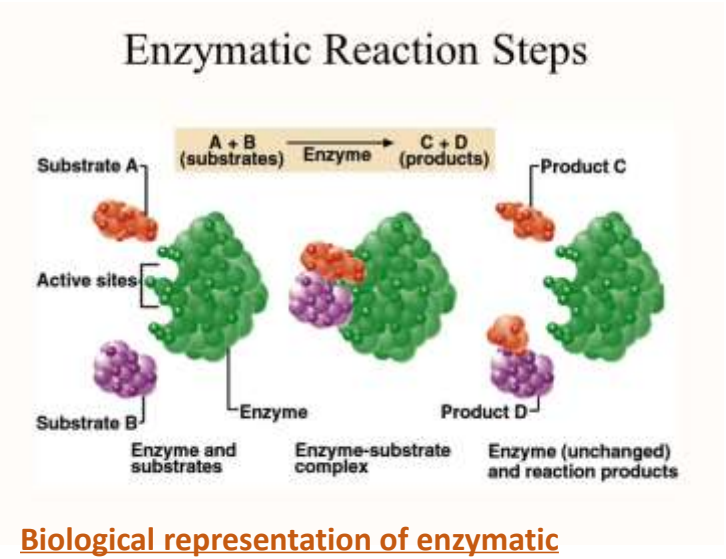
Mathematical models

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

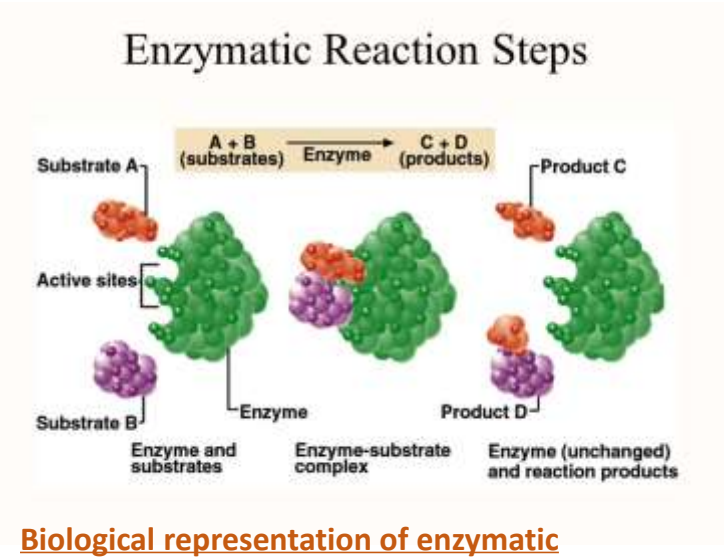


Biological representation of enzymatic reaction

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

Mathematical representation of enzymatic reaction

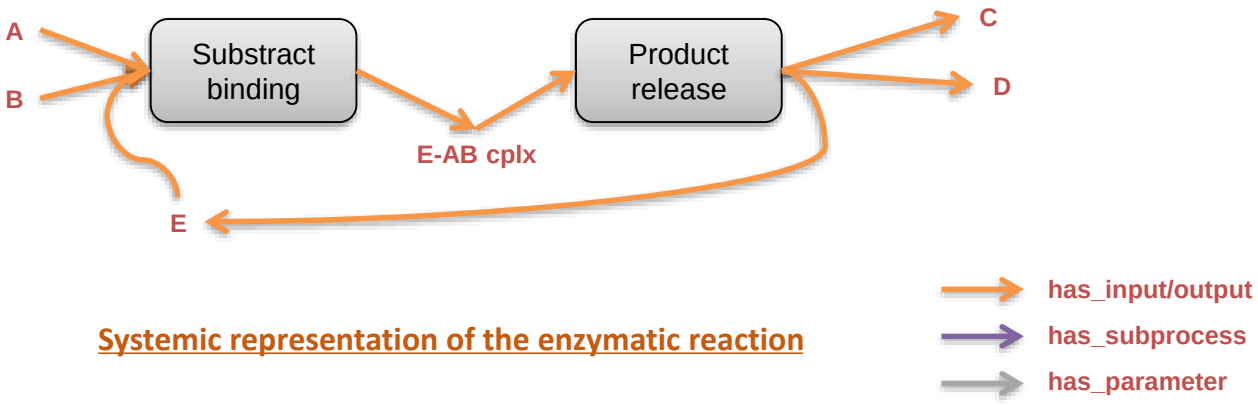
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge



