Modélisation métabolique

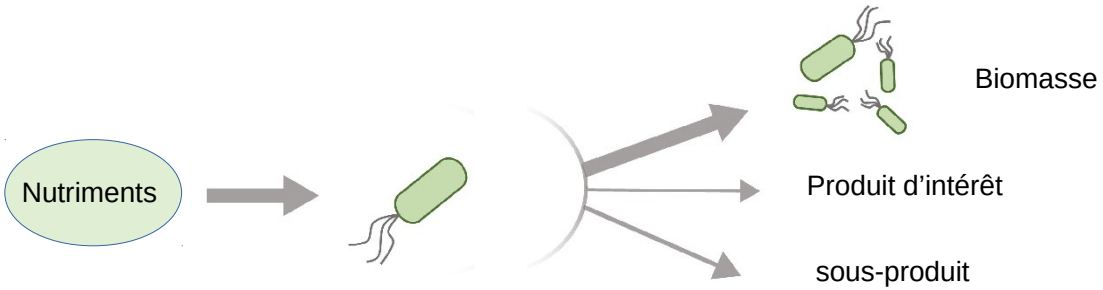
Sabine Peres

sabine.peres@univ-lyon1.fr

Biometry and Evolutionary Biology
University Claude Bernard Lyon 1 - France

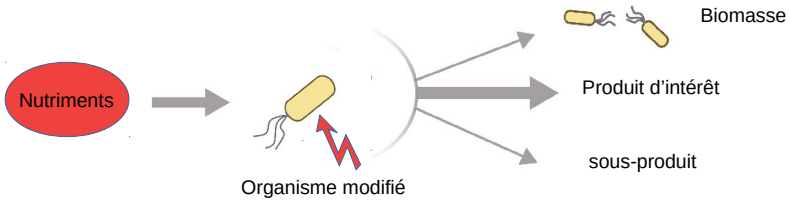
Developp control strategies for living organisms

1. Concevoir de **nouvelles approches fondamentales de modélisation** pour décrypter le fonctionnement des organismes.

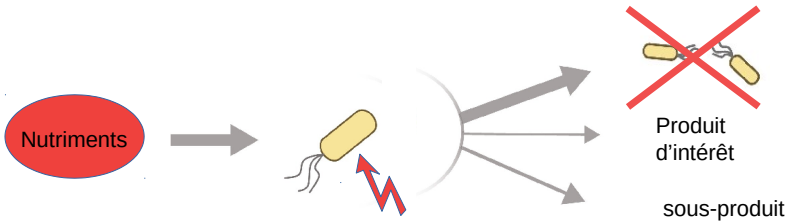


Mahout, *et al. Processes. 2020*

2. Développer des **approches prédictives** des systèmes biologiques pour proposer des **cibles** thérapeutiques, en agroalimentaire, etc.



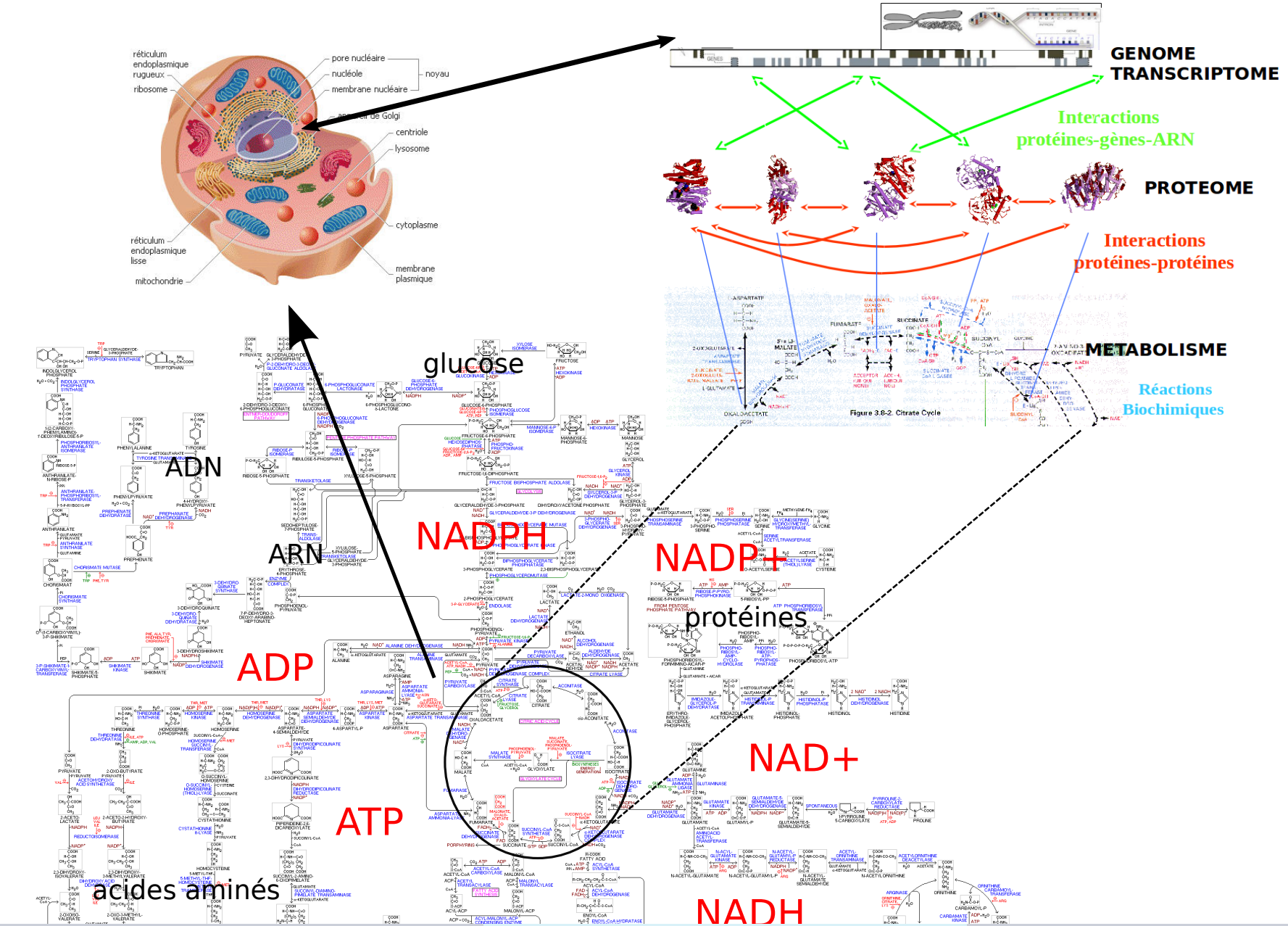
Da Veiga Moreira, *et al. Sci. Rep. 2021*



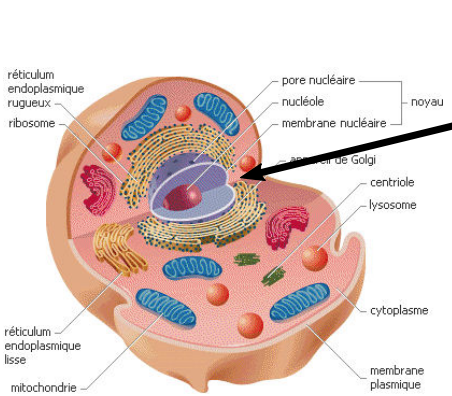
Da Veiga Moreira, *et al. Sci. Rep. 2019*

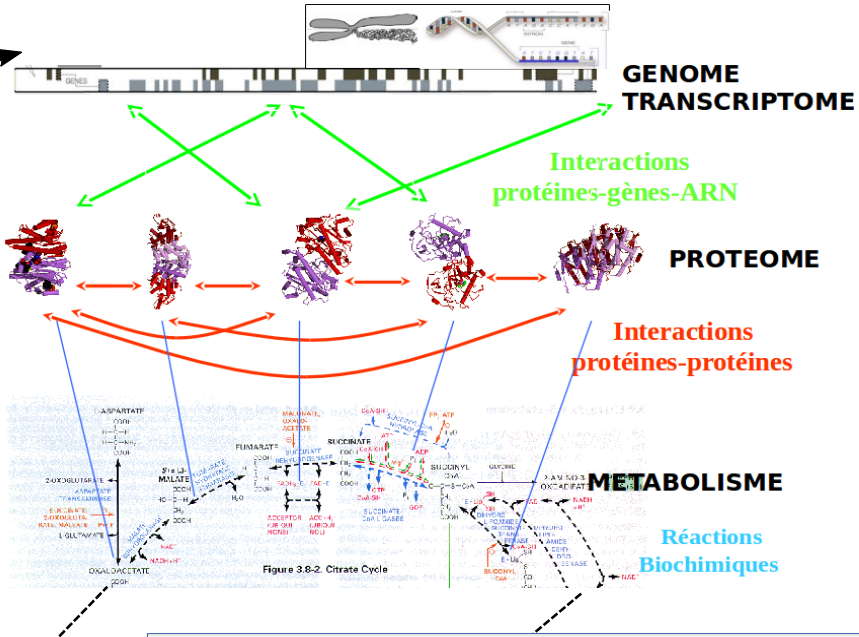
3. Comprendre la stratégie des organismes face à des **modifications génétiques ou environnementales**.

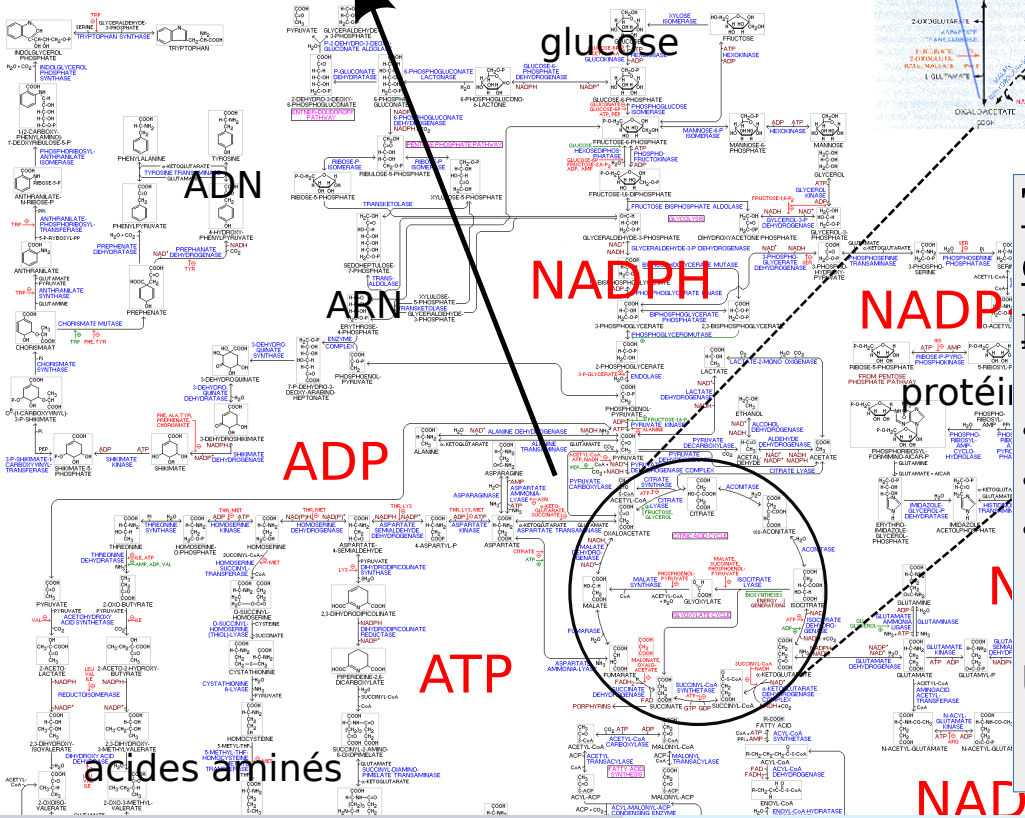
Metabolism: Biochemical reactions producing the energy and molecules of a cell necessary for its survival



Metabolism: Biochemical reactions producing the energy and molecules of a cell necessary for its survival







<u>The control of metabolism is a major challenge for biology because it allows the production of molecules of interest:</u>

- pharmaceuticals to produce therapeutic molecules
- health to repair a pathology
- food to increase the yield of a molecule

Understand the capabilities of an organism
⇒ analyze its metabolic pathways

Thèmes abordés

- Introduction métabolisme
- Formalisation des réseaux métaboliques
- Modélisation à base de contraintes des réseaux métaboliques : Flux Balance Analysis (FBA), dynamic FBA (dFBA)
- Modes élémentaires de flux (EFM)
- Minimal Cut Sets (MCS)
- Ajout de contraintes thermodynamique, cinétique
- (Modélisation dynamique des réseaux métaboliques)

On identifiera différents régimes de fonctionnement, les voies optimales et les cibles pour optimiser/empêcher la production de molécules. On prédira ensuite les comportements face à des perturbations.

Outils

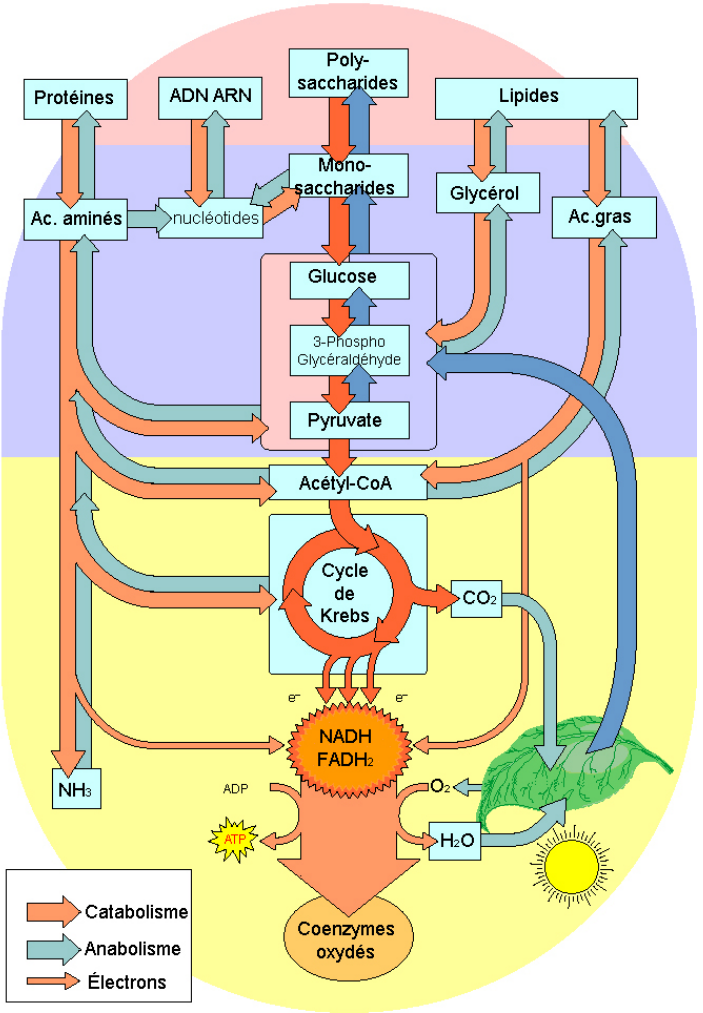
- Python ($\geq$ 3.8)
- Conda / Jupyter Notebook
- CobraPy
- cplex
- Copasi
- Escher map
- Cobamp / ASPefm

Exploiter et transformer les molécules de leur environnement

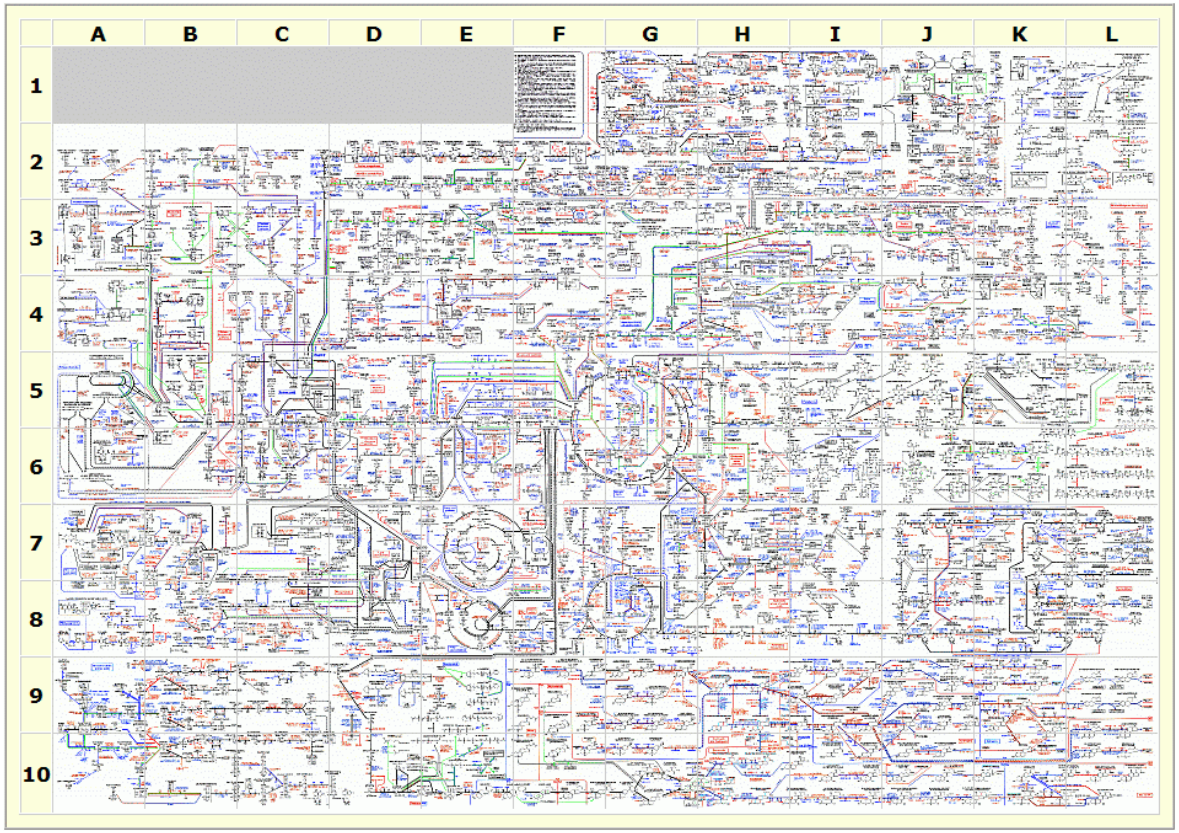
Les cellules des organismes vivant consacrent une grande partie de leurs activités à **exploiter et à transformer** les molécules de leur environnement pour en retirer de l'énergie et créer les molécules qui serviront à leur propre construction.

