Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens

Hybrid approach for explainable metabolic modelling of microbial ecosystems'

Présenté par Maxime LECOMTE

November 20, 2023

Membres du jury

Président: SIMON Laurent

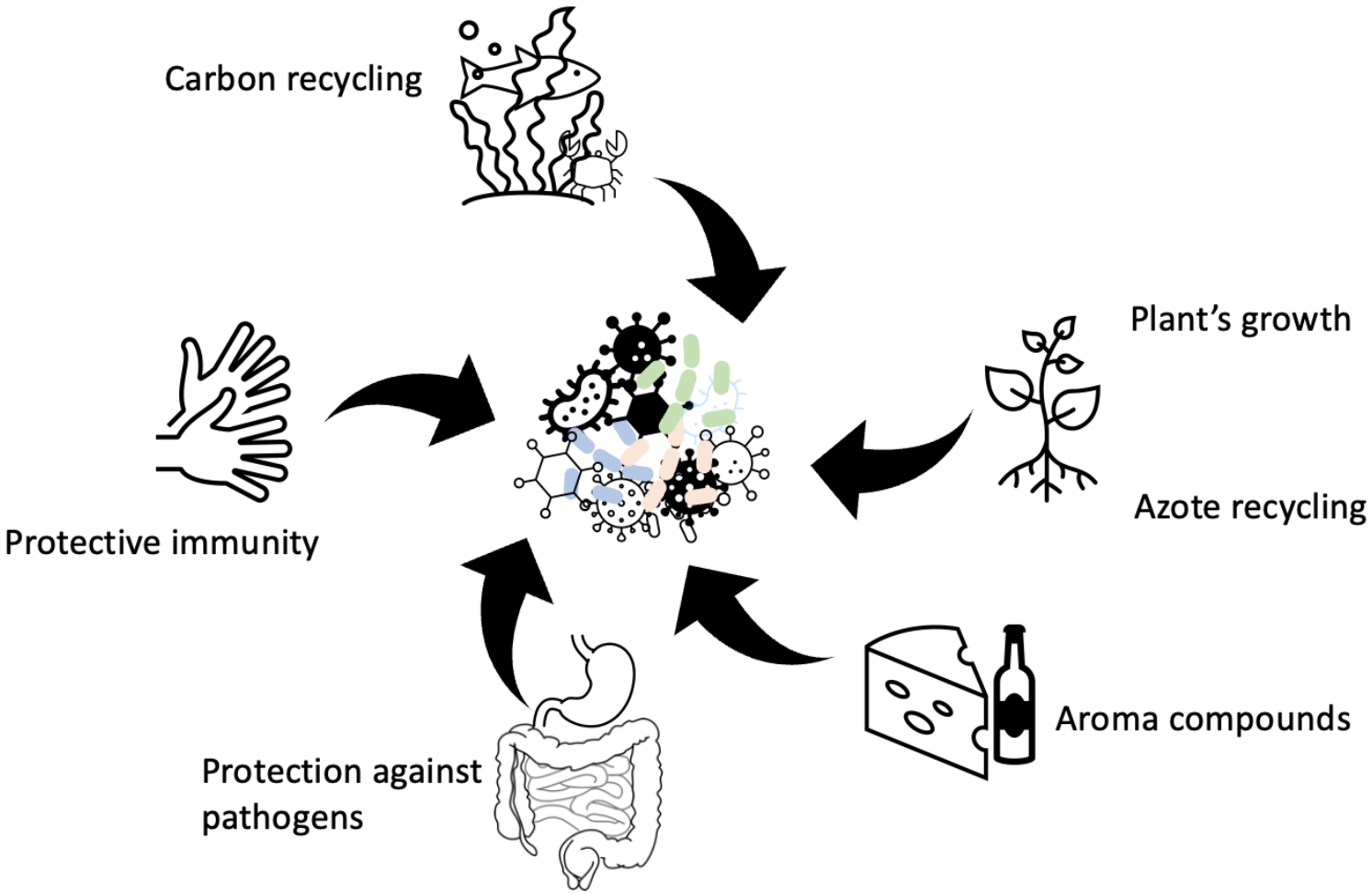
Rapportrices: BAROUKH Caroline
COCAIGN-BOUSQUET Muriel

Examinateurs: COTTRET Ludovic
LAROCHE Béatrice
MARKOV Gabriel

Co-direction: David SHERMAN et Hélène FALENTIN
Encadrement: Clémence FRIOUX

École doctorale Mathématiques et informatique — Université de Bordeaux — INRAE — Inria — LaBRI

Why the study of microorganisms is relevant ?



- High diversity of microorganisms
- Microorganisms roles specific to the environment (Royet and Plailly, 2004; Belkaid and Hand, 2014; Zhang et al., 2015; Hoorman, 2011; McSweeney and Sousa, 2000)