Biological representation of enzymatic reaction

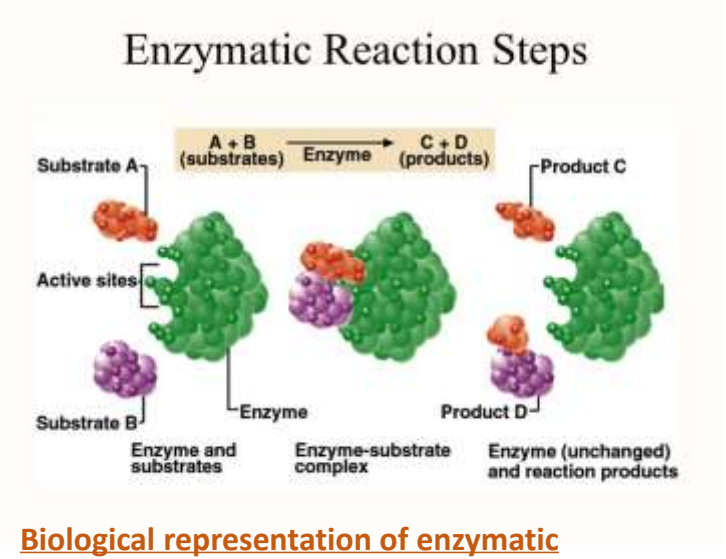
$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

Mathematical representation of enzymatic reaction



Systemic representation of the enzymatic reaction

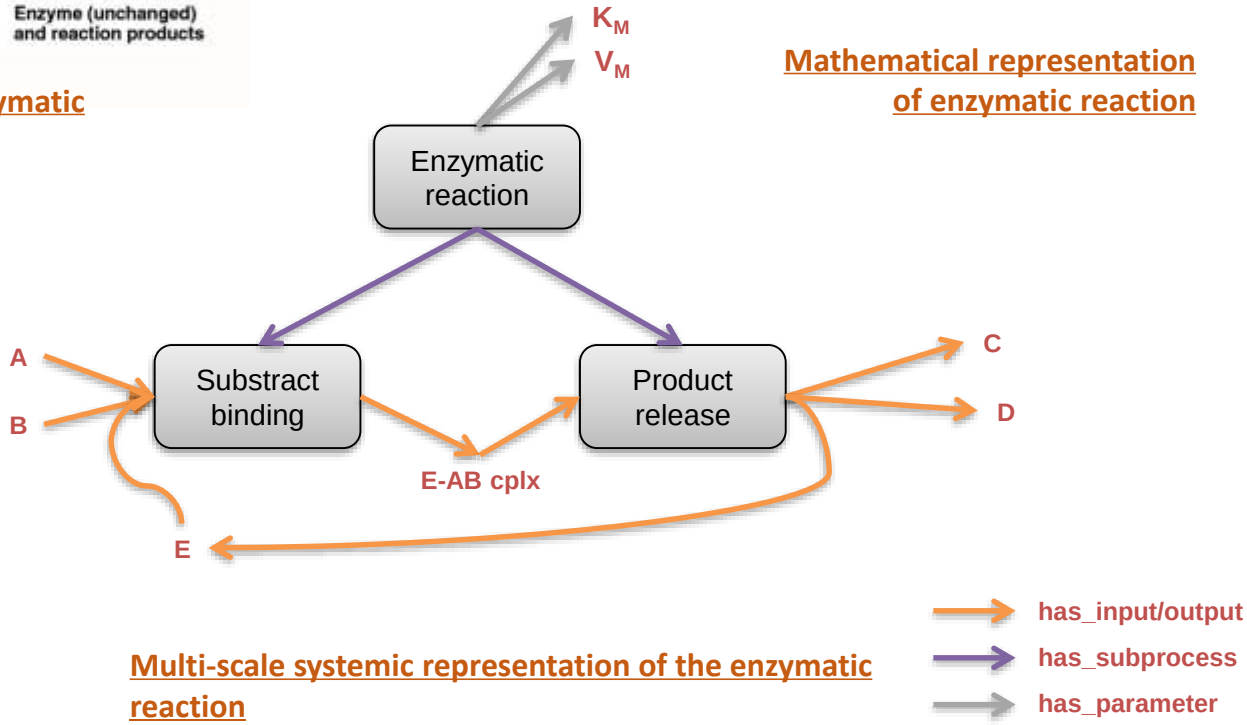
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge



Biological representation of enzymatic reaction

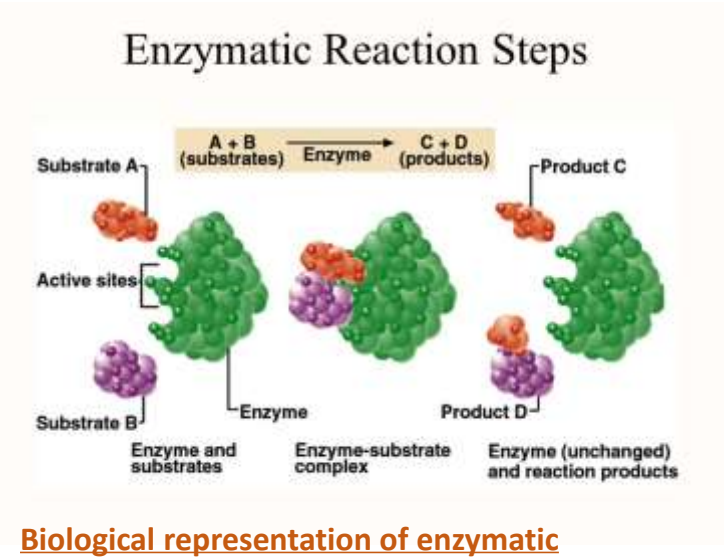
$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

Mathematical representation of enzymatic reaction



Multi-scale systemic representation of the enzymatic reaction

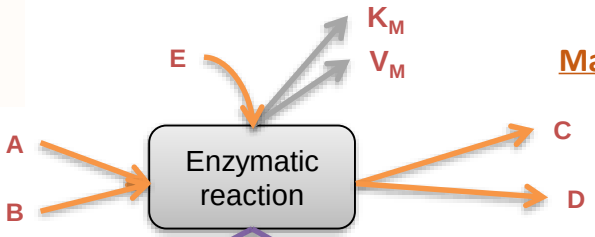
Systemic approach appears to be a good integrative framework to support and relate the representation of biological and mathematical knowledge

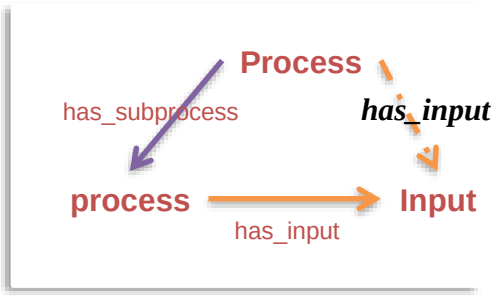


Biological representation of enzymatic reaction

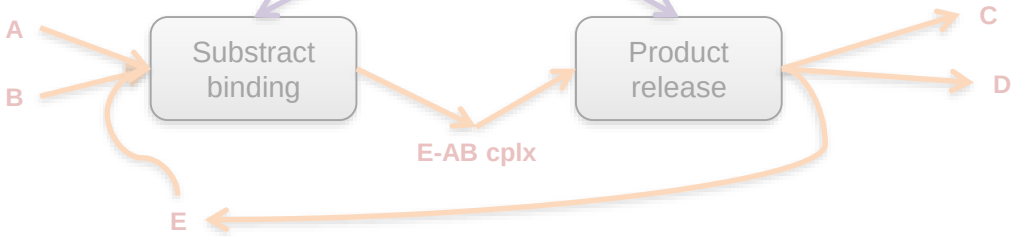
$$V_0 = \frac{V_{max}[S]}{(K_M + [S])}$$

Mathematical representation of enzymatic reaction





Logical rules



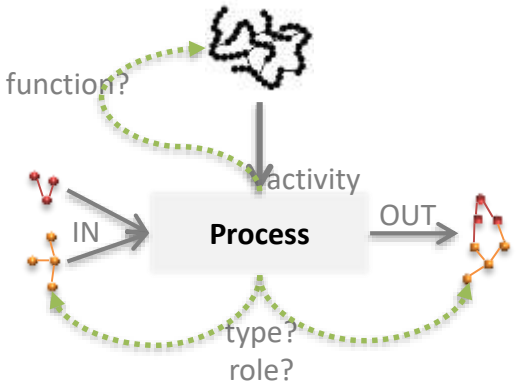
Automatic aggregation of the multi-scale systemic representation of the enzymatic reaction