Catabolisme : dégradation des molécule pour produire de l'énergie

Anabolisme : synthèse de molécules (muscle, tissue, lipide...)

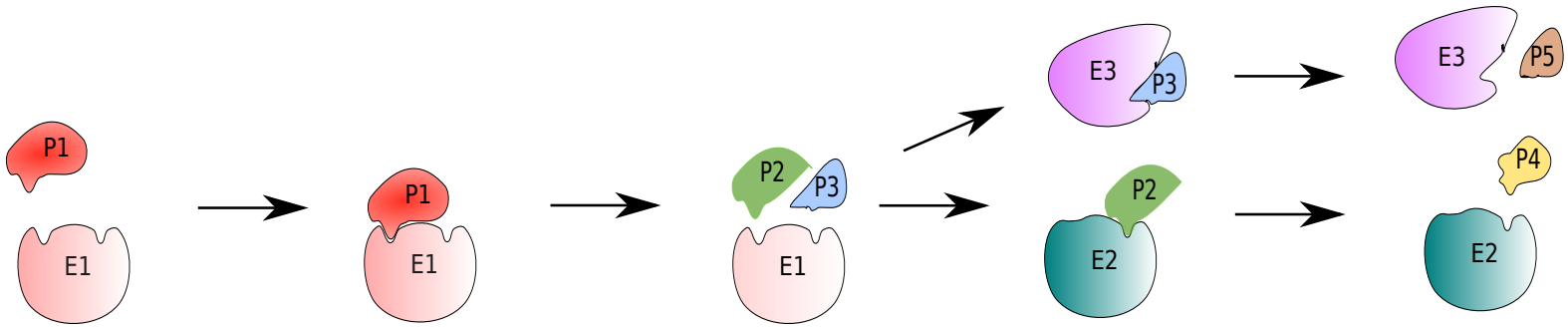


Le métabolisme: la chimie du vivant



Ensemble des réactions biochimiques qui se déroulent au sein d'une cellule ou d'un organisme produisant de l'énergie et des molécules nécessaires à son bon fonctionnement.

Voies de synthèse de produits : réseaux métaboliques

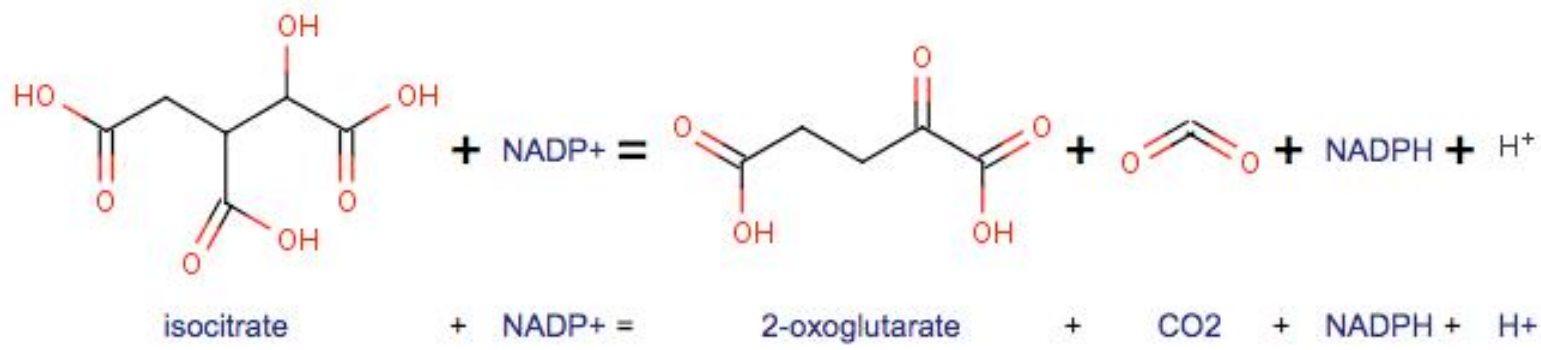


Enchaînements de réactions qui transforment étape par étape les métabolites, formant des chemins de conversion dans le métabolisme.

Les acteurs du métabolisme

- **Métabolites**: molécules impliquées dans le métabolisme cellulaire. En grande majorité, elles sont organiques, composées de carbone, d'hydrogène, d'oxygène, azote, de phosphore et de soufre.
- **Réactions**: transformation chimique des métabolites.
- **Enzymes**: protéines jouant le rôle de catalyseurs sans lesquels la plupart des réactions métaboliques ne pourraient se dérouler à des vitesses compatibles avec la vie de la cellule.
- Cinétique des réactions métaboliques
- Contrôle des réactions métaboliques
- Aspects thermodynamiques

Réactions

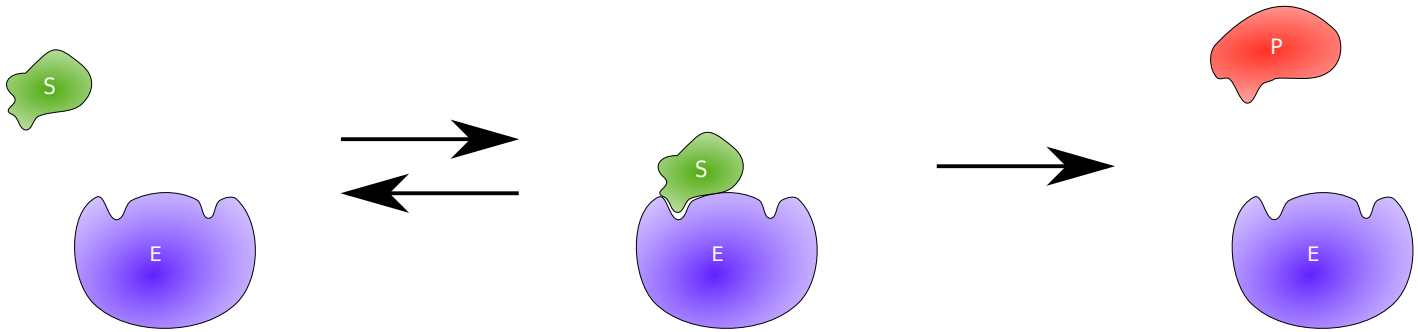


- Des métabolites **substrats** réagissent entre eux pour donner des métabolites **produits**. On représente généralement la réaction par son équation bilan, laquelle met en évidence la **stœchiométrie** de la réaction, c'est-à-dire les proportions dans lesquelles les métabolites sont consommés et produits.

Réactions enzymatiques

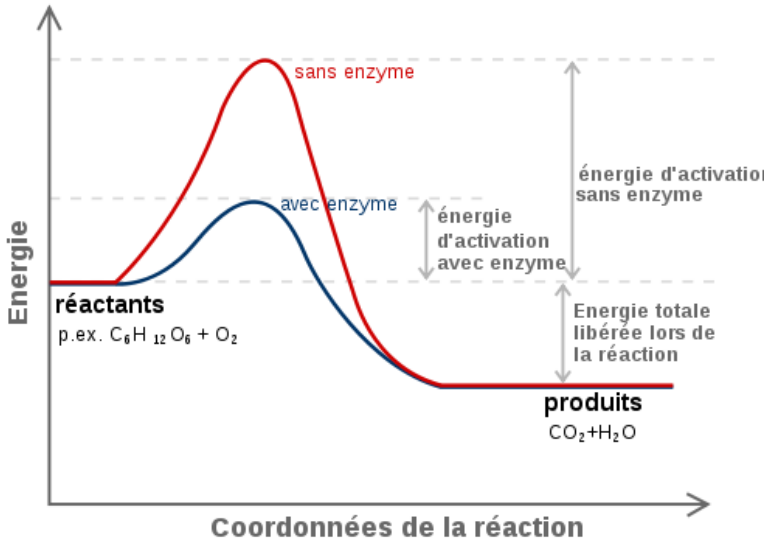
$$E + S \longrightarrow ES \longrightarrow E + P$$

- Fixation d'un enzyme sur un (ou plusiers) substrat(s) avec formation du complexe enzyme-substrat(s) suivie de la formation d'un produit et relarguage de l'enzyme libre.



Enzymes

- Les enzymes favorisent spécifiquement une réaction particulière, agissent à faible concentration et augmentent considérablement la vitesse des réactions.

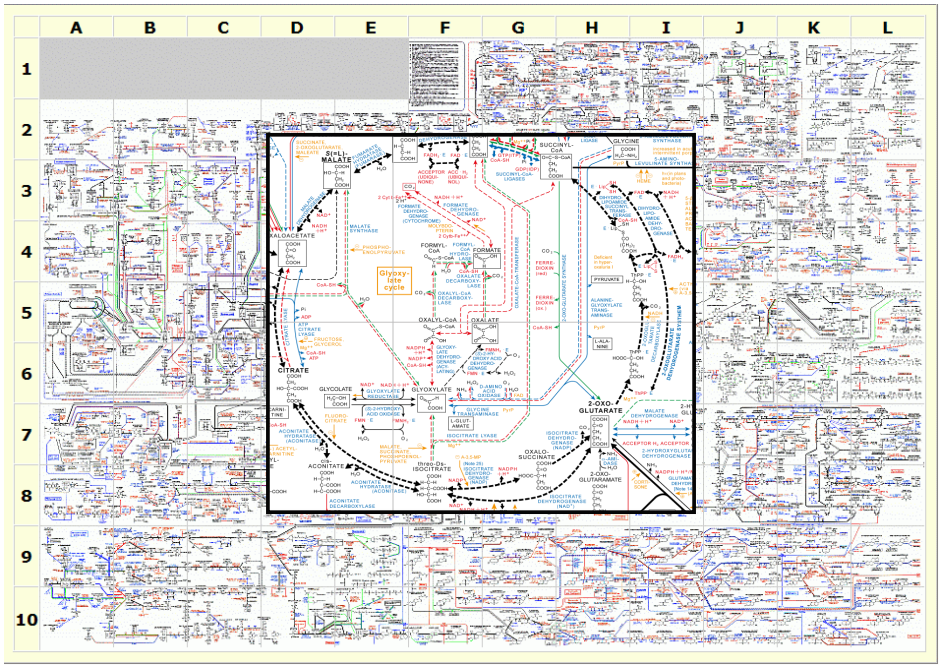


- Ils se retrouvent intactes en fin de réaction.
- Le double aspect spécificité et accélération de la catalyse enzymatique donne à l'organisme le contrôle des transformations métaboliques se déroulant dans la cellule.

Cinétique des réactions métaboliques

- Une bonne grandeur pour décrire le fonctionnement du métabolisme est la vitesse des réactions métaboliques, également appelée **flux**.
- Les flux des réactions renseignent directement sur les conversions métaboliques ayant lieu dans la cellule.
- La vitesse d'une réaction enzymatique dépend de nombreux facteurs : concentration des substrats et produits, concentration de l'enzyme, efficacité catalytique de l'enzyme, température, pH, pression, etc.

Le métabolisme est un système complexe

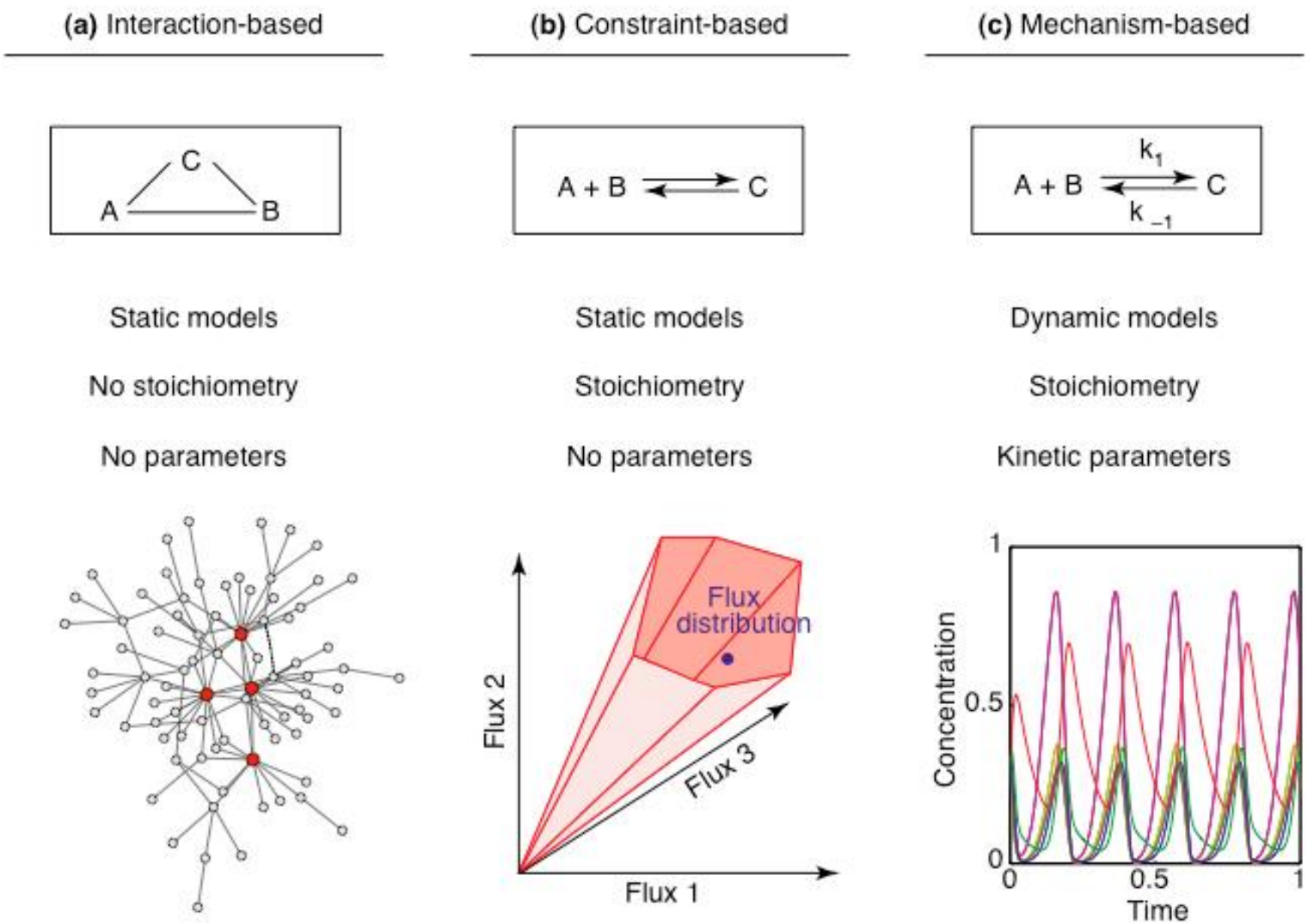


?