Bacterial interaction are responsible of the observed roles

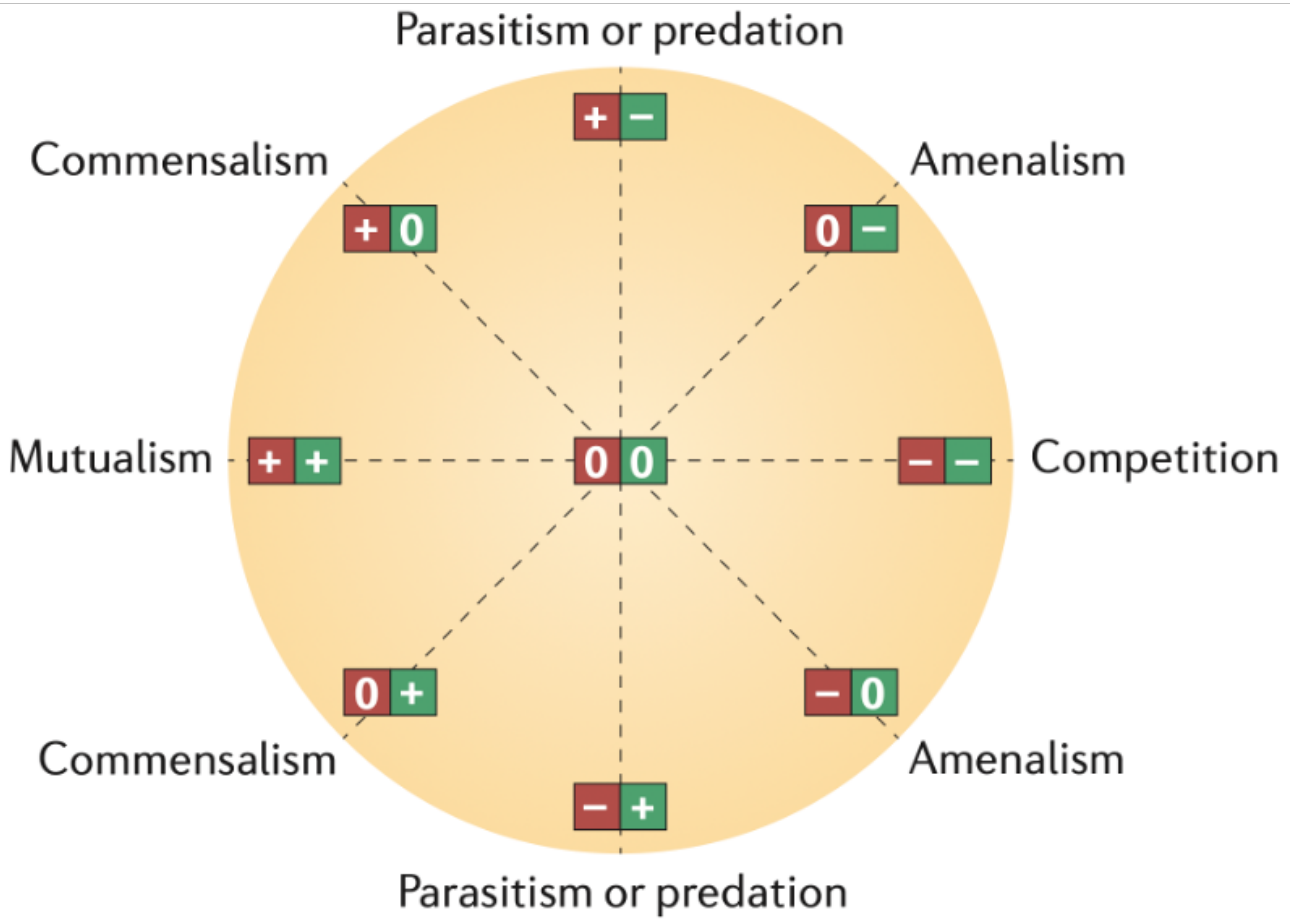


Figure 1: List of different types of bacterial interactions (Faust and Raes, 2012)

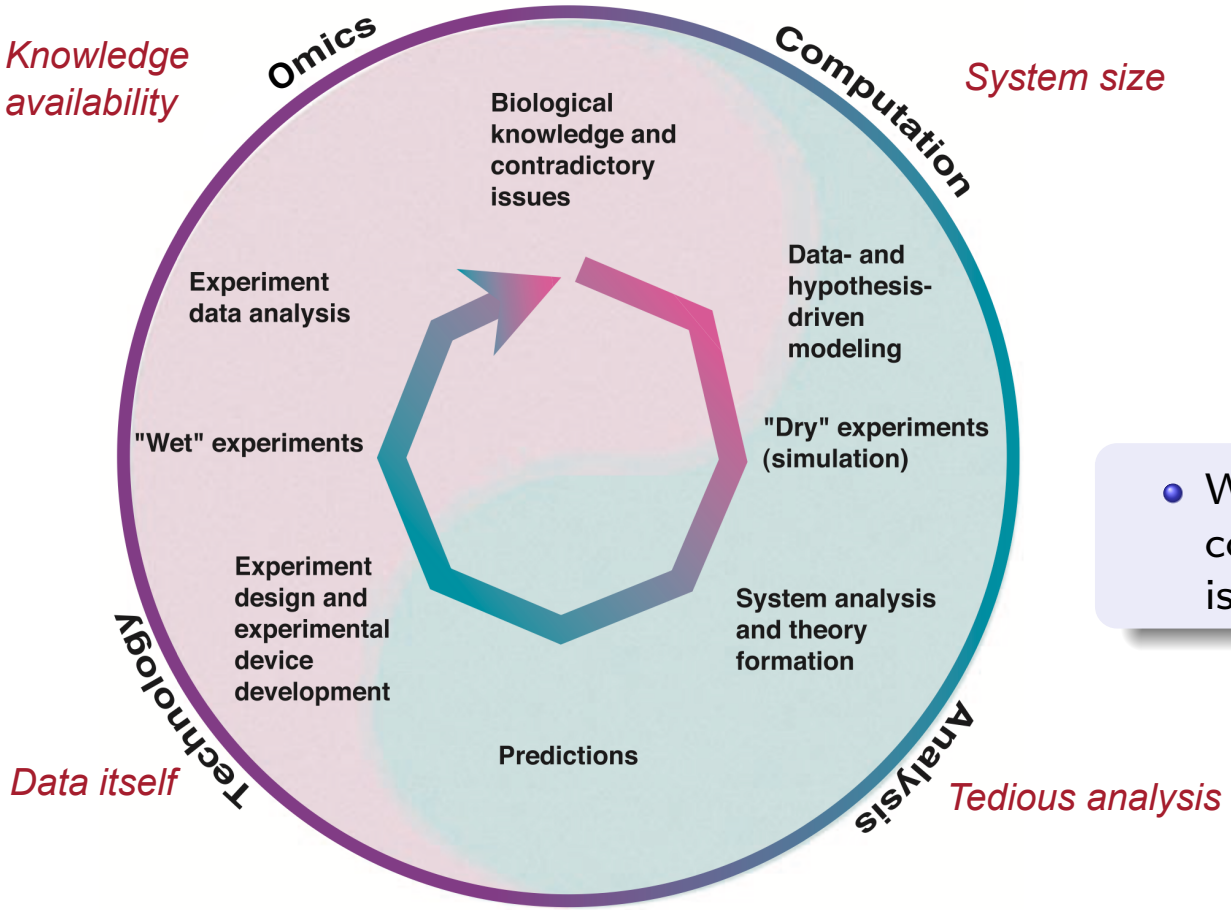
- Bacterial interactions are distinguishable within two species
- And within ecosystems composed of thousand of species ? → need of informatic

How can we combine biological knowledge and infomatic program ?

Systems biology

System biology

Associate an organism to a system and study the all system (Kitano, 2002)



- Which type of computation model is involved ?

Figure 2: System biology modified from Kitano, 2002

Metabolism as a starter pack for analysing bacterial interactions

Metabolism

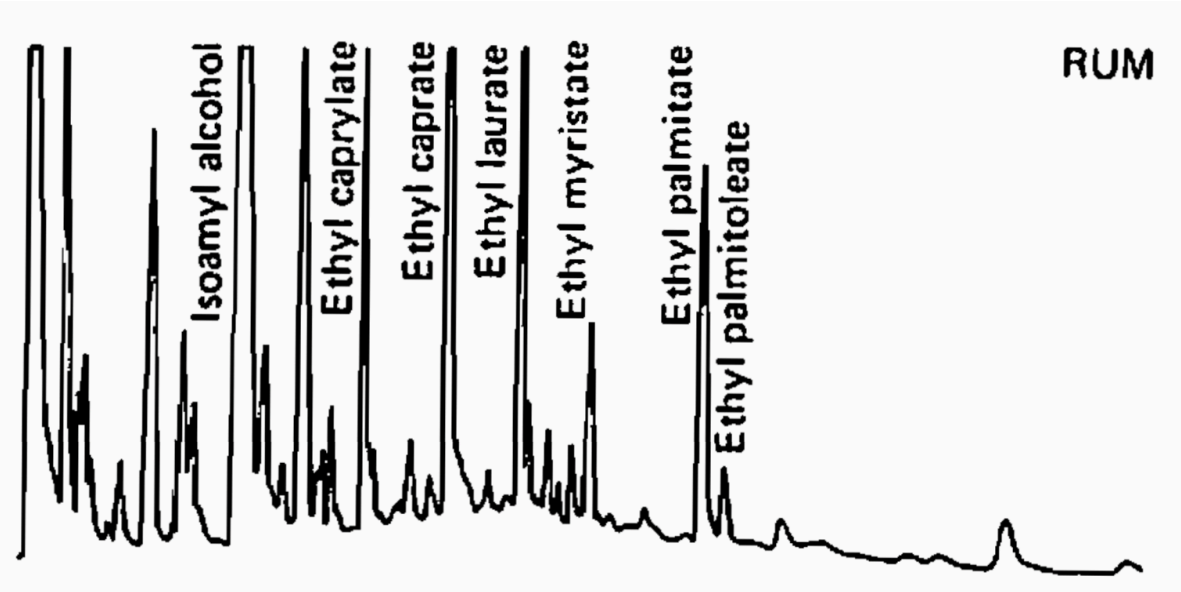


Figure 3: Gas chromatograms of the major aroma compounds isolated from rum (from Suomalainen and Lehtonen, 1978)