has_input/output/mediator
has_subprocess
has_parameter

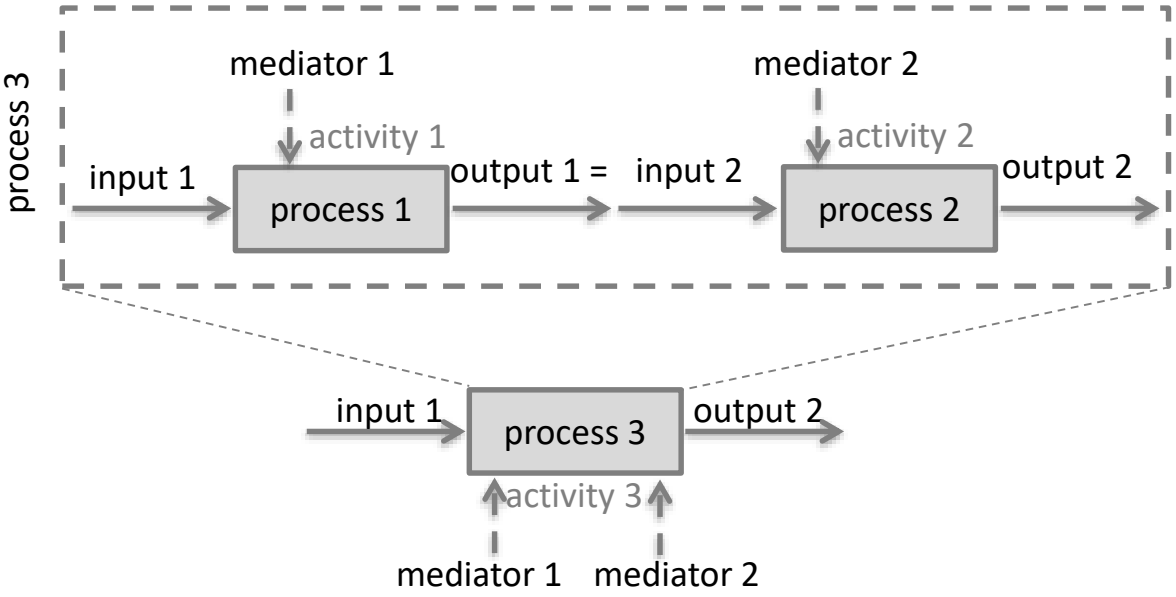
The main hypothesis of the approach

- 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

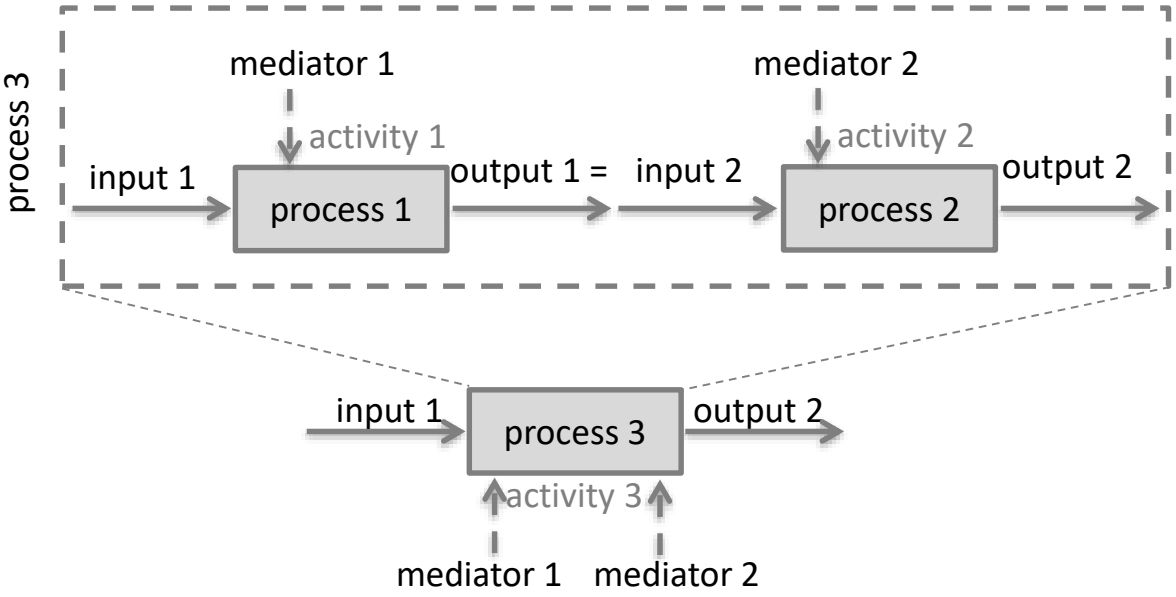
A fine description of biological processes as an instances should automatically conferred properties to its participants