Il est impossible de prévoir son comportement par la seule connaissance de ses composants élémentaires.

⇒ **Modéliser pour Analyser et Modifier son comportement**

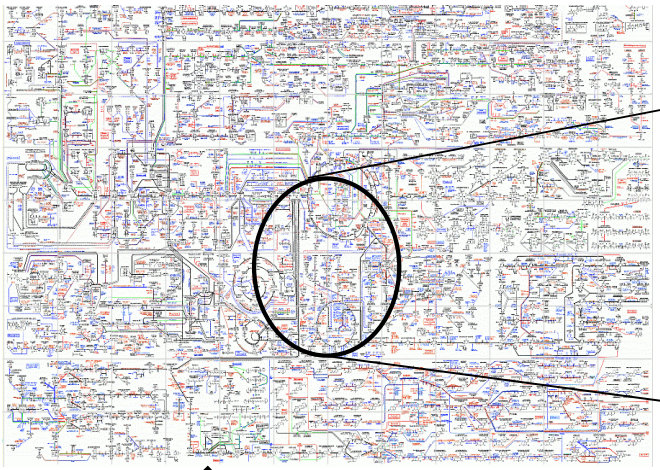
Il est important de choisir un type de modélisation adapté aux questions biologiques posées !



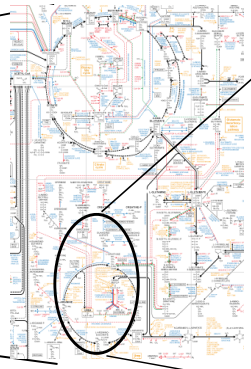
Stelling, Curr Opin Microbiol. 2004 Oct;7(5):513-518

Analyse des voies métaboliques

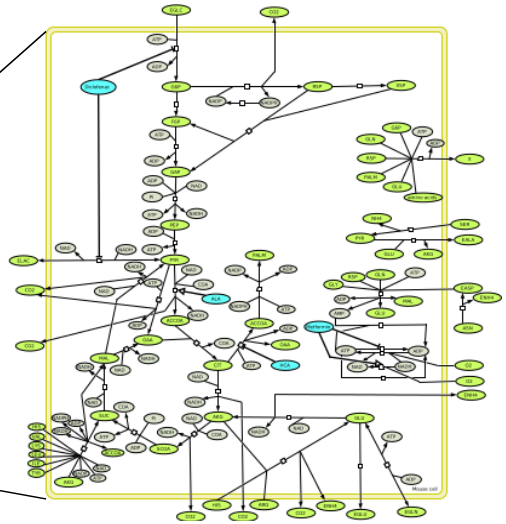
Genome-scale metabolic model



Medium-scale metabolic model



Small-scale metabolic model



Steady state analysis

Constraints-based modelling

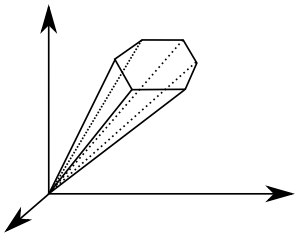
reactions

metabolites

2	-2	-1	0	0	0	0	0
1	-1	0	-1	1	0	0	0
0	0	-1	2	-2	0	0	0
0	0	1	0	0	-1	2	0
-2	2	1	0	0	0	0	0
-1	1	0	1	-1	0	0	0
0	0	1	-2	2	0	0	0
0	0	-1	0	0	1	-2	0

Sint · v1 v2 . . . **V** = 0

Solution space

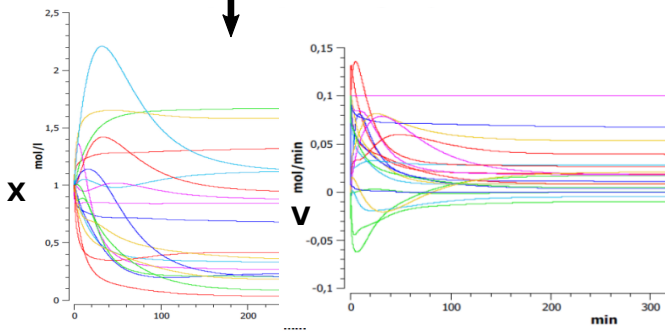


Kinetics modelling

Ordinary differential equations

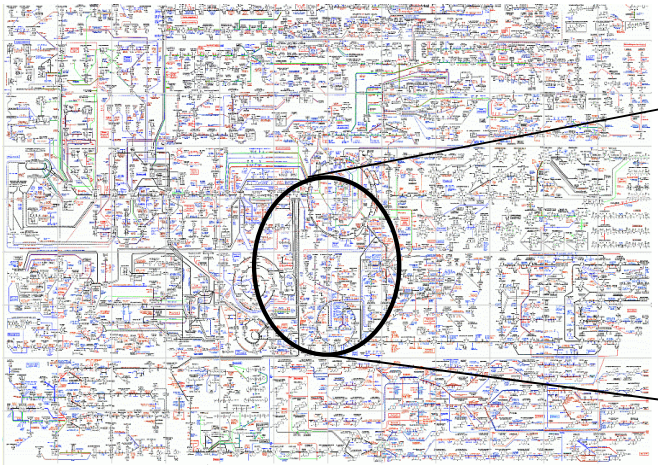
$$\frac{dX(t)}{dt} = \begin{cases} S_{int}.V(X(t)) - \mu X_{int} \\ S_{ext}.V(X(t)).\mu \\ \mu.biomass \end{cases}$$

S : stoichiometric matrix
V : vector of reactions rates (i.e Flux)
X : vector of metabolites
μ : growth rate

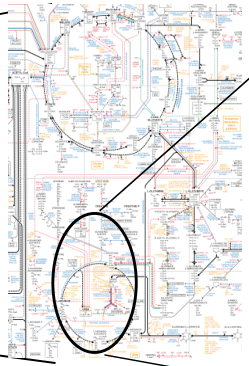


Analyse des voies métaboliques

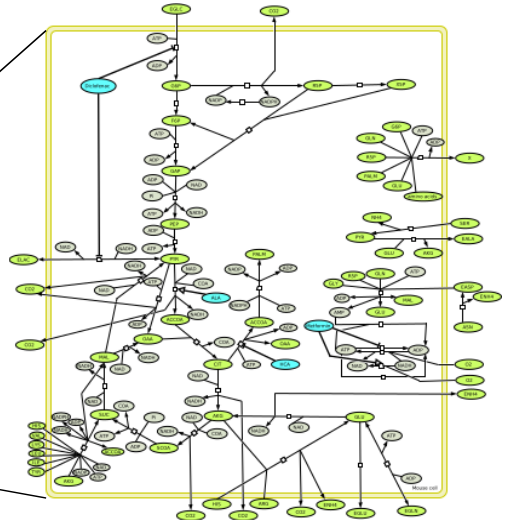
Genome-scale metabolic model



Medium-scale metabolic model



Small-scale metabolic model



Steady state analysis
Constraints-based modelling

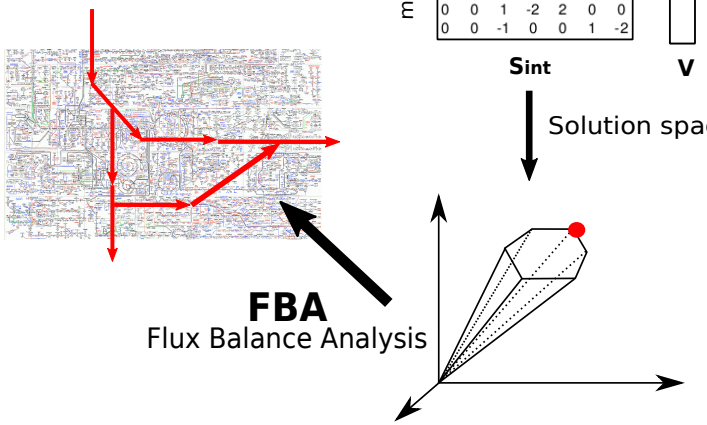
reactions

metabolites

2	-2	-1	0	0	0	0	v1
1	-1	0	-1	1	0	0	v2
0	0	-1	2	-2	0	0	.
0	0	1	0	0	-1	2	.
-2	2	1	0	0	0	0	.
-1	1	0	1	-1	0	0	
0	0	1	-2	2	0	0	
0	0	-1	0	0	1	-2	

$\mathbf{S_{int}} \cdot \mathbf{V} = 0$

Solution space

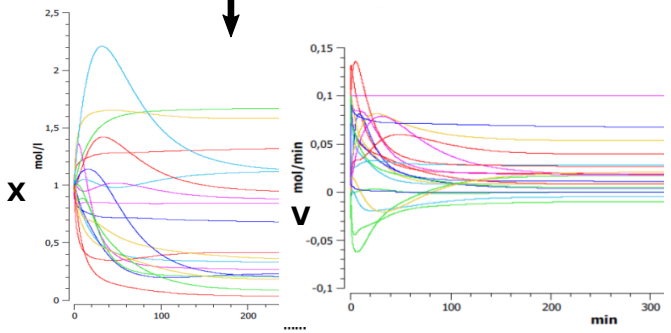


FBA
Flux Balance Analysis

Kinetics modelling
Ordinary differential equations

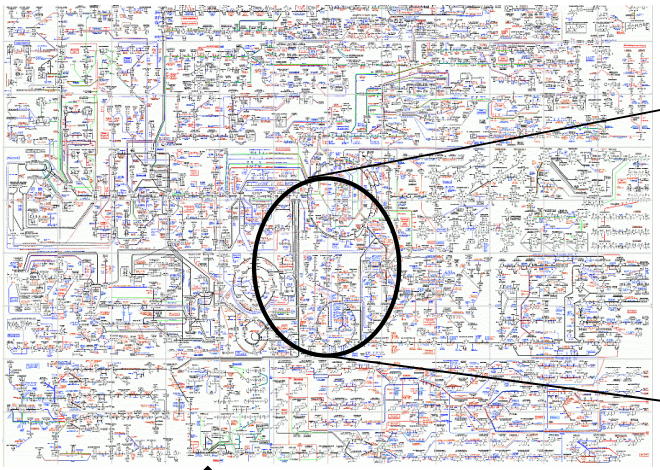
$$\frac{dX(t)}{dt} = \begin{cases} S_{int}.V(X(t)) - \mu X_{int} \\ S_{ext}.V(X(t)).\mu \\ \mu.biomass \end{cases}$$

S : stoichiometric matrix
V : vector of reactions rates (i.e Flux)
X : vector of metabolites
μ : growth rate

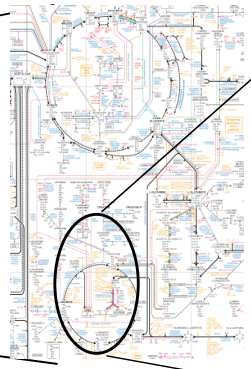


Analyse des voies métaboliques

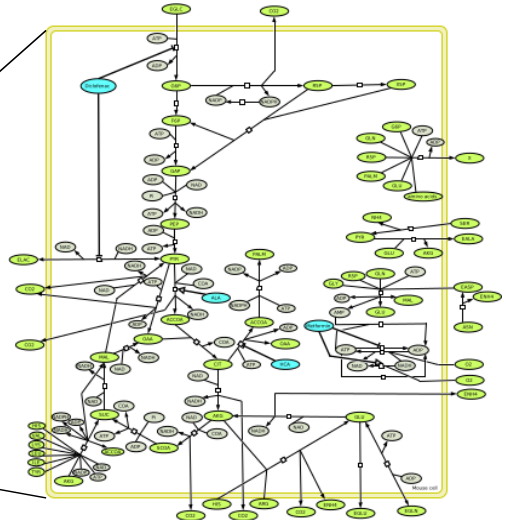
Genome-scale metabolic model



Medium-scale metabolic model



Small-scale metabolic model



Steady state analysis
Constraints-based modelling

reactions

metabolites

2	-2	-1	0	0	0	0
1	-1	0	-1	1	0	0
0	0	-1	2	-2	0	0
0	0	1	0	0	-1	2
-2	2	1	0	0	0	0
-1	1	0	1	-1	0	0
0	0	1	-2	2	0	0
0	0	-1	0	0	1	-2

Sint

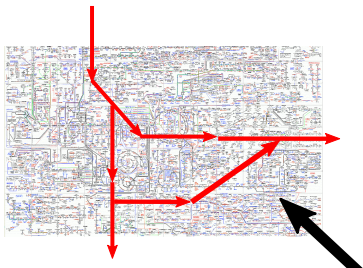
$\cdot$ v1 v2 . . . v

$= 0$

Kinetics modelling
Ordinary differential equations

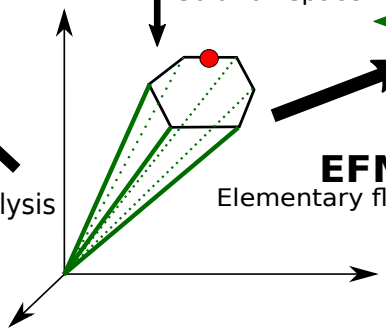
$$\frac{dX(t)}{dt} = \begin{cases} S_{int}.V(X(t)) - \mu X_{int} \\ S_{ext}.V(X(t)).\mu \\ \mu.biomass \end{cases}$$

S : stoichiometric matrix
V : vector of reactions rates (i.e Flux)
X : vector of metabolites
μ : growth rate

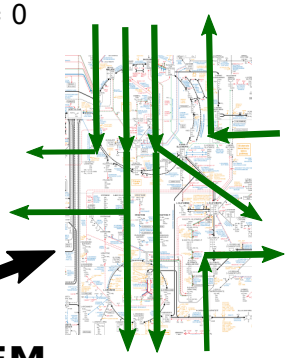


FBA
Flux Balance Analysis

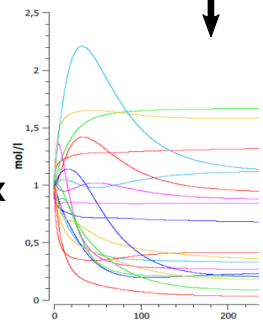
Solution space



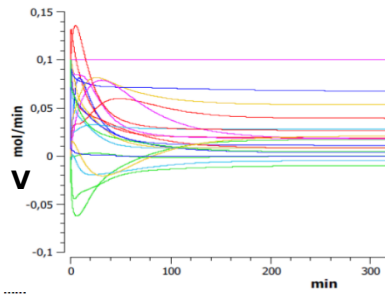
EFM
Elementary flux modes



X



V



Exchange format

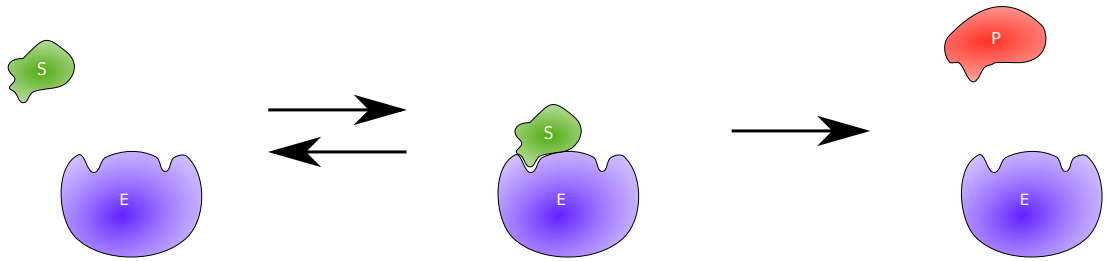
SBML, Systems Biology Marke-up Language:

- a representation format based on XML
- reaction focused and designed for modelling
- list of compartments, metabolites, reactions. . .
- Database: BIGG, Metabolicatlas, Biomodels, ...