Metabolism as a starter pack for analysing bacterial interactions

Metabolism

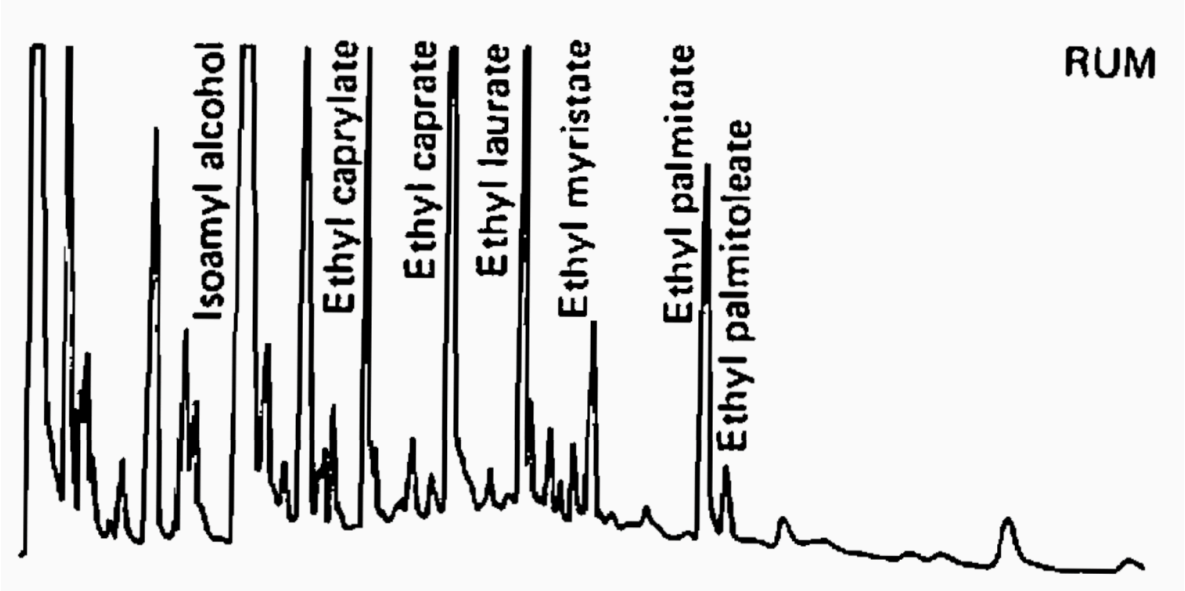


Figure 3: Gas chromatograms of the major aroma compounds isolated from rum (from Suomalainen and Lehtonen, 1978)

What is metabolism ?

Set of all biochemical reactions occurring in the cell of an organism that permit the production of energy and metabolic goods. (Sánchez López de Nava A, 2023)

What underlying mechanisms are responsible of the observed activity ?

Metabolism and Bacterial interactions

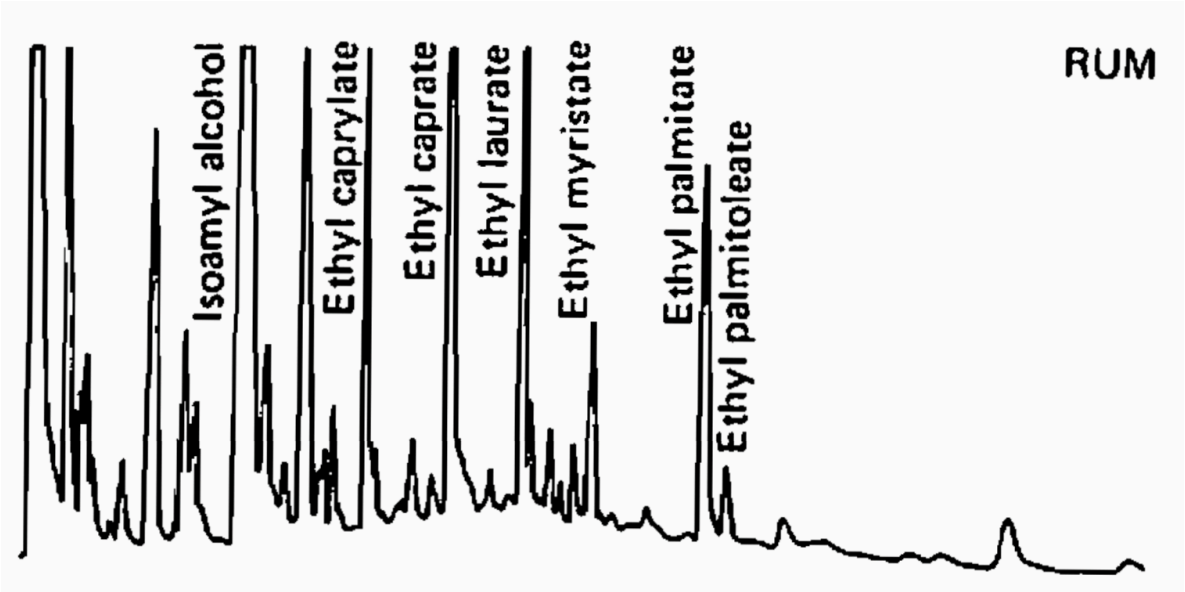


Figure 3: Gas chromatograms of the major aroma compounds isolated from rum (from Suomalainen and Lehtonen, 1978)

What is metabolism ?

Set of all biochemical reactions occurring in the cell of an organism that permit the production of energy and metabolic goods. (Sánchez López de Nava A, 2023)

a fatty acyl-CoA + ethanol → ethyl palmitate + coA

What underlying mechanisms are responsible of the observed activity ?