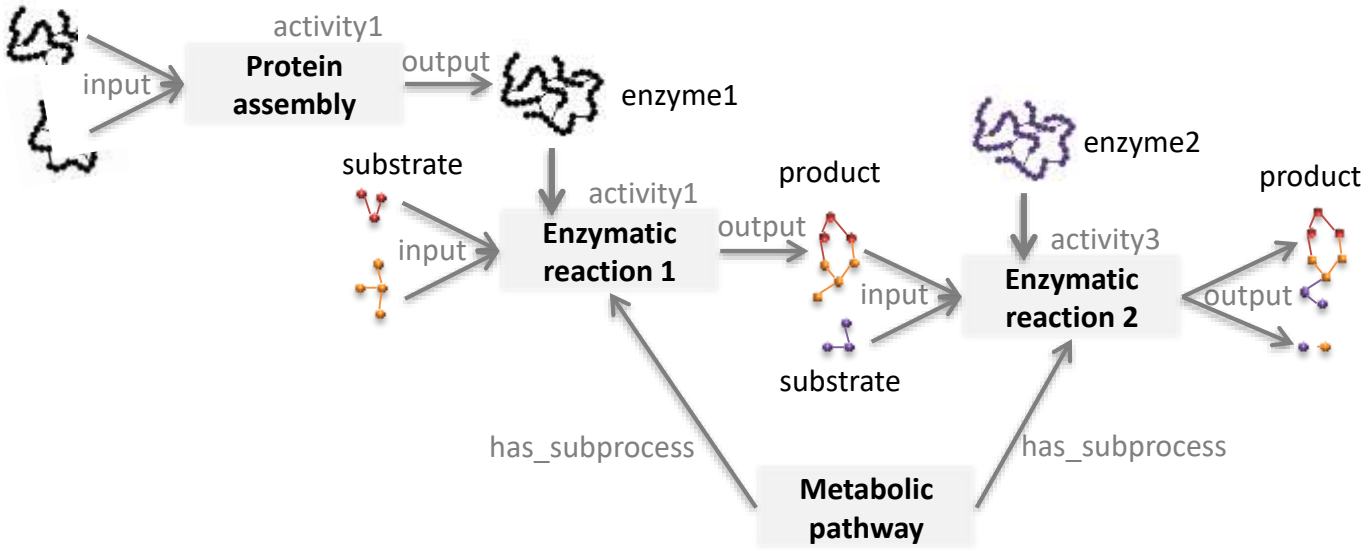


Systemic approach: a process-centered representation of systems



System biology: a process-centered representation of biology



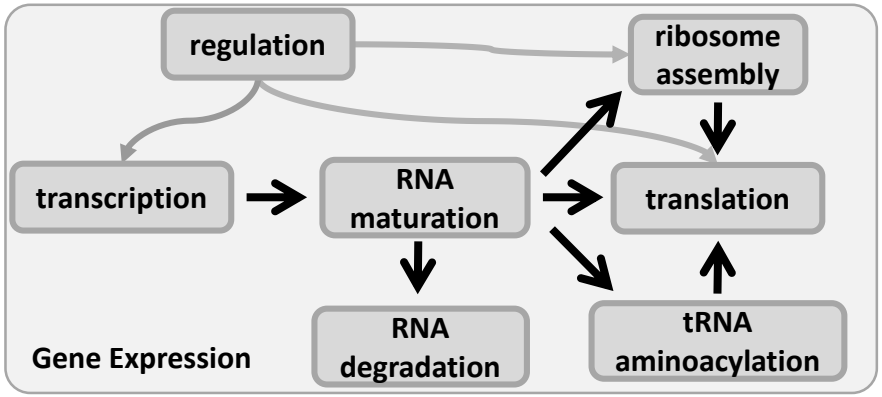


System biology: a process-centered representation of biology

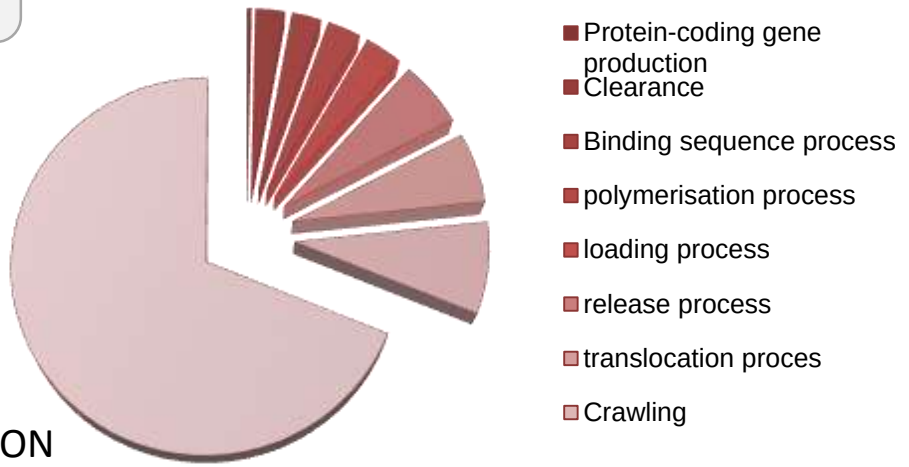
Bacterial interlocked Process ONtology (BiPON)

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