Formalisation des réactions enzymatiques

- Fixation d'un enzyme sur un (ou plusieurs) substrat(s) avec formation du complexe enzyme-substrat(s) suivie de la formation d'un produit et relarguage de l'enzyme libre.

$$E + S \longrightarrow ES \longrightarrow E + P$$

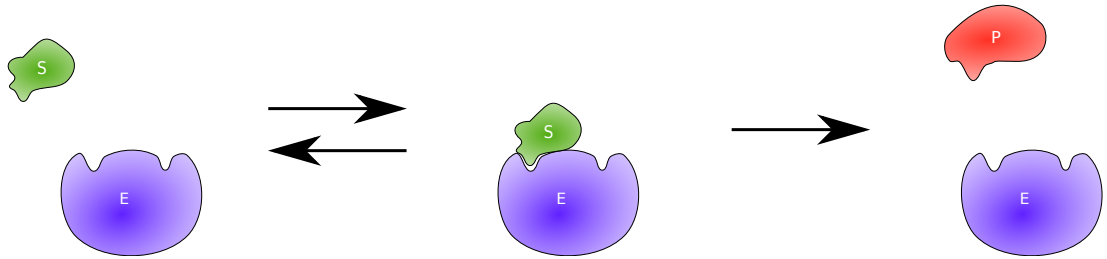


$$E : S \longrightarrow P$$

Formalisation des réactions enzymatiques

- Fixation d'un enzyme sur un (ou plusieurs) substrat(s) avec formation du complexe enzyme-substrat(s) suivie de la formation d'un produit et relarguage de l'enzyme libre.

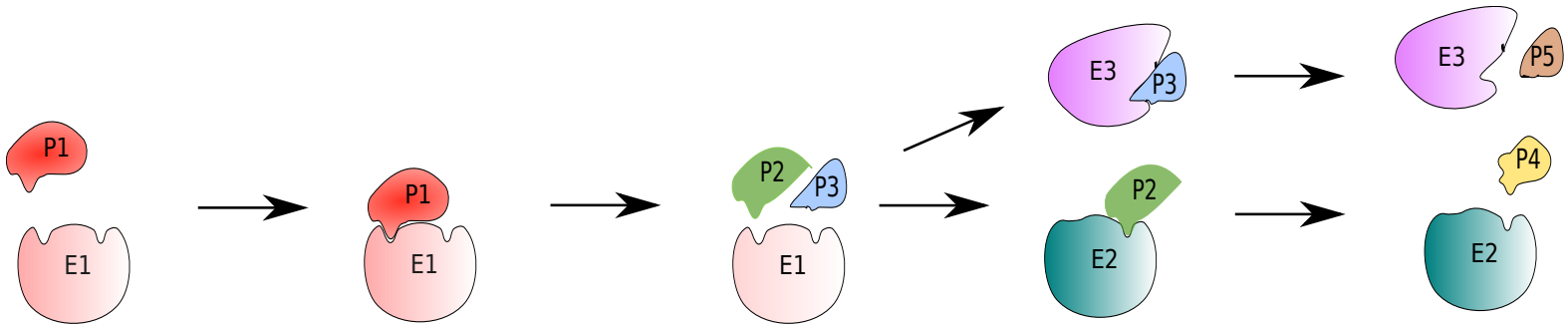
$$E + S \longrightarrow ES \longrightarrow E + P$$



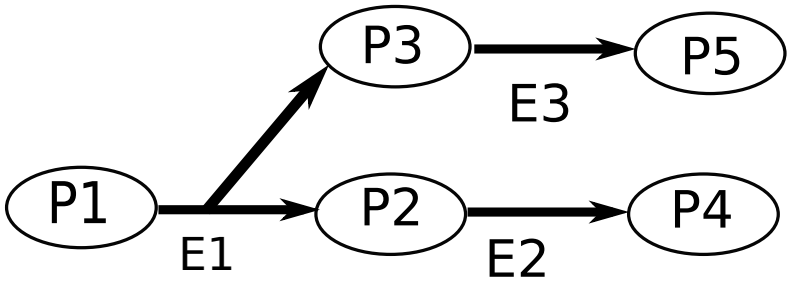
$$E : S \longrightarrow P$$

$$E : \sum_{i=1}^{nbSubtrats} s_i S_i \longrightarrow \sum_{i=1}^{nbProducts} p_i P_i$$

Représentation : hypergraphe orienté



Nœuds: métabolites
Arêtes: réactions

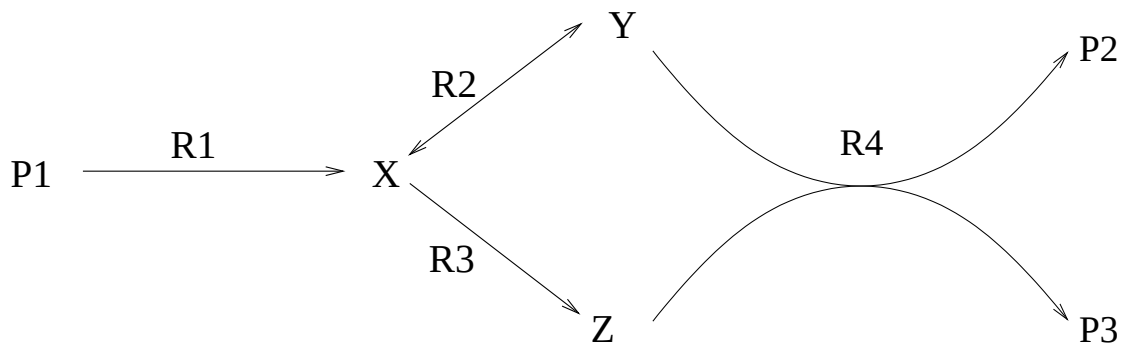


Exemple

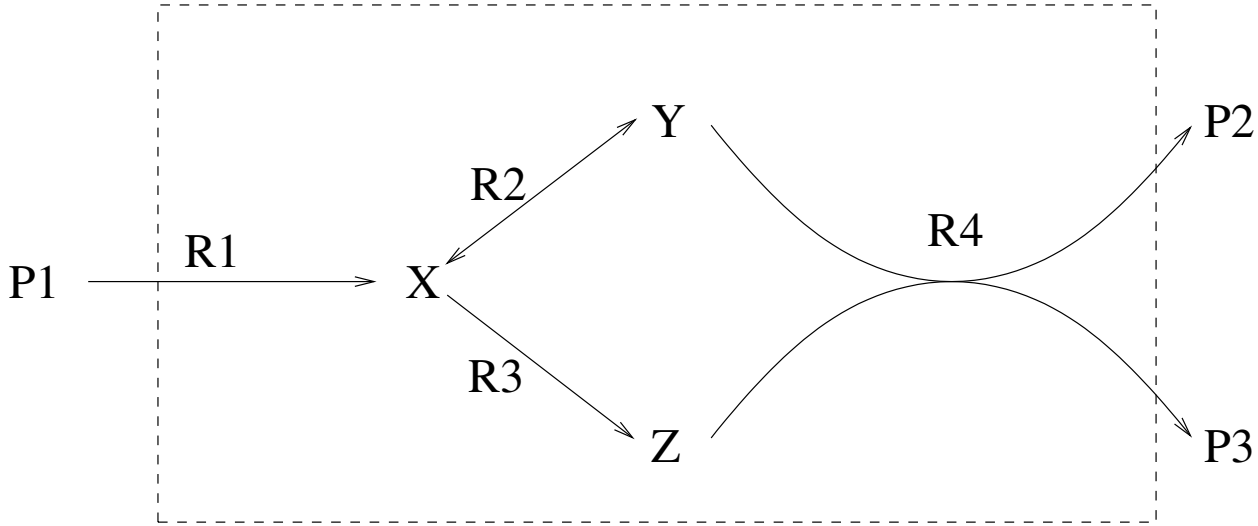
R1 : $P_1 \rightarrow X$
R2 : $X \leftrightarrow Y$
R3 : $X \rightarrow Z$
R4 : $Y + Z \rightarrow P_2 + P_3$

Exemple

R1 : $P_1 \rightarrow X$
R2 : $X \leftrightarrow Y$
R3 : $X \rightarrow Z$
R4 : $Y + Z \rightarrow P_2 + P_3$



Bornes du système



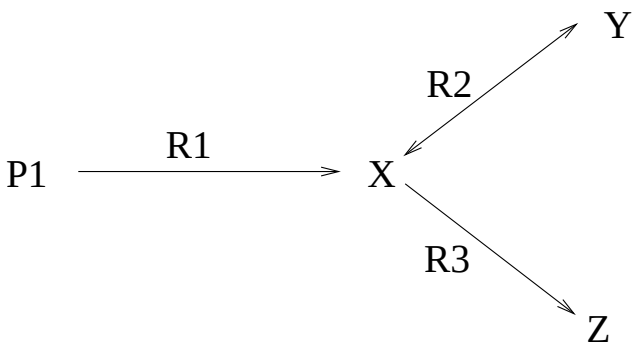
- Métabolites externes : P1, P2, P3
- Métabolites internes : X, Y, Z

Matrice de stœchiométrie

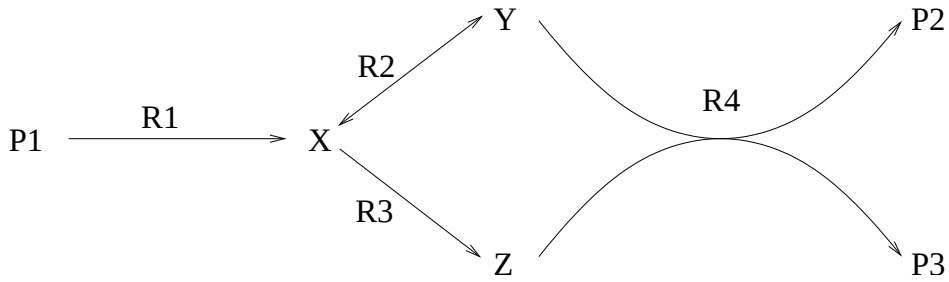
- Tous les réseaux métaboliques avec m métabolites et r réactions peuvent être représentés par une matrice de stœchiométrie N de m lignes et r colonnes telle que :

$$N_{ij} = \begin{cases} a & \text{si la réaction } j \text{ produit } a \text{ molécules de } i. \\ -a & \text{si la réaction } j \text{ consomme } a \text{ molécules de } i. \\ 0 & \text{sinon} \end{cases}$$

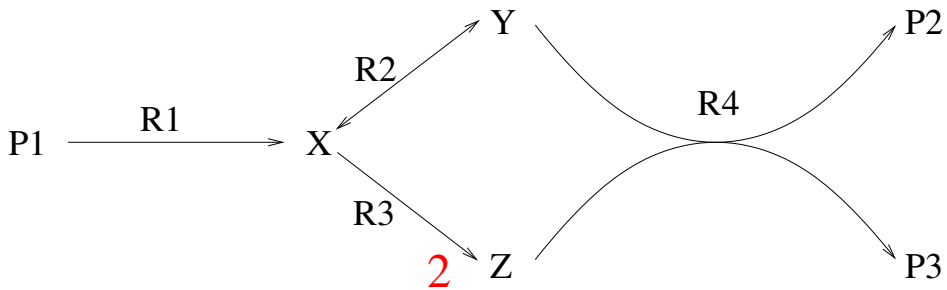
- Exemple:



$$N = \begin{matrix} R1 & R2 & R3 \\ (+1 & -1 & -1) \end{matrix} \quad X$$

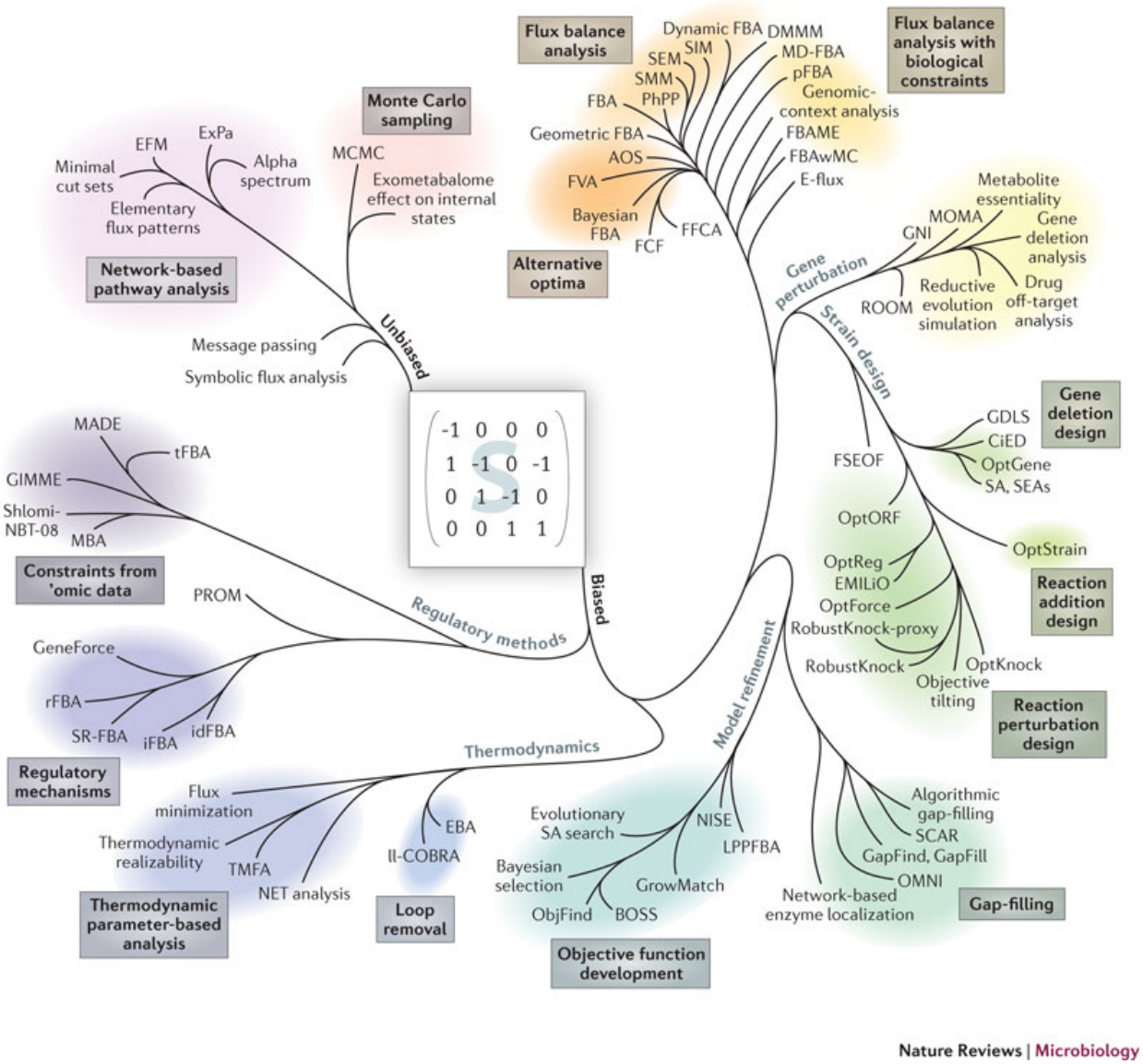


$$N = \begin{array}{c} \begin{array}{cccc} R1 & R2 & R3 & R4 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array} \begin{array}{c} X \\ Y \\ Z \end{array}$$



$$N = \begin{array}{c} \begin{array}{cccc} R1 & R2 & R3 & R4 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 2 & -1 \end{pmatrix} \end{array} \begin{array}{c} X \\ Y \\ Z \end{array}$$

Modelling method based on metabolic network structure



Lewis *et al.* Nat Rev Microbiol. 2012

Dynamique d'un réseau métabolique

- Description des transformations biochimiques du réseau.

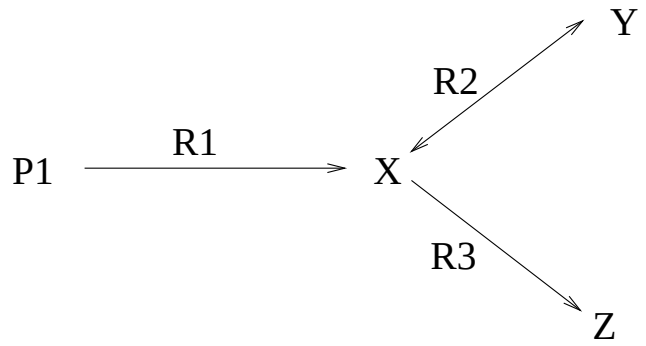
$$\frac{dX(t)}{dt} = N.V(X(t),P) \text{ avec } \begin{cases} N & : \text{matrice de stœchiométrie} \\ V & : \text{vecteur de réactions} \\ X & : \text{vecteur de métabolites internes} \\ P & : \text{vecteur de paramètres}(k_m, v_m, pH, etc.) \end{cases}$$

- Hypothèse de l'état stationnaire : la concentration de chaque métabolite est quasi-constante :

$$\frac{dX(t)}{dt} = 0 \Rightarrow N.V(X(t),P) = 0.$$

- **L'espace de solution est dans le noyau de N** ($\ker N$).