Metabolism and Bacterial interactions

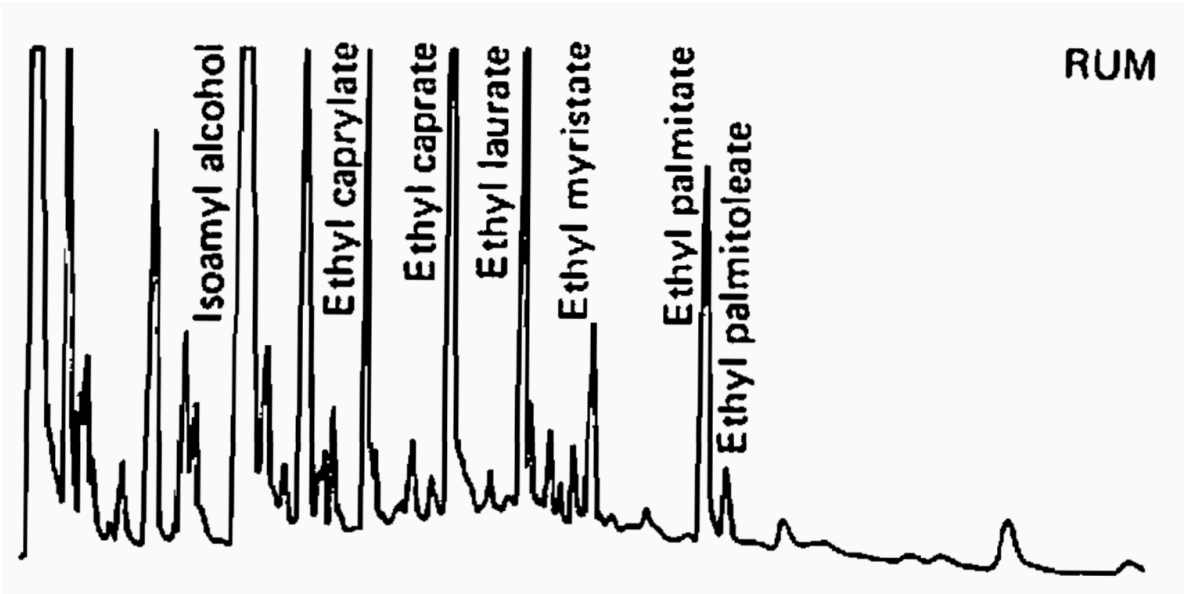


Figure 3: Gas chromatograms of the major aroma compounds isolated from rum (from Suomalainen and Lehtonen, 1978)

What is metabolism ?

Set of all biochemical reactions occurring in the cell of an organism that permit the production of energy and metabolic goods. (Sánchez López de Nava A, 2023)

a fatty acyl-CoA + ethanol → ethyl palmitate + coA

- Metabolism of an organism explain observable phenotype
- Is impacted by bacterial interactions

How is the metabolism represented?

r_1 :2 pyr $\rightarrow$ 1 acetoL + 1 CO_2
r_2 :1 acetoL $\rightarrow$ 1 diac + 1 CO_2
r_3 :1 acetoL $\rightarrow$ 1 acetoin + 1 CO_2
r_4 :1 diac $\rightarrow$ 1 acetoin
r_5 :1 acetoin $\rightarrow$1 butanediol

How is the metabolism represented?

r_1 :2 pyr $\rightarrow$ 1 acetoL + 1 CO_2
r_2 :1 acetoL $\rightarrow$ 1 diac + 1 CO_2
r_3 :1 acetoL $\rightarrow$ 1 acetoin + 1 CO_2
r_4 :1 diac $\rightarrow$ 1 acetoin
r_5 :1 acetoin $\rightarrow$1 butanediol

Stoichiometry matrix

$$
\begin{array}{c} \\ \text{pyr} \\ \text{acetoL} \\ \text{diac} \\ \text{CO}_2 \\ \text{acetoin} \\ \text{butanediol} \end{array}
\begin{array}{c}
\begin{array}{ccccc} r_1 & r_2 & r_3 & r_4 & r_5 \end{array} \\
\begin{pmatrix}
-2 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 \\
1 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

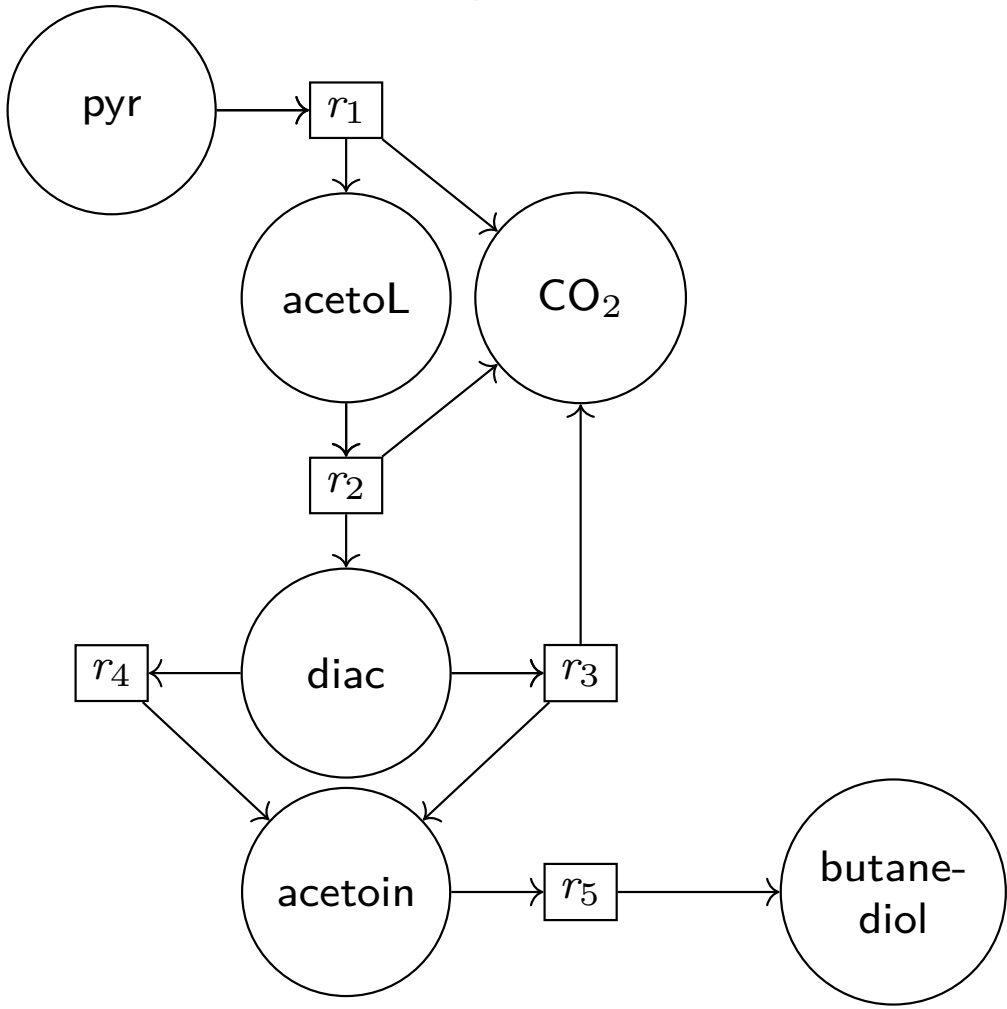
How is the metabolism represented?

r_1 :2 pyr $\rightarrow$ 1 acetoL + 1 CO_2
r_2 :1 acetoL $\rightarrow$ 1 diac + 1 CO_2
r_3 :1 acetoL $\rightarrow$ 1 acetoin + 1 CO_2
r_4 :1 diac $\rightarrow$ 1 acetoin
r_5 :1 acetoin $\rightarrow$1 butanediol

Stoichiometry matrix

	r_1	r_2	r_3	r_4	r_5
pyr	-2	0	0	0	0
acetoL	1	-1	-1	0	0
diac	0	1	0	-1	0
CO_2	1	1	1	0	0
acetoin	0	0	1	1	-1
butanediol	0	0	0	0	1

bipartite graph



Stoichiometry matrix is commonly used for quantitative analysis instead of **graph**, more focused on topology analysis

How is the metabolism reconstructed?