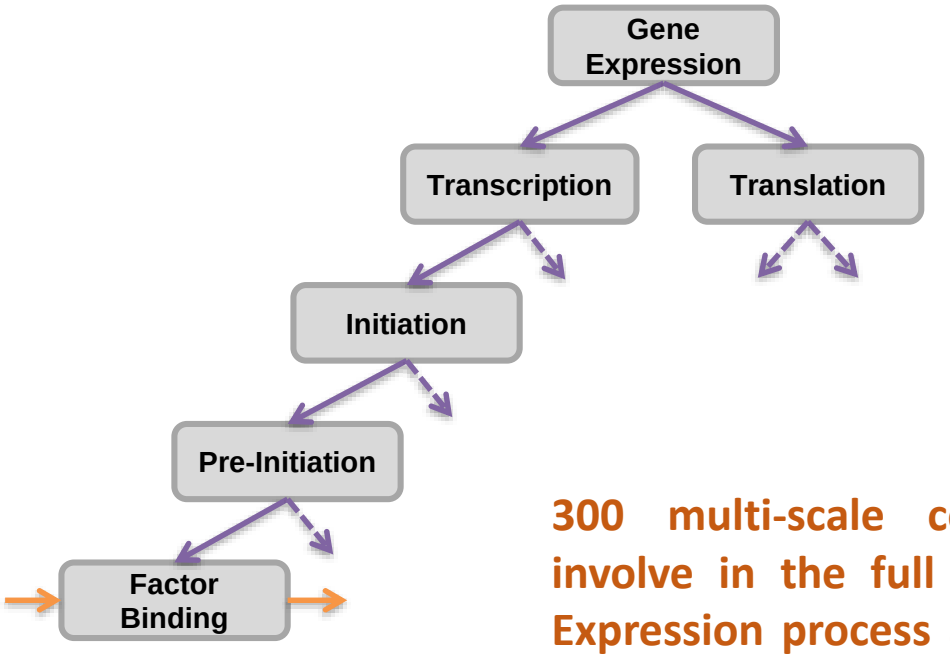
The bacterial interlocked process ONtology (BiPON): a systemic multi-scale unified representation of biological processes in prokaryotes

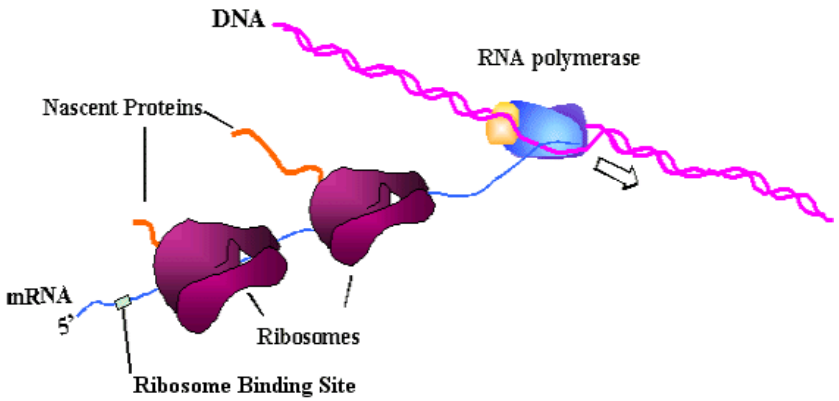
Vincent J. Henry [1,2†], Anne Goelzer [2*†], Arnaud Ferré [1], Stephan Fischer [2], Marc Dinh [2], Valentin Loux [2], Christine Froidevaux [1] and Vincent Fromion [2]