Exemple



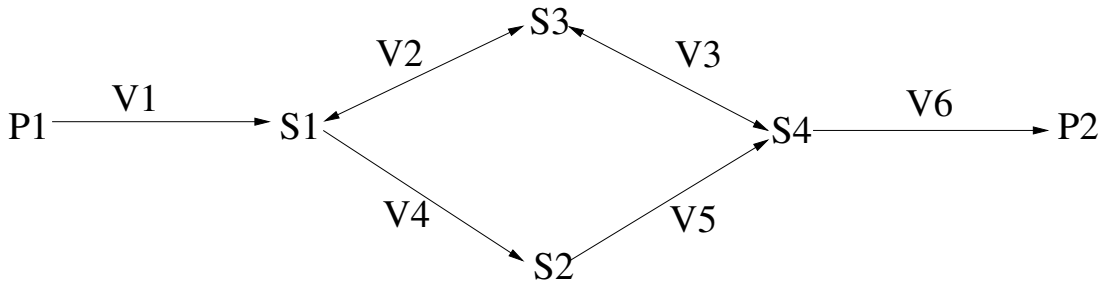
$$K = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{matrix} R1 \\ R2 \\ R3 \end{matrix}$$

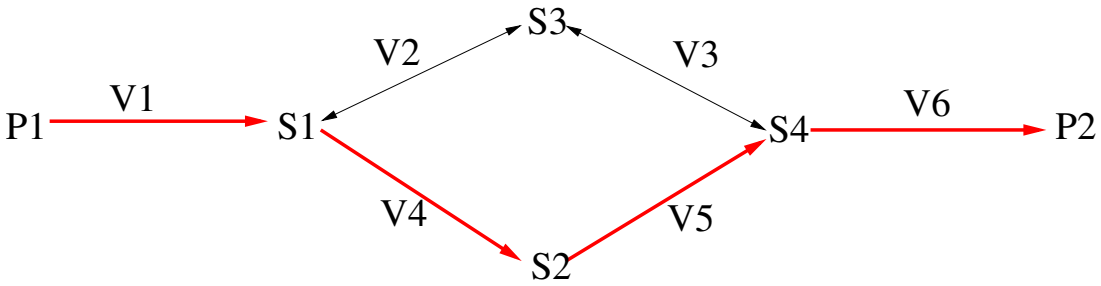
On a:

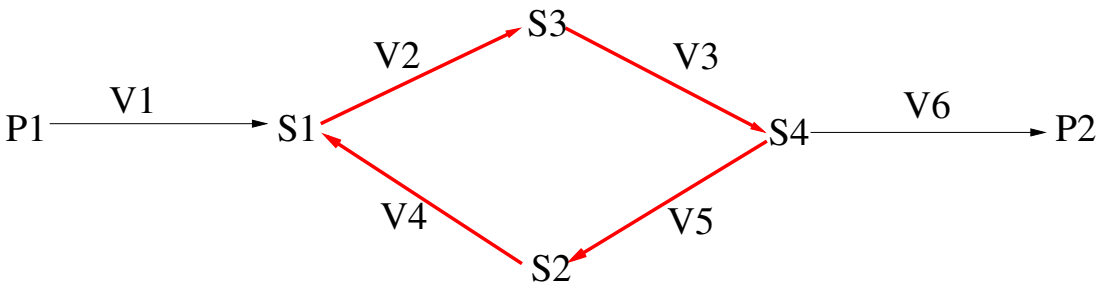
$$N.K_1 = 0 \text{ et } N.K_2 = 0$$

$$(+1-1-1).\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 0 \text{ et } (+1-1-1).\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0$$

ker N







$$K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$$

Reaction correlations

Enzyme subsets

- Set of reactions which can operate only together.
- Detection: Rows in K differ only by scalar factor.
- Strong coupling $\rightarrow$ Indication of common regulation.
- **Network compression !!!**

$$K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ -1 & 1 \\ 0 & 1 \end{pmatrix}$$

overall reaction

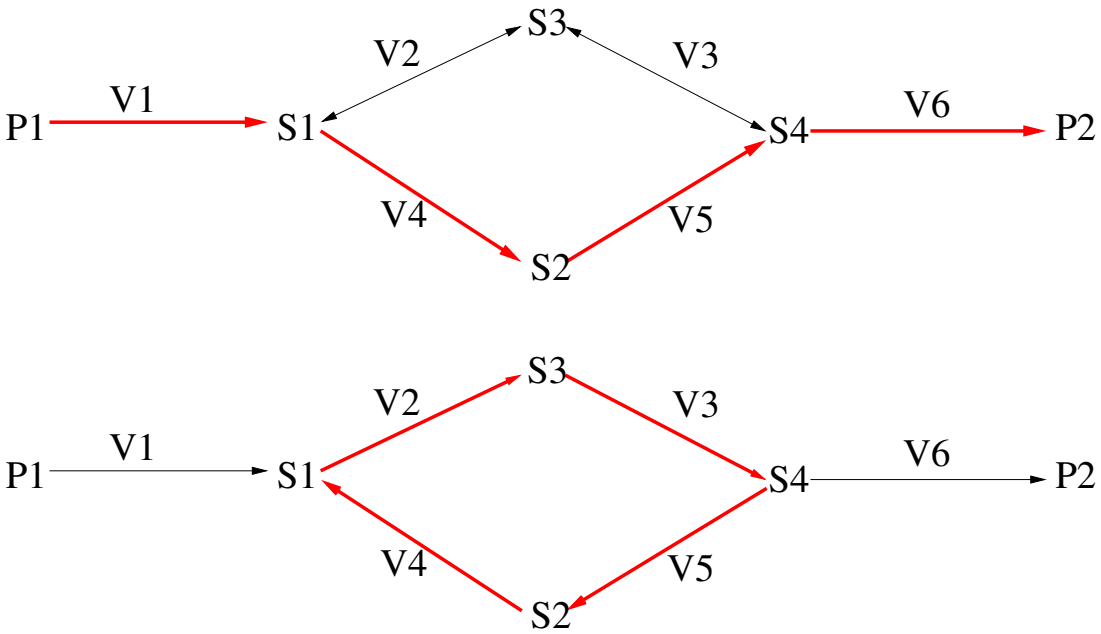
1: S4 + P1 = S1 + P2

2: S1 = S4

3: S1 = S4

Constraintes sur l'espace de solutions

- Tous les flux possibles de ce réseau à l'état stationnaire sont dans $\ker N$.



$\rightarrow$ Il faut tenir compte de l'irréversibilité de certaines réactions.

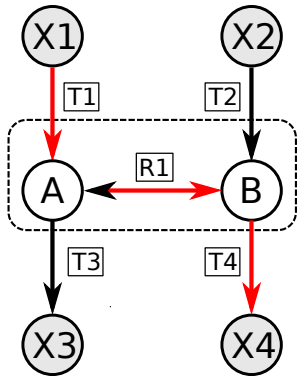
Constraints-based modelling (CBM)

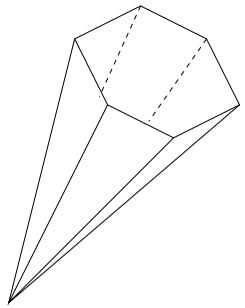
- **Steady state hypothesis**: metabolite concentration is constant (i.e. there is no accumulation of internal metabolites in the system).

$$\frac{dm(t)}{dt} = 0 \Rightarrow S.V(m(t)) = 0$$

- **Reaction directionality constraints**:

$$v_j \geq 0, \text{if } j \text{ is irreversible}$$





$\Rightarrow$ **Solution space**: $P = \{v \in \mathbb{R}^r \mid S.v = 0 \text{ and } v_i \geq 0 \forall i \text{ irreversible reaction}\}$ is a **polyhedral convex cone**

Constraint-based modeling of metabolic networks

"Because biological information is incomplete, it is necessary to take into account the fact that cells are subject to certain constraints that limit their possible behaviors. By imposing these constraints in a model, one can then determine what is possible and what is not, and determine how a cell is likely to behave, but never predict its behavior precisely."

Palsson (2000)

Constraints on fluxes

1. Steady state constraint

$$Sv = 0$$

   Fluxes constrained to **subspace**

2. Irreversibility constraints on some fluxes (from thermodynamics/heuristics/empirical evidence)

$$v_i \geq 0, \forall i \ \textit{irreversible}$$

   Fluxes constraint to **flux cone**

3. Flux bounds from capacity constraints, maintenance, ...

$$v_{i,min} \leq v \leq v_{i,max}$$

   Fluxes constraint to **convex polytope**

Setting up the constraint based model (CBM)

Constraint based model useful if non-trivial steady state fluxes exist

- The steady state equation

$$Sv = 0$$

  should have a non-zero solution v (non-trivial steady state flux space)
- We need rank $S < m$; most models have more reactions than metabolites anyway.

Metabolite / flux units

- In CBMs, metabolites are usually considered in molar amounts per dry biomass: **mmol/g**
- Fluxes are then in **mmol/gh)**

Biomass reaction

- Biomass reaction formalizes consumption of metabolites to generate biomass

$$v_{bio} : \sum_{i=1}^{n} c_i X_i \rightarrow \ 1 \text{ g dry biomass}$$

- Based on pre-determined constant biomass composition
- Coefficients c_i commonly in mmol / g dry biomass
- Unit of v bio becomes 1/h: interpretable as dry biomass growth rate μ

biomass : 0.05 GLU-6-P + 0.69 RIBOSE-5-P + 0.41 ERYTH-4-P + 2.1 PEP + 3.13 PYR + 0.95 ACETYL-CoA + 1.88 OXALO + 82.9 ATP + 3.26 NAD + 13.53 NADPH + 2.77 L-glutamine + 4.4 L-glutamate + 72.83 H2O => BIOMASS + 0.95 CoA + 82.9 ADP + 3.26 NADH + 13.53 NADP + 68.88 H + 6.04 AKG + 84.93 P

Optimization principle

Constraint based model

$$Sv = 0$$

$$v_{i,min} \leq v_i \leq v_{i,max}$$

- Underdetermined system of equalities / inequalities: flux polytope
- How do we determine fluxes v that we expect to occur in nature?

Add an optimization objective

- **Hypothesis:** Cells regulate fluxes within constraints to achieve an "optimal" configuration from an evolutionary perspective.

$$max\ J(v)$$

$$s.t.\ Sv = 0$$

$$v_{i,min} \leq v_i \leq v_{i,max}$$

Use mathematical optimization to predict phenotype $\Rightarrow$ **Linear Programming**

Linear Programming (LP): What is it?

Linear programming is a technique for the optimization of a linear objective function, subject to linear equality and linear inequality constraints.

- We state an objective function that measures what we are interested in.
- try to find the best value for this objective function under the given constraints.
- optimal solution normally lies in a corner of the polytope or along a whole edge.
- Solvers: cplex, glpk, gurobi, etc.

Ecriture générale d'une programmation linéaire

On peut écrire ainsi un programme linéaire avec n variables $x_1, \ldots, x_n$ et m contraintes.

$$\begin{array}{rc} \max & \sum_{i=1}^{n} c_i x_i \\ \text{sous les contraintes} & \sum_{i=1}^{n} a_{ij} x_i \leq b_j, (j = 1, \ldots, m) \\ & x_i \in \mathbb{R}, (i = 1, \ldots, n) \end{array}$$

- **Linéarité :** Objectif et contraintes sont des fonctions linéaires des variables de décision (les coefficients c_i et a_{ij} des variables sont constants)
- **Continuité :** Les variables peuvent prendre n'importe quelle valeur réelle respectant les contraintes linaires

Règle de réécriture

Toute contrainte d'égalité peut s'écrire comme deux inégalités :

$$\sum_{i=1}^{n} a_i x_i = b \equiv \begin{cases} \sum_{i=1}^{n} a_i x_i \leq b \\ \sum_{i=1}^{n} a_i x_i \geq b \end{cases}$$

Toute contrainte $\geq$ peut s'écrire comme une contrainte $\leq$:

$$\sum_{i=1}^{n} a_i x_i \geq b \equiv \sum_{i=1}^{n} -a_i x_i \leq -b$$

Tout problème de minimisation peut s'écrire comme un problème de maximisation :

$$\max \sum_{i=1}^{n} c_i x_i \equiv \min \sum_{i=1}^{n} -c_i x_i$$

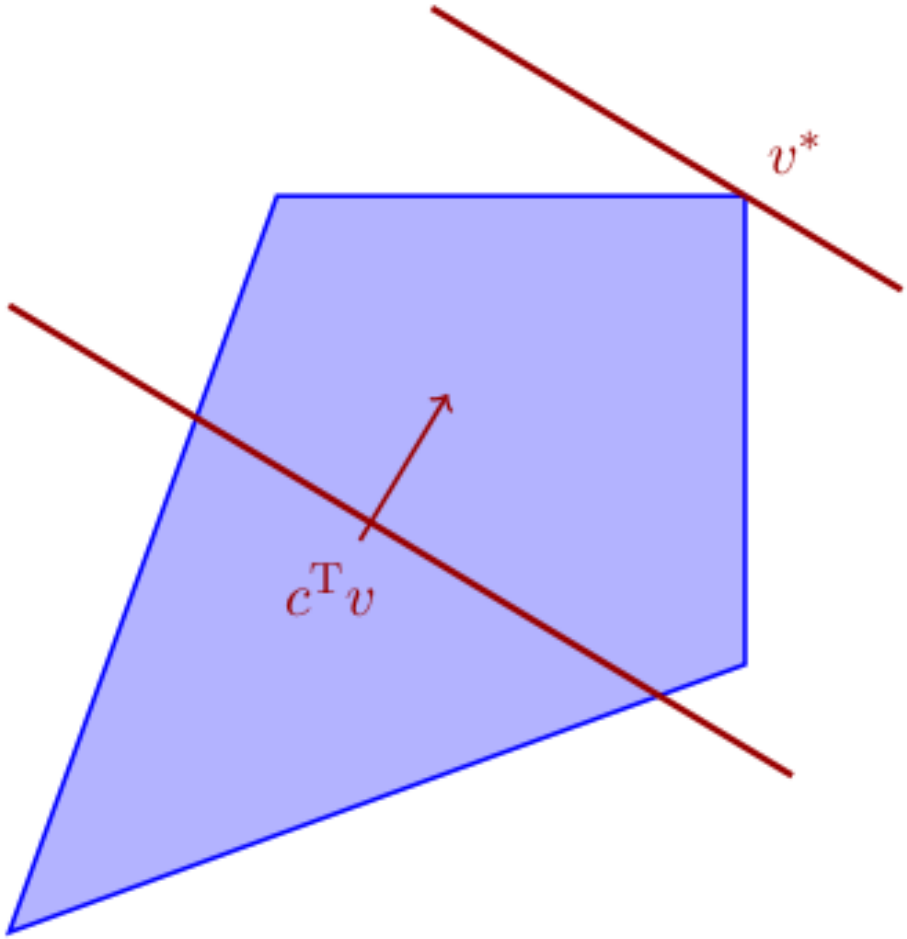
Exemples simples de programmes non linéaires

$$\begin{array}{rc} \min & \sum_{i=1}^{n} x_i x_i \\ \text{sous les contraintes} & \sum_{i=1}^{n} a_{ij} x_i \leq b_j, (j = 1, \ldots, m) \\ & x_i \in \mathbb{R}, (i = 1, \ldots, n) \end{array}$$