Genome-scale metabolic network (GEMs) reconstruction

Genome-scale metabolic network (GSMNs

Contain metabolic reactions predicted from the entire genomic content through gene-protein-reaction (GPR) relationships (Thiele and Palsson, 2010)

$$r1 : (g1 \wedge g2) \vee (g1 \wedge g3)$$

$$r2 : g1 \wedge (g2 \vee g3)$$

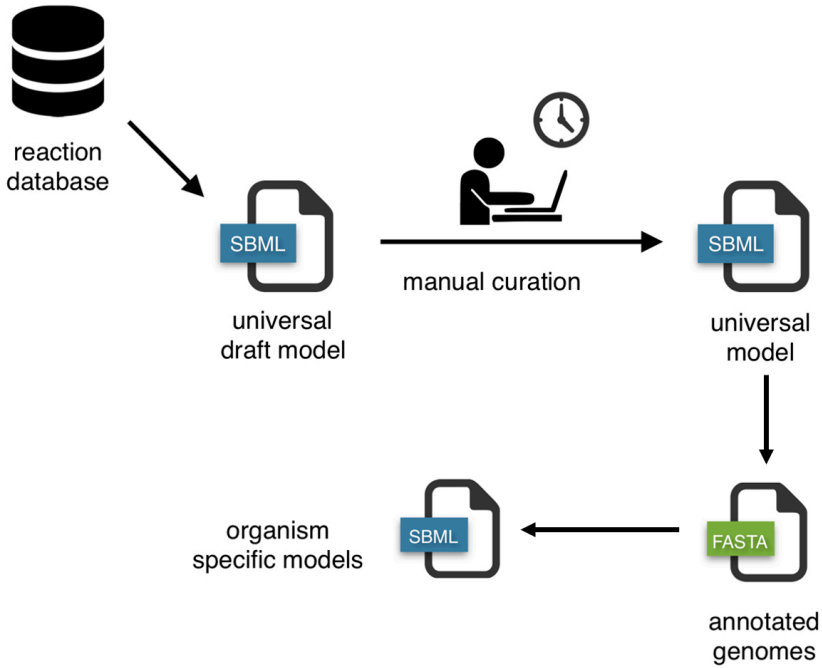


Figure 4: Top down genome-scale metabolic network reconstruction approach (modified from Machado et al., 2018)

- For bacteria: average of 1500 reactions, 1000 genes, 800 metabolites
- Informatic can help to resolve combinatorial problem

Reasoning-based metabolic analysis

Definition

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

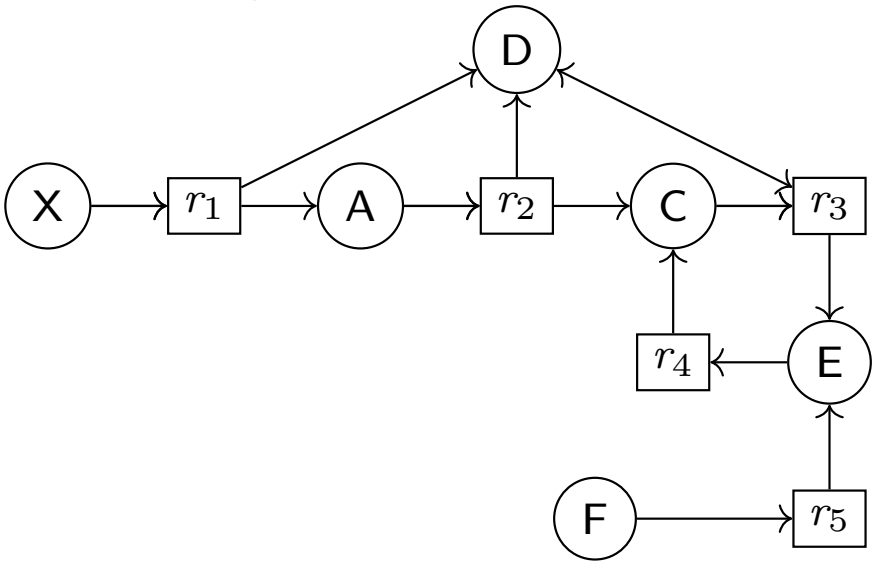
Reasoning-based metabolic analysis

Definition

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



How to compute metabolic capability ?

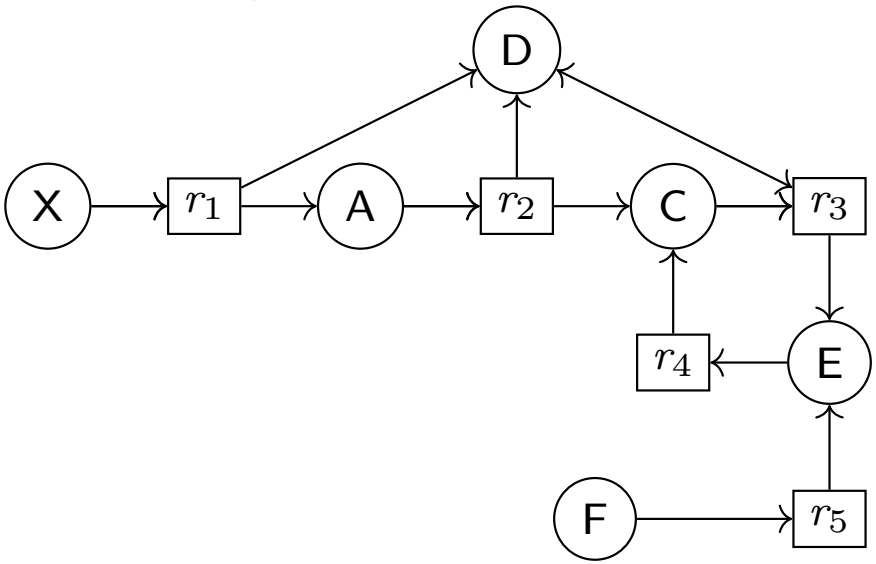
Reasoning-based metabolic analysis

Definition

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

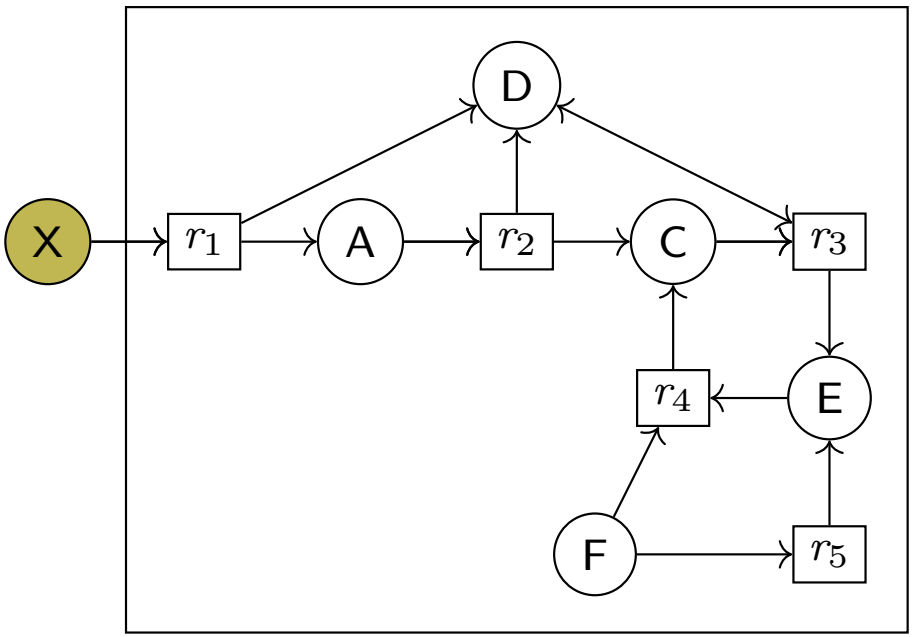
The scope, *i.e.* the metabolic capacity, a network is reached in 2 logical rules (Ebenhöh, Handorf, and Heinrich, 2004)