→ 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**

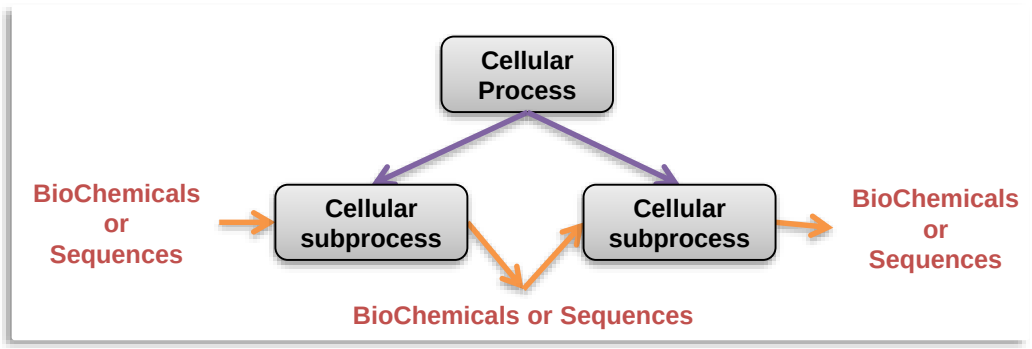
Modeling heterogeneous and multi-scale processes of bacterial gene expression

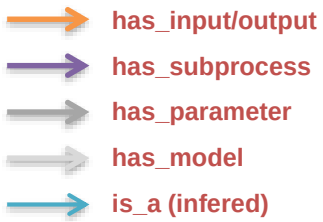




300 multi-scale cellular processes involve in the full aggregated Gene Expression process and its regulation processes

Biological knowledge representation

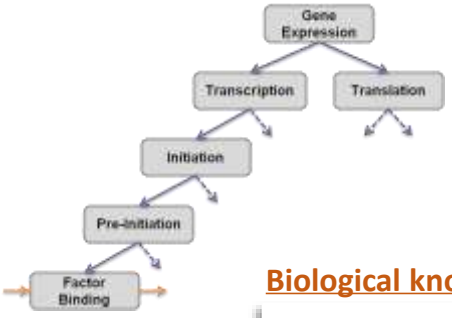




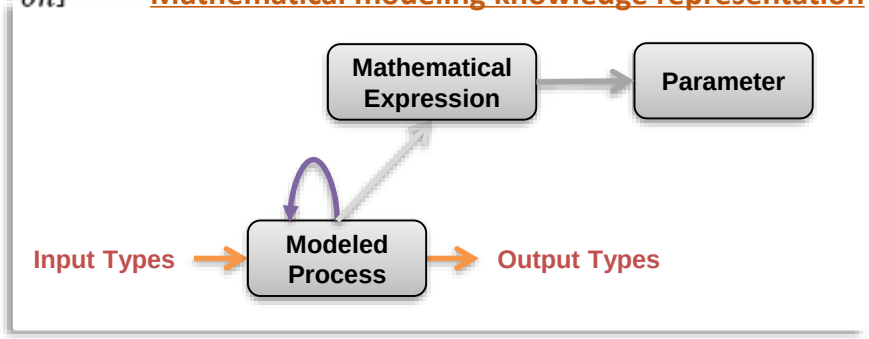
Modeling heterogeneous and multi-scale processes of bacterial gene expression

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

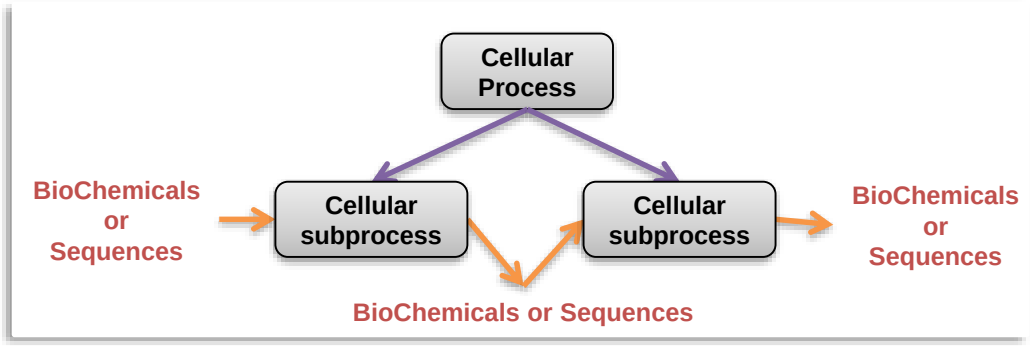
9 multi-scale Modeling Processes related to their mathematical expressions and parameters