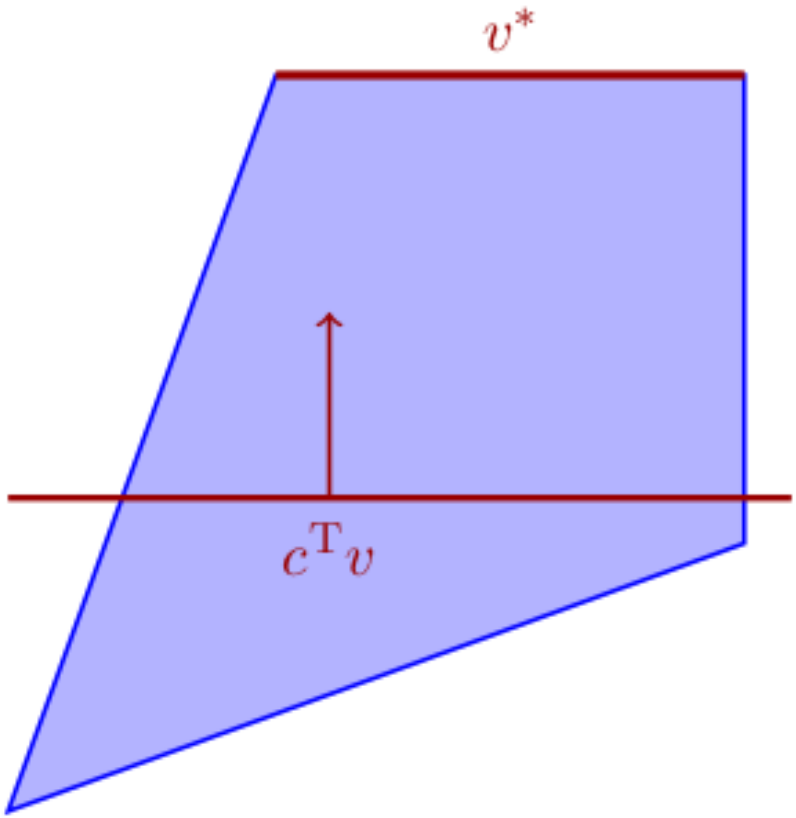
$$\begin{array}{rc} \min & \sum_{i=1}^{n} x_i \\ \text{sous les contraintes} & \sum_{i=1}^{n} a_{ij} x_i \leq b_j, (j = 1, \ldots, m) \\ & x_i \in \mathbb{N}, (i = 1, \ldots, n) \end{array}$$

$$\begin{array}{rc} \min & \sum_{i=1}^{n} c_i x_i \\ \text{sous les contraintes} & \sum_{i=1}^{n} a_{ij} x_i \leq b_j, (j = 1, \ldots, m) \\ & x_1 = x_2 \text{ ou } x_1 = x_3 \\ & x_i \in \mathbb{R}, (i = 1, \ldots, n) \end{array}$$

Generalized geometrical interpretation

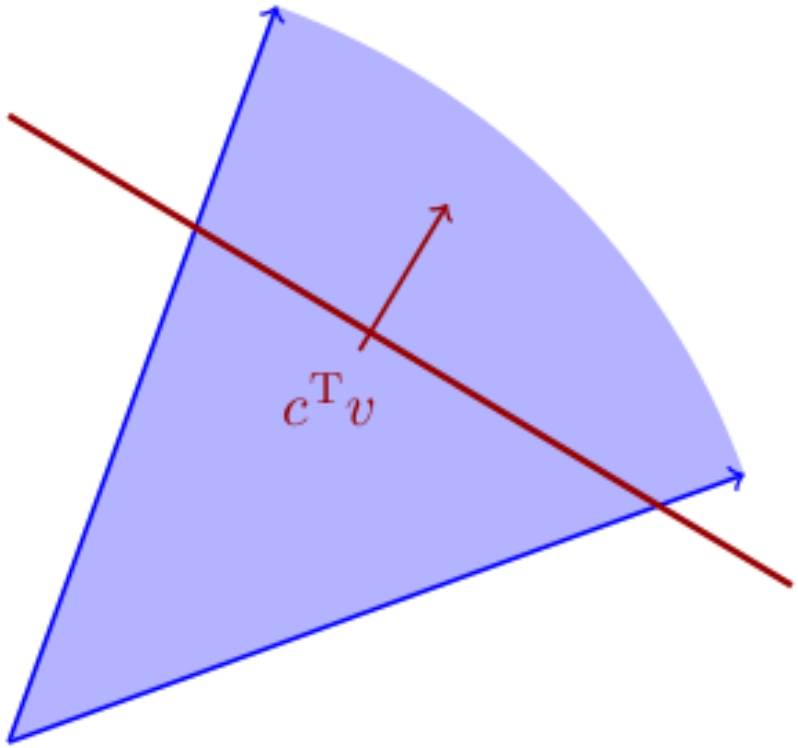


Non-uniqueness of optimal solutions



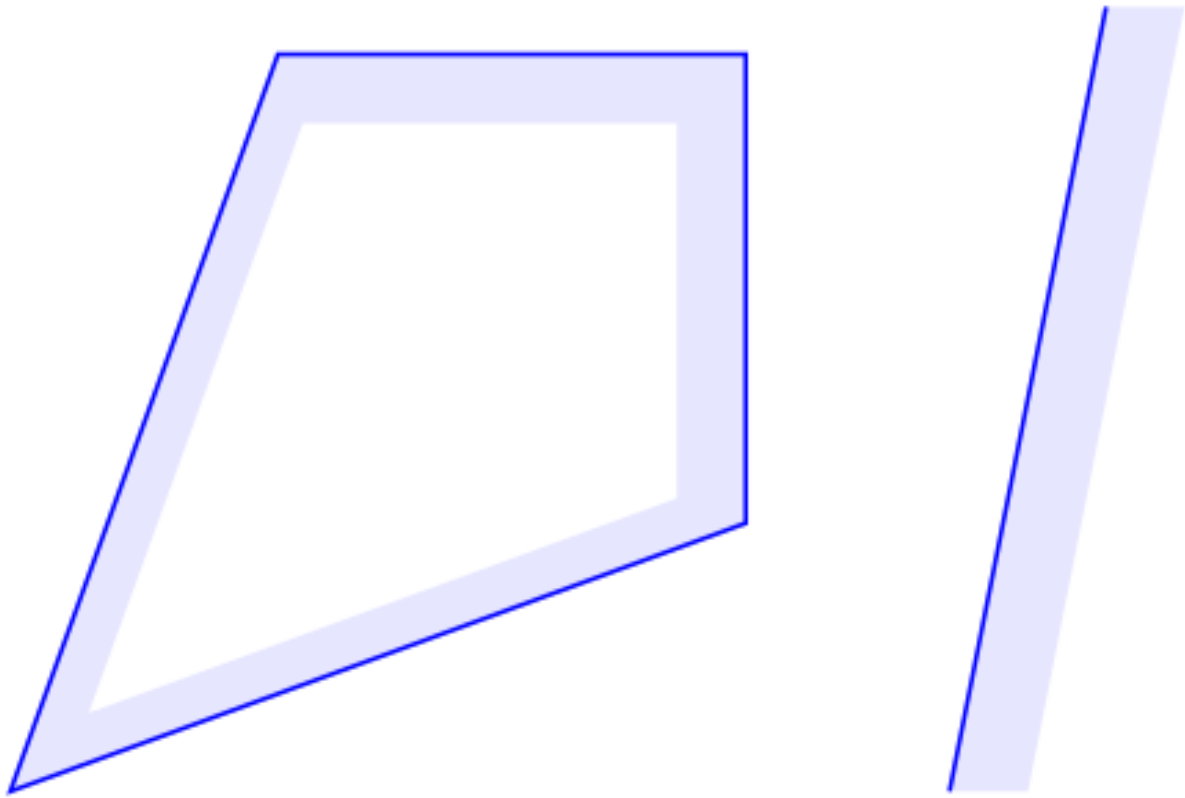
Set of optimal solutions is a face of the polytope

Unboundedness



Unboundedness: $max\ c^T v = \infty$

Infeasibility: Constraint set is empty



$$v_1 + v_2 \leq -1$$

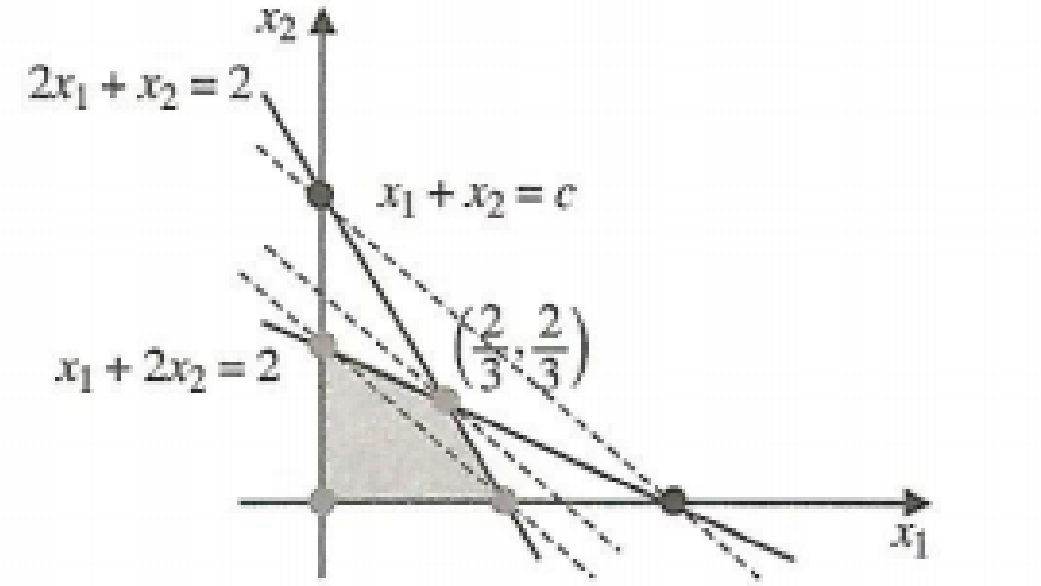
$$v_1, v_2 \geq 0$$

Exercice

$$\max\ x_1 + x_2$$
subject to
$$x_1 + 2x_2 \leq 2$$
$$2x_1 + x_2 \leq 2$$
$$x_1, x_2 \geq 0$$

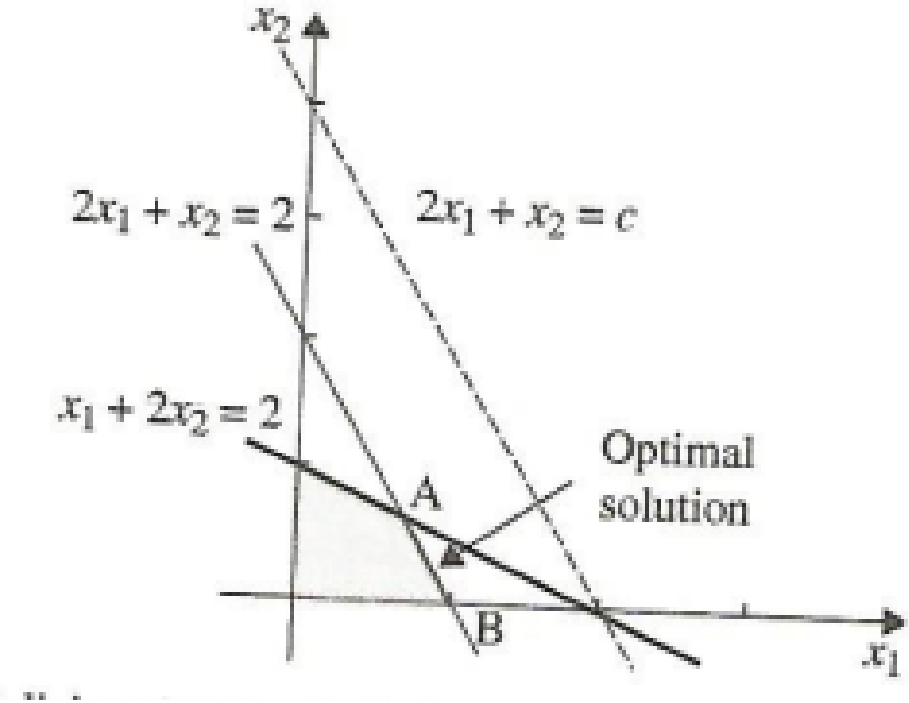
Correction

max $x_1 + x_2$
subject to
$x_1 + 2x_2 \leq 2$
$2x_1 + x_2 \leq 2$
$x_1, x_2 \geq 0$



Correction

max $2x_1 + x_2$
subject to
$x_1 + 2x_2 \leq 2$
$2x_1 + x_2 \leq 2$
$x_1, x_2 \geq 0$



La fonction objectif est coplanaire avec une des contraintes, le nombre de solution optimale est ∞ dans le segment [A B].

Types of Objective Functions

- Physiologically meaningful
  - Max ATP, NADH
  - Max growth $\rightarrow$ biomass objective function
  - Min reactive oxygen species (ROS) production
  - Max photon capture in photosynthesis
  - ...
- Bioengineering
  - Max succinate production in *E. coli*
  - Max malate production in *S. cereviseae*
  - ...
- Exploration
  - Max alanine from succinate
  - ...

Constraints on metabolism

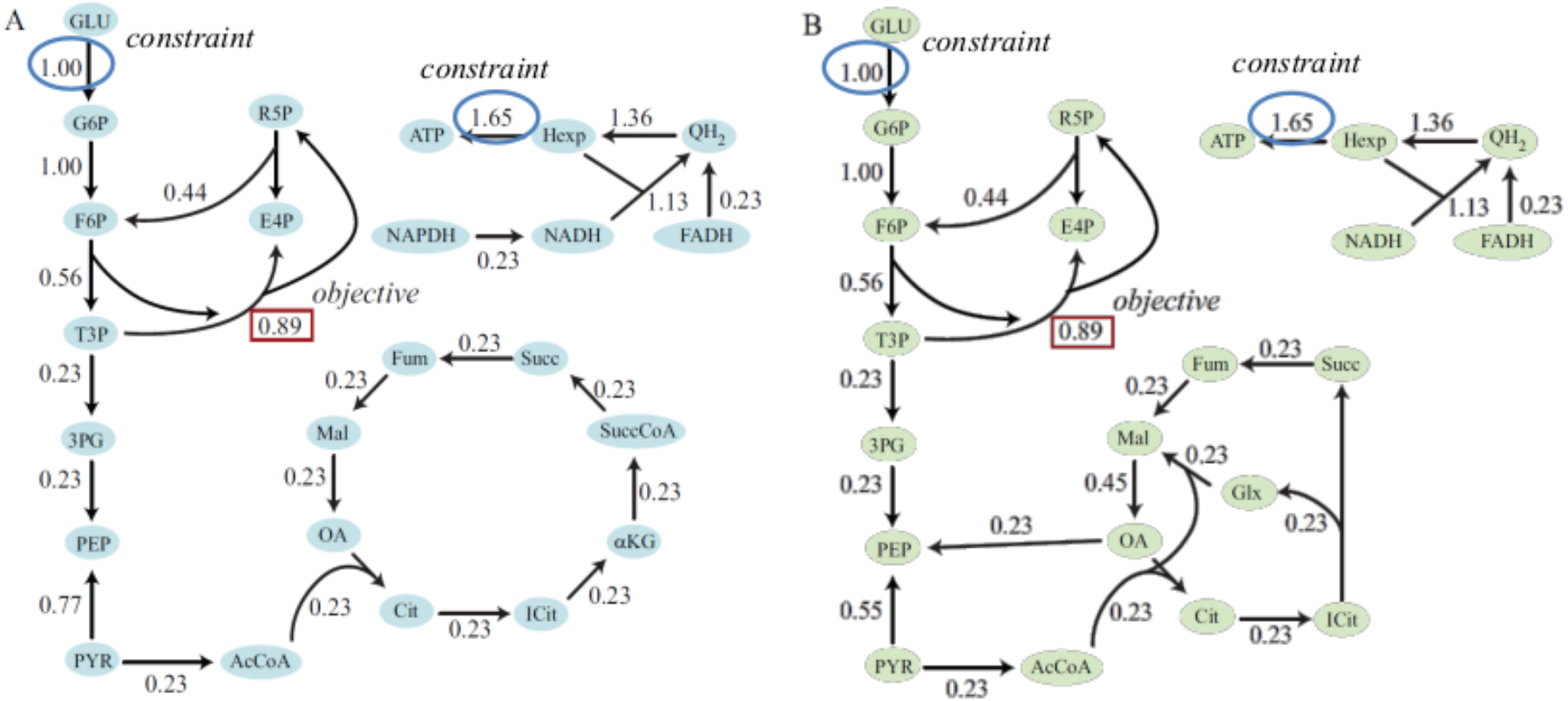
- Physicochemical
  - Mass conservation
  - Thermodynamics
  - Solvent capacity
  - Enzyme kinetics
- Spatial or topological
  - Limited space in membrane
  - DNA is tightly packed
  - Limited amount of mRNA
  - Limited amount of protein
  - Limited diffusion rate
- Environmental
  - Nutrient availability
  - pH
  - Temperature
  - Osmolarity
- Regulatory
  - Change in amount
    - Regulate transcription or translation
  - Change in activity
    - Chemical transformation (e.g. phosphorylation)
    - Binding of ligand, inhibitor, activator

What do you do with FBA ?