Reasoning-based metabolic analysis

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



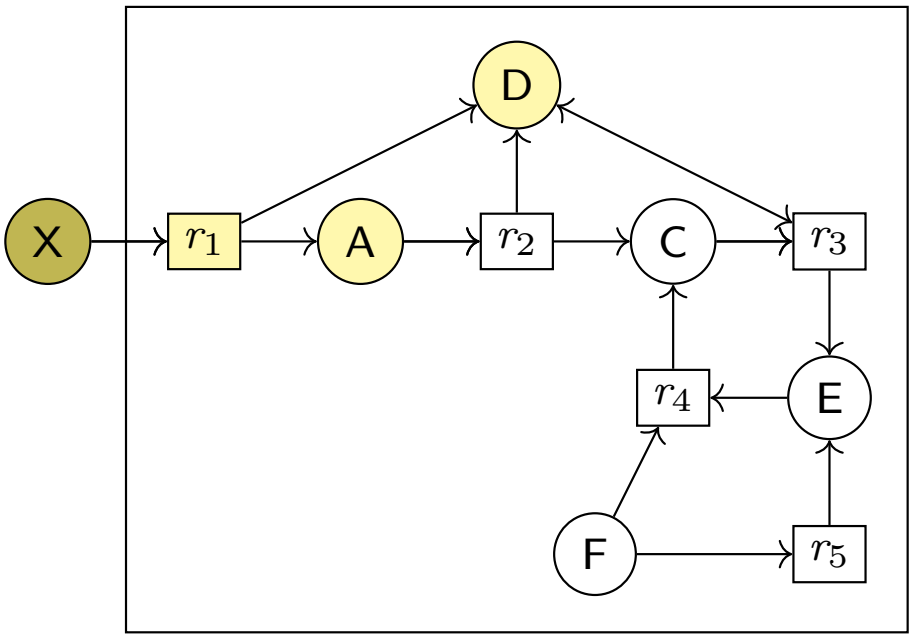
X is a seed

Reasoning-based metabolic analysis

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



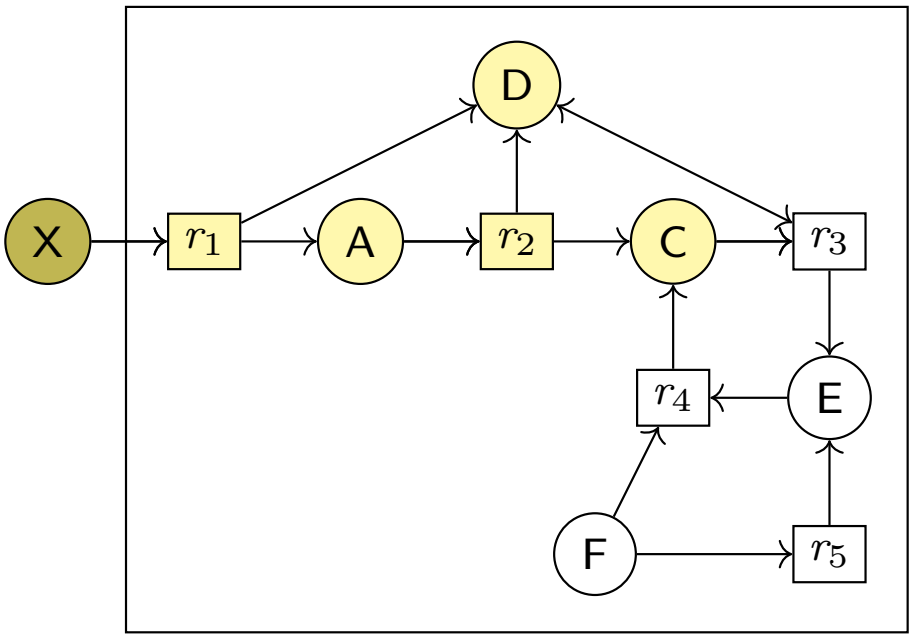
X is available; r_1 is activated and A,D are producible

Reasoning-based metabolic analysis

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



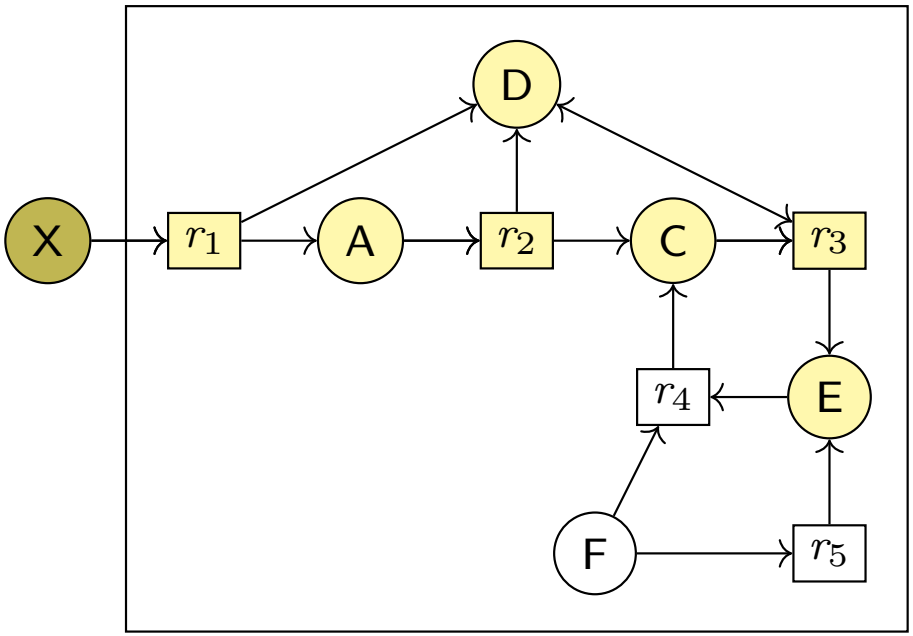
A,D are available; r_2 is activated and C is producible

Reasoning-based metabolic analysis

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



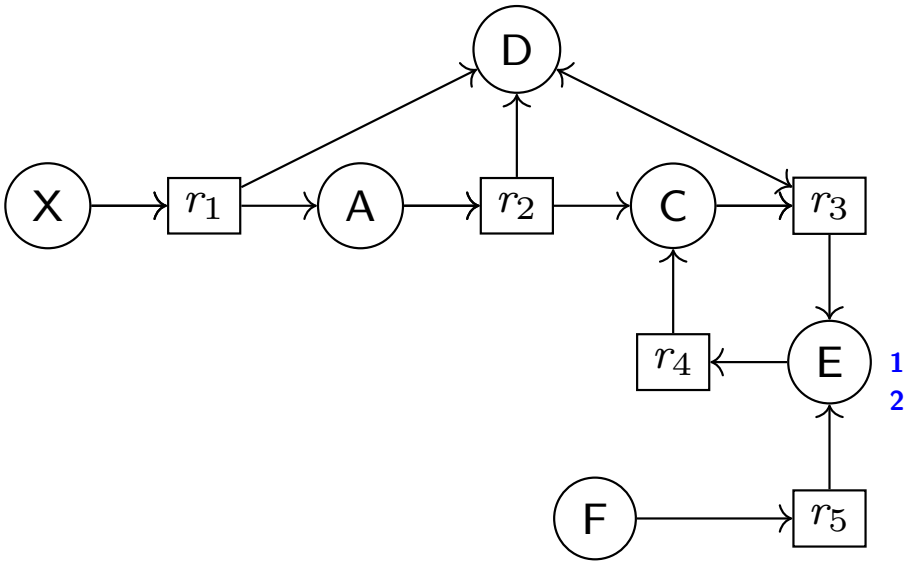
- The potential metabolic capability and topology dependant