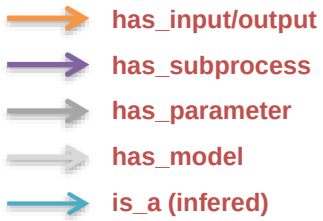


Mathematical modeling knowledge representation



Biological knowledge representation

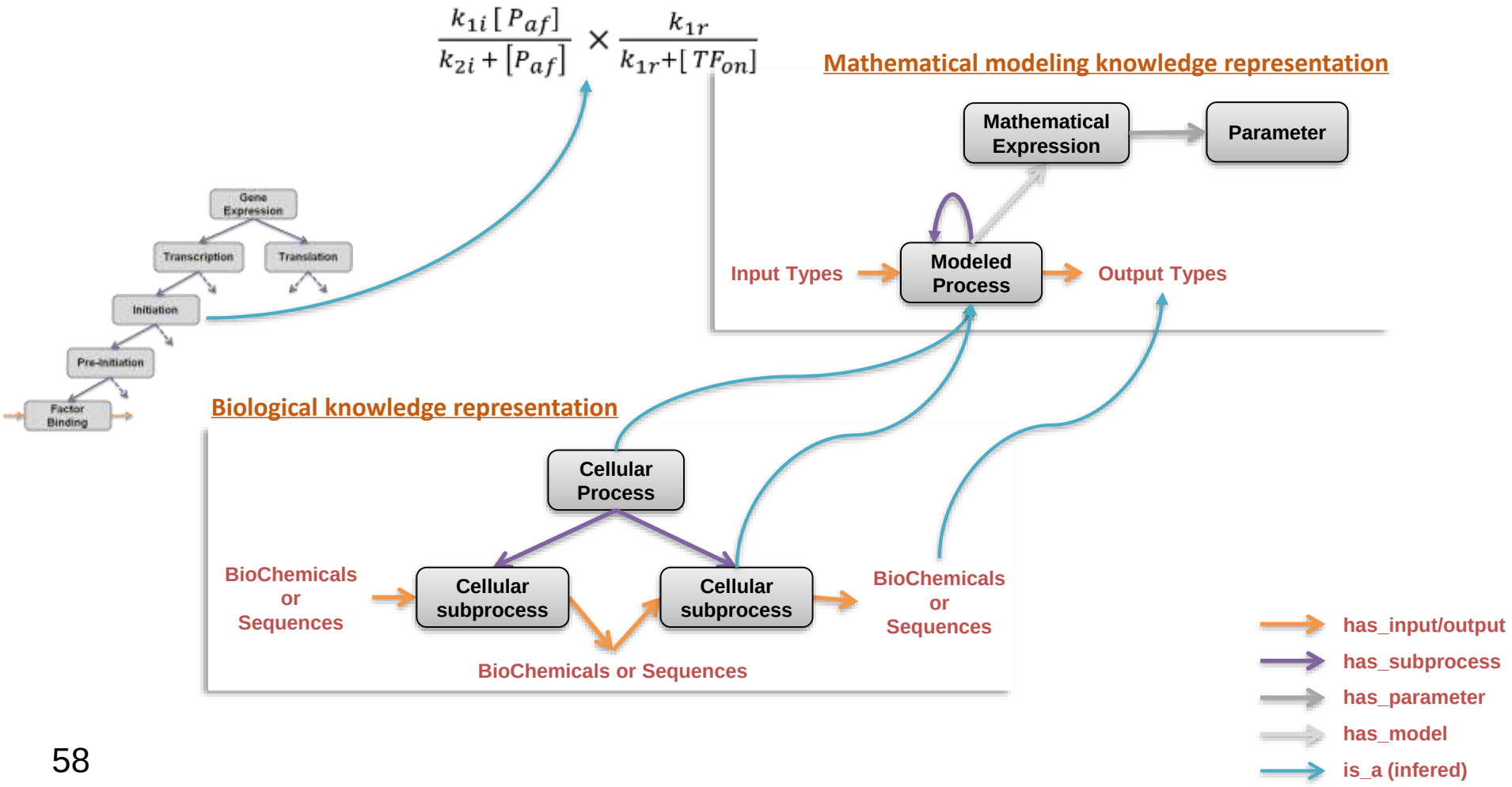




Modeling heterogeneous and multi-scale processes of bacterial gene expression

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)