- Predicting the effects of gene deletions
  - the values in the stoichiometric matrix of the deleted gene become 0
  - by recalculating the objective function, you can predict the effect of a gene knockout
- Predicting the effects of media composition changes
  - add additional species to include the new chemicals
  - see what happens to your target product as a result
- Predicting the yields of important cofactors such as ATP, NADH...
- ...

Alternative solutions

Multiple solutions can result in the same value of the objective function



Flux Variability Analysis (FVA)

- On recherche un encadrement de toutes les solutions au problème FBA
- On calcul la valeur optimale avec FBA
- Pour chaque vitesse de réaction v_i , on recherche le minimum et le maximum que peut admettre cette réaction, lorsque la fonction objective du problème FBA a atteint sa valeur optimale:

max v_i	min v_i
st	st
$v_b = v_{opt}$	$v_b = v_{opt}$
$S.v = 0$	$S.v = 0$
$v_i \geq 0$ si v_i irreversible	$v_i \geq 0$ si v_i irreversible

Rem Si le réseau a n réactions, on résoud $1 + 2 * n$ LP

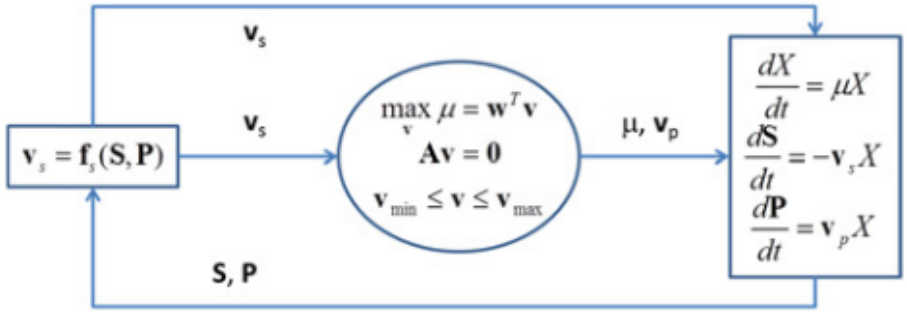
Dynamic FBA

FBA

$$\begin{array}{lll} \max & C^T.v & \text{(Objective fct)} \\ \text{s. t.} & S.v = 0 & \text{(steady state)} \\ & l \leq v \leq u & \text{(flux constraints)} \end{array}$$

Differential system

$$\begin{array}{lcl} \frac{dC(t)}{dt} & = & S_c.v(C(t)).X(t) \\ \frac{dX(t)}{dt} & = & \mu(t).X(t) \end{array}$$



- At each time step, FBA is used to predict growth (μ), nutrient uptake and by-product secretion rates (V_c)
- These rates are then used to calculate biomass and nutrient concentrations in the culture at the end of the time step.
- The concentrations can, in turn, be used to calculate maximum uptake rates of nutrients for the next time step.

Yarrowia Lipolytica

Lipids

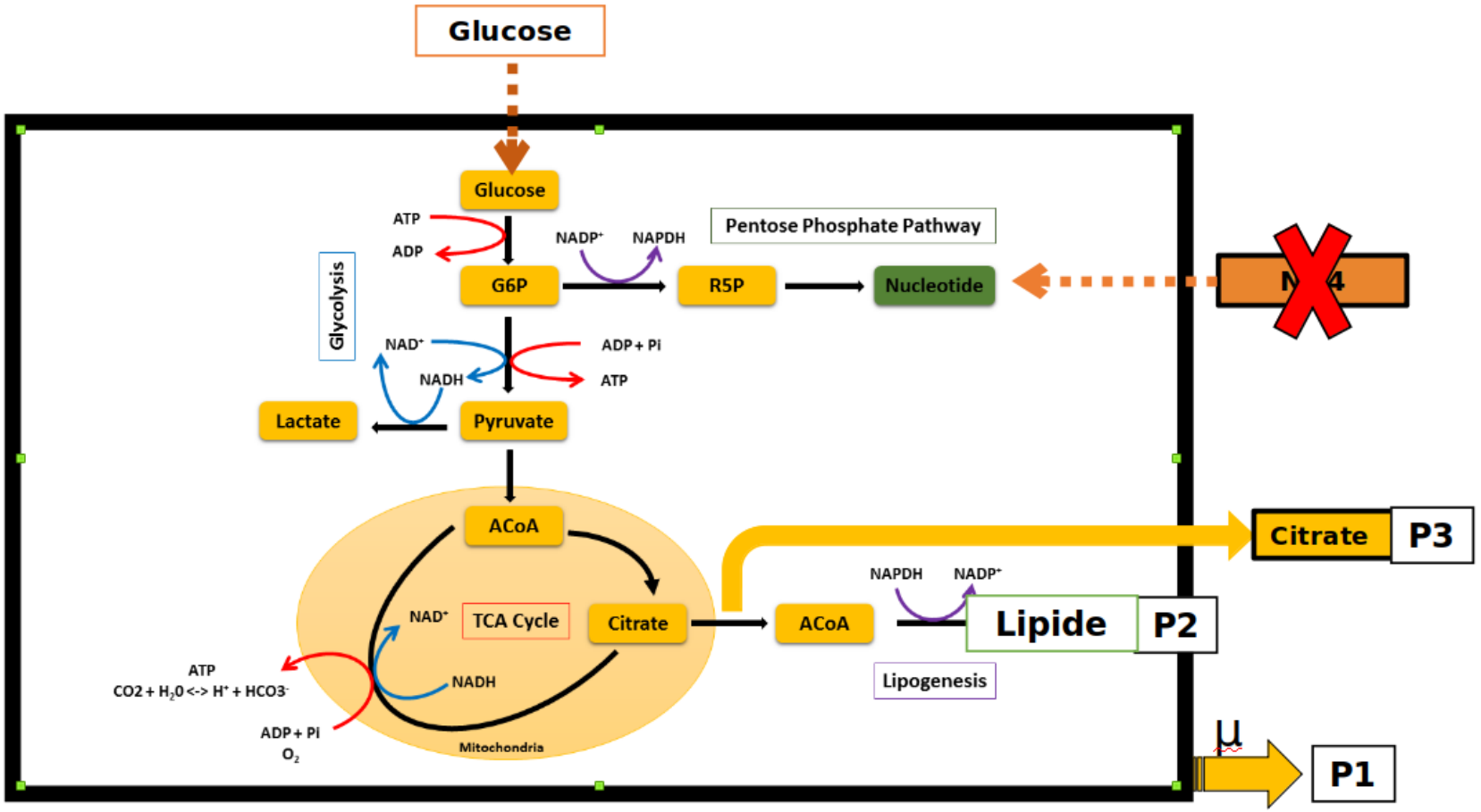
Citric Acid

Food additive (E330)
Acidifier (soft drinks)
Acidity corrector
Flavor precursor

Using computational approaches to adress industrial problems.

Find alternative strategy (without GMO) to maximize citrate production.

Dynamic FBA for triggering citrate overproduction in *Yarrowia lipolytica*

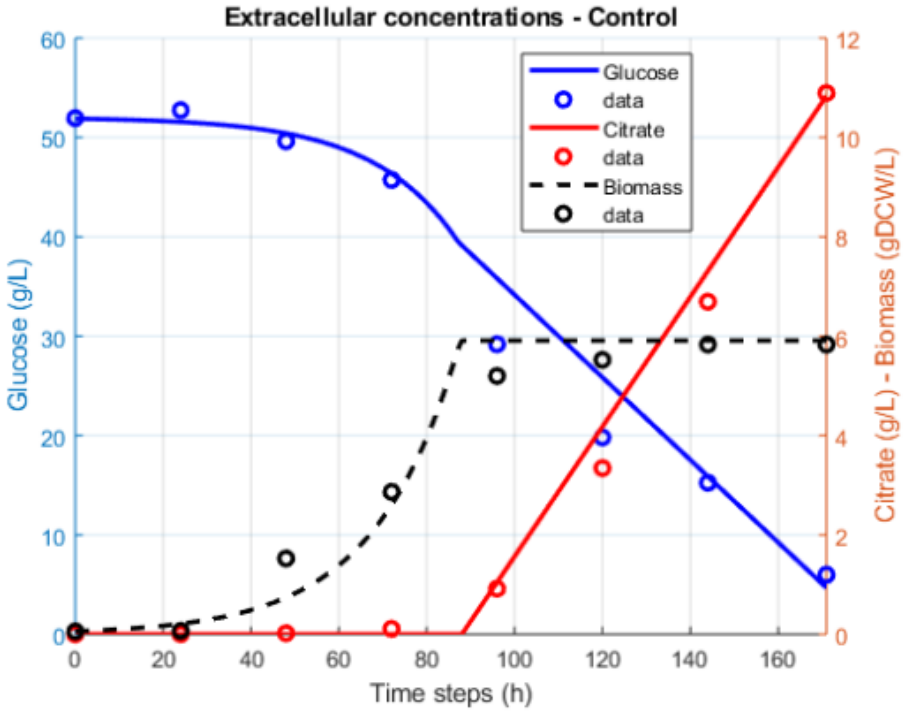


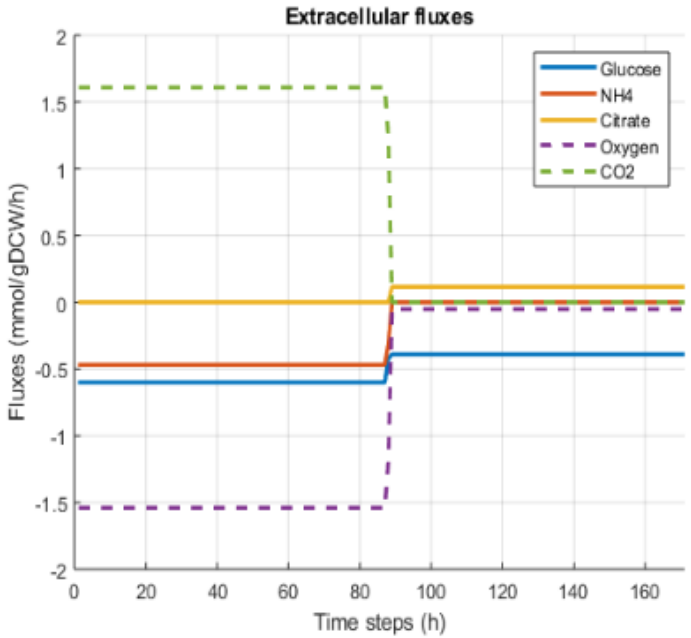
Genome-scale metabolic model : **1985 reactions, 1683 metabolites**.

Dynamic FBA to identify metabolic levers for citrate optimization.

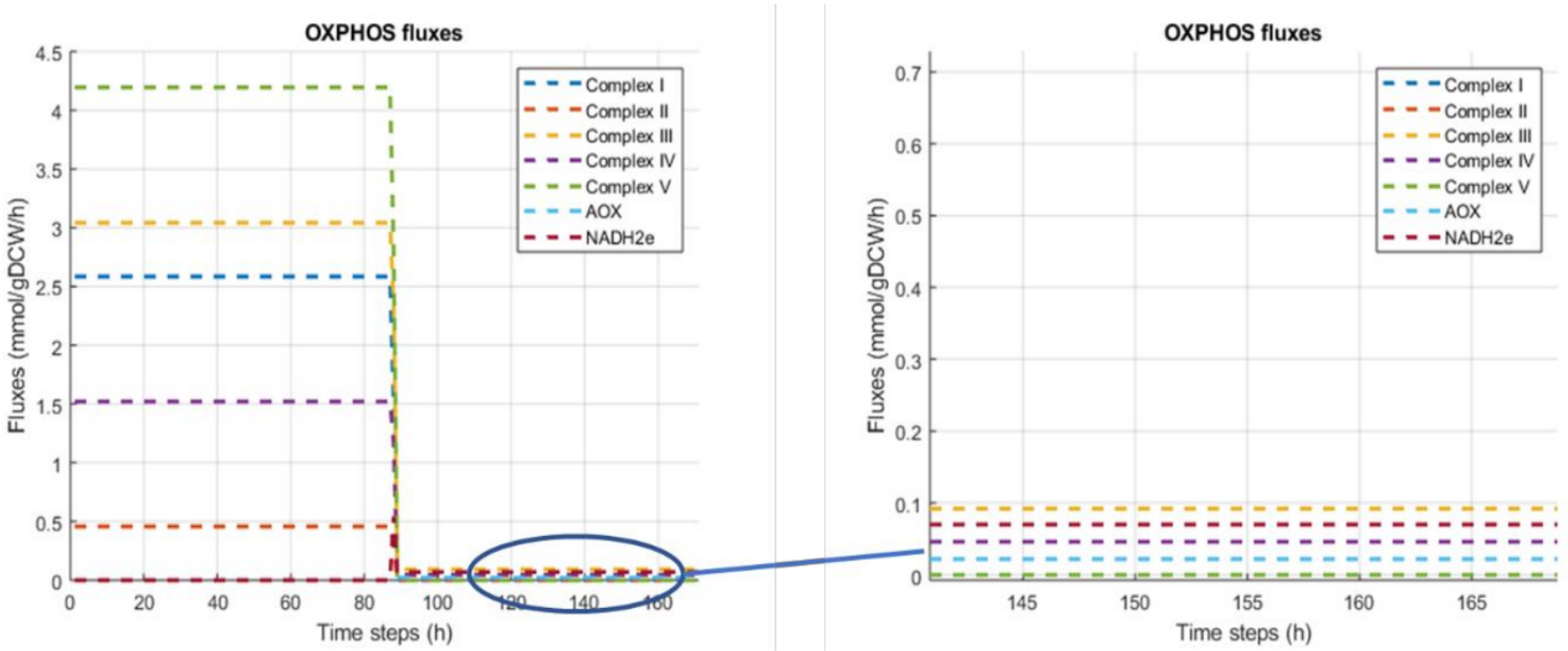
Da Veiga Moreira, Jolicœur, Schwartz & Peres. (Scientific Reports 2021)

Model Calibration with experimental data and fluxes analysis



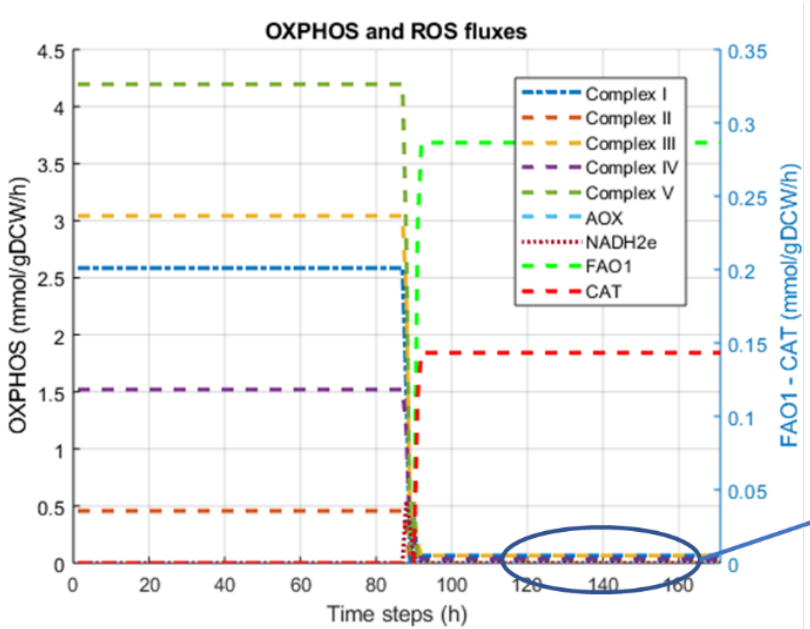


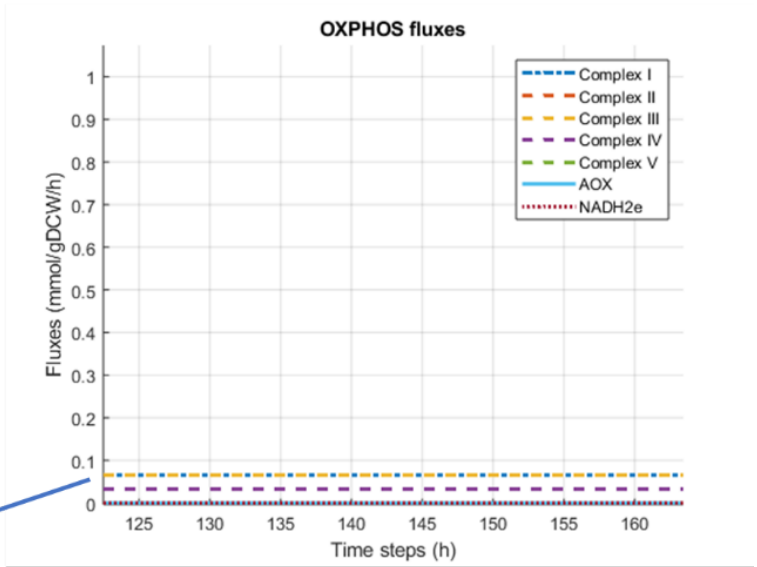
Activité de la chaîne respiratoire en phase μ et stat