Logical rules implementation

Reasoning-based

Allow us to infer qualitative models from logical rules based on biological knowledge

topological-based approaches



- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

```
1 scope(M) :- seed(M).
2 scope(M) :- bacteria(B), product(M,R,B),
       reaction(R,B), scope(M2) : reactant(M2,R,
       B).
```

Logical rules are embedded using Answer Set Programming (**Lifschitz2008**)

Why using Answer Set Programming

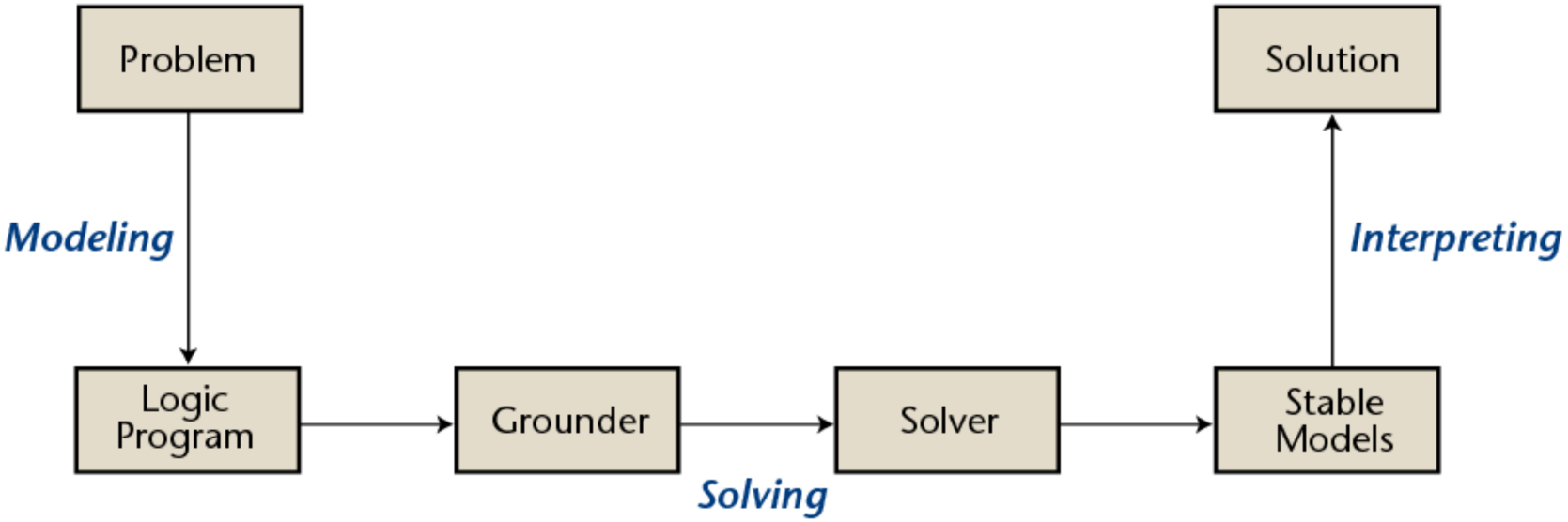


Figure 5: The workflow of Answer set programming (ASP) (Kaufmann et al., 2016)

- Close assumption
- Explainable model
- Solve combinatorial problems

Numerical metabolic model of the metabolism

definition

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic content

Numerical metabolic model of the metabolism

Flux Balance Analysis (Orth, Thiele, and Palsson, 2010)

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic content

Constraint-based approaches

$$\text{maximiser/minimiser } f_{obj}$$
$$\text{tel que } (S.v)_{int} = 0$$
$$\text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}}$$

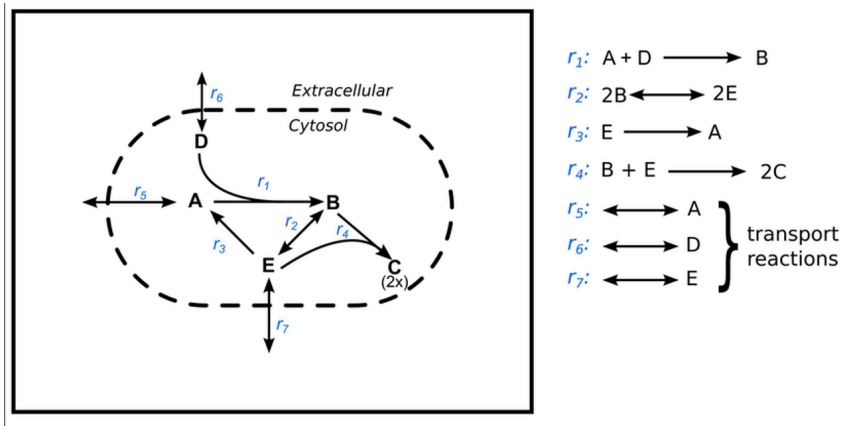


Figure 6: Example of metabolic network

$$\boldsymbol{S} = \begin{array}{c|ccccccc} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ \hline A & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ B & 1 & -2 & 0 & -1 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ D & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ E & 0 & 2 & -1 & -1 & 0 & 0 & 1 \end{array}$$

(Stoichiometric values)

$$\vec{\boldsymbol{v}} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}$$

(Metabolic flux values)

Figure 7: Stoichiometry matrix representation and the flux vector v

Numerical metabolic model of the metabolism

Flux Balance Analysis (Orth, Thiele, and Palsson, 2010)

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic content

Constraint-based approaches

maximiser/minimiser f_{obj}

tel que $(S.v)_{int} = 0$

et $v_{i_{min}} \leq v_i \leq v_{i_{max}}$

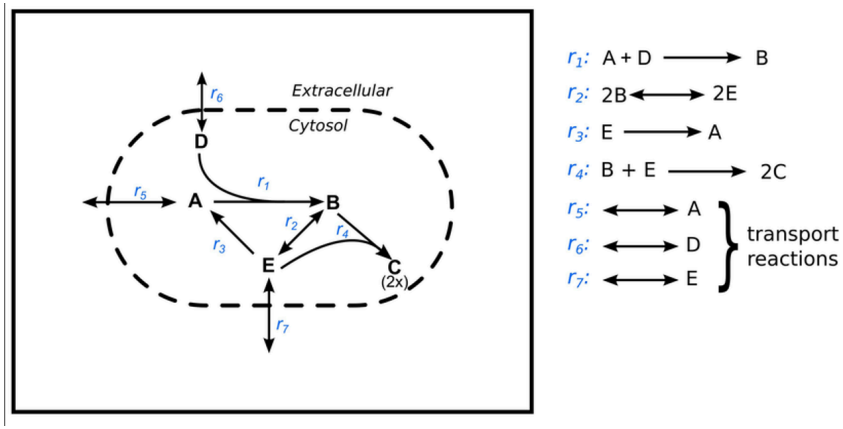


Figure 6: Example of metabolic network

Subject to

$$S\vec{v} = \vec{0} = \begin{cases} \frac{dA}{dt} = -v_1 + v_3 + v_5 \\ \frac{dB}{dt} = v_1 - 2v_2 - v_4 \\ \frac{dC}{dt} = 2v_4 \\ \frac{dD}{dt} = -v_1 + v_6 \\ \frac{dE}{dt} = 2v_2 - v_3 - v_4 + v_7 \end{cases}$$

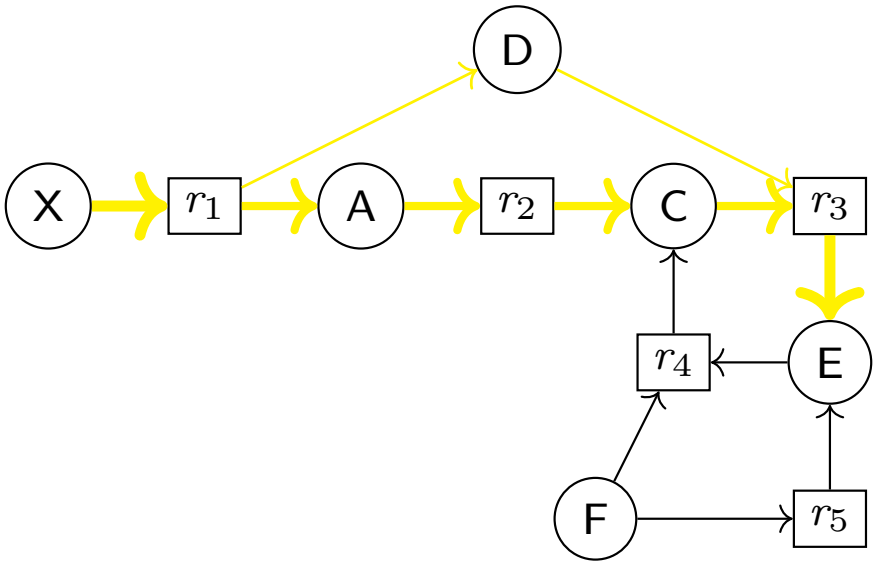
(Steady state system)

$$\begin{array}{l} 0 \leq v_1 < \infty \\ -\infty < v_2 < \infty \\ 0 \leq v_3 < \infty \\ 0 \leq v_4 < \infty \\ 0 \leq v_5 \leq \infty \\ -\infty < v_6 < \infty \\ 0 \leq v_7 \leq \infty \end{array}$$

(Reaction bounds)

Figure 7: Linear programming problem.

Flux application



- Topology and stoichiometry of metabolic goods
- Can explain metabolic observations through reaction fluxes
- difficult to apply in larger scale

Contributions and objective

Objective

Contribute to analyzing metabolic interactions of bacterial communities associated to two use cases: controlled and uncontrolled environment

- Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling (*submitted article*)
  - Highlight pathway of interest
  - Generate mechanistic hypothesis
  - Integrate heterogeneous data
  - Generate robust GSMN
- Reasoning-based metabolic modelling of cooperation and competition potentials in large-scale microbial communities (*in preparation*)
  - Generate mechanistic hypothesis
  - Bacterial interaction rules
  - Generic approach
- Enrichment of discrete models
  - Highlight pathway
  - Generate mechanistic hypothesis
  - Integrate heterogeneous constrains
  - Generate logical rules

blablabla

Belcour, Arnaud et al. (2020). "Metage2metabo, microbiota-scale metabolic complementarity for the identi1cation of key species". In: *eLife*. ISSN: 2050084X. DOI: 10.7554/eLife.61968.

Belkaid, Yasmine and Timothy W. Hand (2014). "Role of the Microbiota in Immunity and Inflammation". In: *Cell* 157.1, pp. 121–141. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2014.03.011. URL: https://www.sciencedirect.com/science/article/pii/S0092867414003456.

Ebenhöh, Oliver, Thomas Handorf, and Reinhart Heinrich (2004). "Structural analysis of expanding metabolic networks.". In: *Genome informatics. International Conference on Genome Informatics* 15.1, pp. 35–45. ISSN: 09199454.

Faust, Karoline and Jeroen Raes (2012). "Microbial interactions: From networks to models". In: *Nature Reviews Microbiology* 10.8, pp. 538–550. ISSN: 17401526. DOI: 10.1038/nrmicro2832. URL: http://dx.doi.org/10.1038/nrmicro2832.

Frioux, Clémence et al. (2018). "Scalable and exhaustive screening of metabolic functions carried out by microbial consortia". In: *Bioinformatics* 34.17, pp. i934–i943. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty588. URL: https://doi.org/10.1093/bioinformatics/bty588.

Hoorman, James J (2011). "The Role of Soil Bacteria". In: *The Ohio State University Extension*, pp. 1–4.

Kaufmann, Benjamin et al. (2016). "Grounding and Solving in Answer Set Programming". In: *AI Mag.* 37, pp. 25–32. URL: https://api.semanticscholar.org/CorpusID:30782765.

 Kitano, Hiroaki (2002). "Systems biology: A brief overview". In: *Science* 295.5560, pp. 1662–1664. ISSN: 00368075. DOI: 10.1126/science.1069492.

 Machado, Daniel et al. (2018). "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities". In: *Nucleic Acids Research*. ISSN: 13624962. DOI: 10.1093/nar/gky537.

 McSweeney, Paul L.H. and Maria José Sousa (2000). "Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review". In: *Lait* 80.3, pp. 293–324. ISSN: 00237302. DOI: 10.1051/lait:2000127.

 Orth, Jeffrey D., Ines Thiele, and Bernhard O. Palsson (2010). *What is flux balance analysis?* DOI: 10.1038/nbt.1614.

 Royet, Jean-Pierre and Jane Plailly (Oct. 2004). "Lateralization of Olfactory Processes". In: *Chemical Senses* 29.8, pp. 731–745. ISSN: 0379-864X. DOI: 10.1093/chemse/bjh067. eprint: https://academic.oup.com/chemse/article-pdf/29/8/731/930410/bjh067.pdf. URL: https://doi.org/10.1093/chemse/bjh067.

 Suomalainen, By Heikki and Matti Lehtonen (1978). "JJu It". In: 85, pp. 149–156.

 Sánchez López de Nava A, Raja A (2023). *Physiology, Metabolism*. URL: www.ncbi.nlm.nih.gov/books/NBK546690/.

 Thiele, Ines and Bernhard O Palsson (2010). "A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction". In: *Nature Protocols* 5.1, pp. 93–121. ISSN: 1754-2189. DOI: 10.1038/nprot.2009.203.

 Zhang, Yu Jie et al. (2015). "Impacts of gut bacteria on human health and diseases". In: *International Journal of Molecular Sciences* 16.4, pp. 7493–7519. ISSN: 14220067. DOI: 10.3390/ijms16047493.