- OXPHOS est réduite en phase stat
- Complexes I & V sont inhibés en phase stat
- La respiration alternative (AOX + NADH2e) est activité en phase stat

Experimental Alternative Oxidase inhibition by adding n-Propyl Gallate enhances citrate production



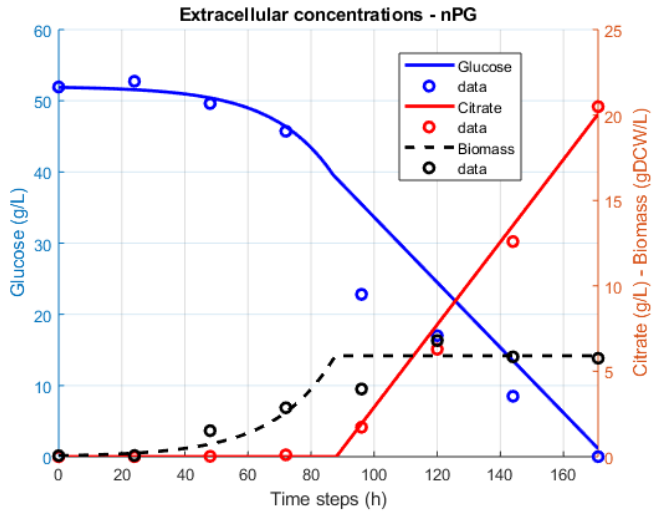


Inhibition of AOX with nPG:

⇒ Activation of complexes I & V

⇒ Inhibition of alternative respiration (AOX + NADH2e)

⇒ Increase citrate production



Modulating mitochondrial activity triggers respiro-fermentative transition

Définition des modes élémentaires

Un vecteur $e = (e_1, ..., e_r)^t \in \mathbb{R}^r$ est un *mode élémentaire* (*EFM*) s'il vérifie les conditions suivantes[1] :

1. *état stationnaire* : $Ne = 0$.

2. *faisabilité* : Pour chaque indice j d'une réaction irréversible $e_j \geq 0$.

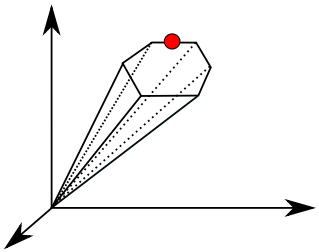
3. *minimalité* : Soit $supp(v) = \{j \in \mathbb{N} : v_j \neq 0\}$. Pour chaque *efm* e' de N, $supp(e') \subseteq supp(e) \Rightarrow \exists \alpha \in \mathbb{R}$ tel que $e' = \alpha e$.

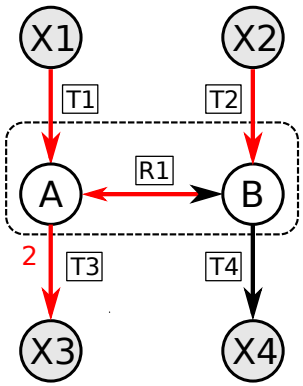
Un *EFM* est un ensemble minimal d'enzymes pouvant opérer à l'état stationnaire en tenant compte des réactions irréversibles.

FBA vs EFMA

FBA
Flux Balance Analysis

$$\begin{array}{lll} \max & C^T.v & \text{(Objective fct)} \\ \text{s. t.} & S.v = 0 & \text{(steady state)} \\ & l \leq v \leq u & \text{(flux constraints)} \end{array}$$

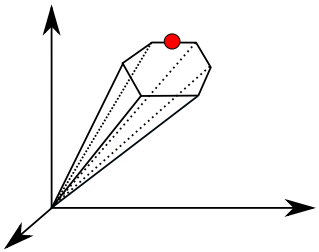


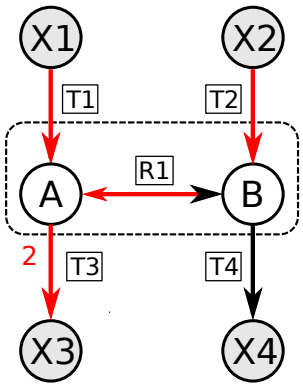


FBA vs EFMA

FBA
Flux Balance Analysis

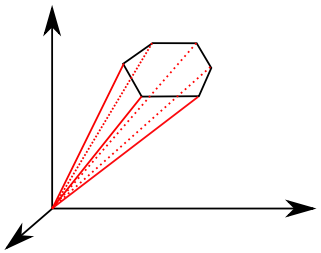
$$\begin{array}{lll} \max & C^T.v & \text{(Objective fct)} \\ \text{s. t.} & S.v = 0 & \text{(steady state)} \\ & l \leq v \leq u & \text{(flux constraints)} \end{array}$$

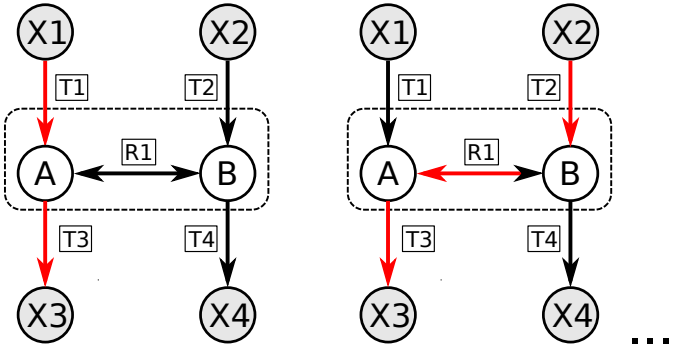




EFMA
Elementary Flux Mode Analysis

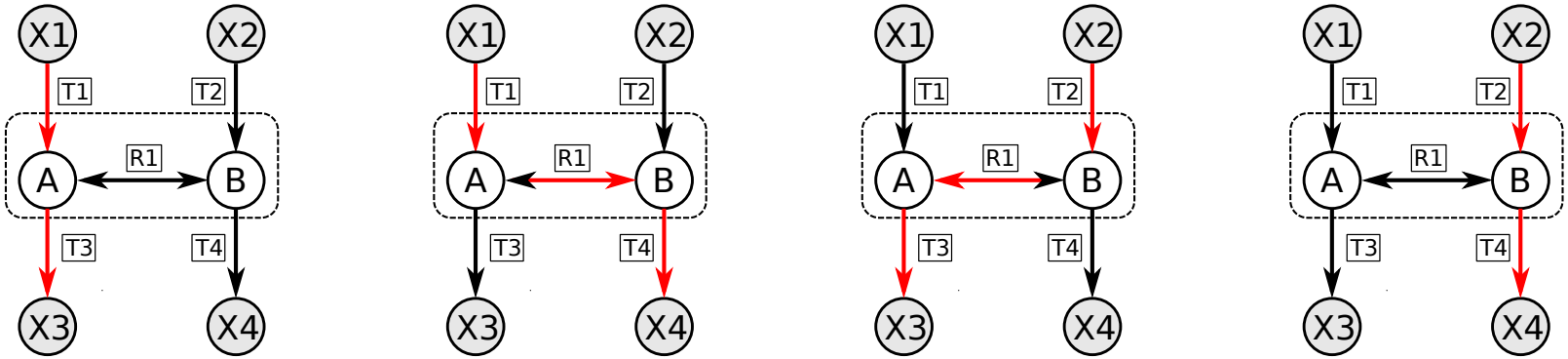
$P = \{v \in \mathbb{R}^r \mid S.v = 0 \text{ and } v_{irrev} \geq 0\}$
EFMs = minimal supports (for inclusion) of P
For $v \in P, supp(v) = \{j \in \mathbb{N} : v_j \neq 0\}$





Exemple

EFMs = **support minimal vectors of the flux cone** that contains all feasible steady-state flux vectors of a given metabolic network

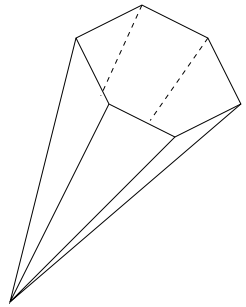


All other possible flux distributions: $r = \sum_{j=1}^{4} \alpha_j . e^j$ with $\alpha_j \geq 0$

Intêrets des *EFMs* dans l'étude de la structure des réseaux métaboliques

- Identification de voies métaboliques.
- Robustesse du réseau (si la production d'un métabolite est toujours possible après suppression de réactions).
- Importance des réactions (fréquence des réactions).
- Corrélation des réactions (enzyme subset : groupe d'enzymes qui opère toujours ensemble avec une proportion de flux fixe dans tous les états stationnaires du système).
- Délétions létales minimales (Minimal Cut Sets)
- Réduction de réseau pour faire des simulations dynamiques

Cône convexe polyédrique



- **Matrice de représentation** A:

$$P = \{x \in \mathbb{R}^d : A.x \geq 0\}$$

- **Matrice génératrice** R:

$$P = \{x \in \mathbb{R}^d : x = R.\lambda \text{ for some } \lambda \geq 0\}$$

Les vecteurs colonnes de R sont les arêtes (« extreme rays ») du cône, c'est-à-dire ses rayons indécomposables.

Double Description (DD) method

- **Théorème de Minkowski**:
  Soit P un cône polyédrique dans $\mathbb{R}^d$ défini par une matrice réelle de représentation $A(m \times d)$:
  $P = P(A) = \{x \in \mathbb{R}^d | Ax \geq 0\} \Rightarrow$ Il existe une matrice réelle $R(d \times n)$ qui engendre P : $P = \{x \in \mathbb{R}^d | x = Ry \text{ avec } y \geq 0\}$.

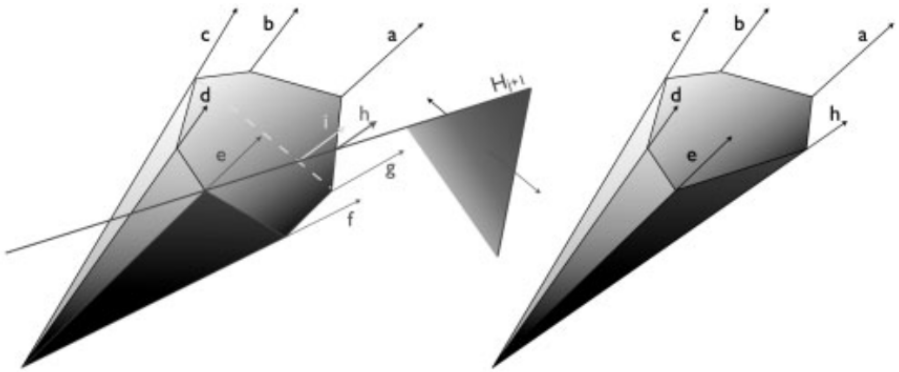
- On appelle la paire (A, R) une « paire de double description » ou « **DD paire** ». La réciproque de ce théorème est également vraie (théorème de Weyl).

Double Description (DD) method

- Soit $A \in \mathbb{R}^{m \times d}$ et $P(A) = \{x : A.x \geq 0\}$
- La méthode DD est un algorithme incrémental permettant de construire $R \in \mathbb{R}^{d \times m}$ tel que (A,R) est une DD paire.
- On assume que P(A) est un cône pointé (pour obtenir un ensemble unique)
- La méthode classique pour calculer R consiste à partir d'une DD paire (A_K, R) telle que A_K soit la sous-matrice de A composée des lignes indexées par un sous-ensemble K d'indices des lignes de A, à ajouter une à une à A_K les lignes restantes de A et à recalculer R à chaque itération, jusqu'à ce que toutes les lignes soient dans A_K.

Double Description (DD) method

- **La paire initiale** (A_K, R):
  - soit partir d'une paire avec $|K| = 1$ (donc d'une seule ligne),
  - soit partir d'une sous-matrice maximale A_K de A dont toutes les lignes sont linéairement indépendantes.
- **Une méthode pour passer de** (A_K, R) **à** (A_{K+i}, R').
  - Géométriquement, ajouter une nouvelle ligne A_i correspond à couper le cône $P(A_K)$ par un hyperplan et à en garder la partie positive.



Double Description (DD) method

- Trouver la matrice R' correspond à déterminer les nouvelles arêtes du cône obtenu. Ces nouvelles arêtes étant sur des anciennes faces du cône peuvent s'exprimer par la combinaison de deux anciennes arêtes (une positive et une négative vis-à-vis de l'hyperplan).
- Remarque: La combinaison se limite à toutes les **paires adjacentes** dont les éléments sont dans un demi-espace différent par rapport à l'hyperplan.

DD pour les réseaux métaboliques

- Contraintes stœchiométriques:

$$N.V = 0 \Rightarrow N.V \geq 0 \ \& \ -N.V \geq 0$$

- Contraintes irréversibilités :
  - Pour chaque réaction irréversible: $v_i \geq 0$
  - On dédouble les réactions réversibles: $v_i \geq 0$ & $v_i^{rev} \geq 0$

$$A = \begin{matrix} Id \\ N \\ -N \end{matrix}$$

Double Description (DD) method

Algorithm 1: DDStandard(**A**)

desc : Standard implementation of double description method
input : Matrix $\boldsymbol{A}$, defining a polyhedral cone $\mathcal{P} = \{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\ ,\ \boldsymbol{x} \geq \boldsymbol{0}\}$
output: Matrix $\boldsymbol{R}$, extreme rays of $\mathcal{P} = \{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{x} = \boldsymbol{R}\boldsymbol{c}\ ,\ \boldsymbol{c} \geq \boldsymbol{0}\}$
begin

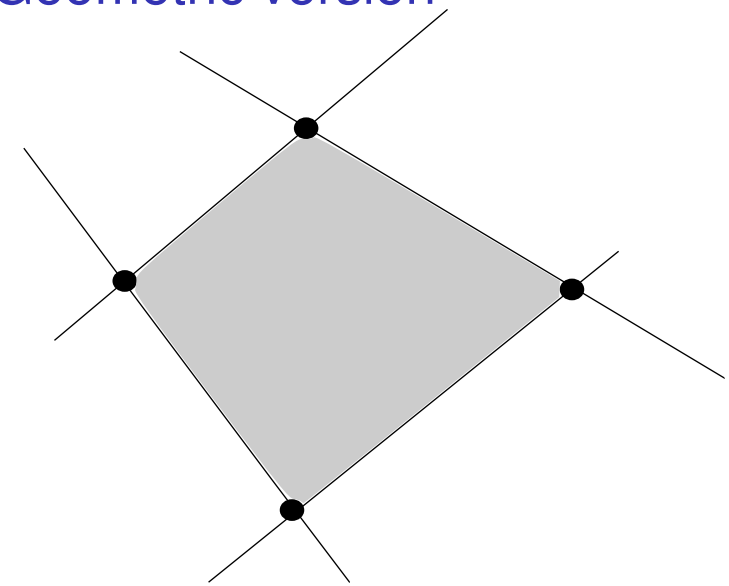
```
    Ak <- maximal submatrix of A consisting of linearly independent row of A
1   R ← Ak^-1
2   ρ ← ···                          /* indices of already considered non-negativity constraints */
3   while (ρ ≠ {1, 2, ··· , d}) do
4       j ← choose from {1, 2, ··· , d} \ ρ
5       τ> ← {i | A j,i > 0}
6       τ0 ← {h | A j,h = 0}
7       τ< ← {k | A j,k < 0}
8       τadj ← {(i, k) | (i, k) ∈ (τ> × τ<) : R*i is adjacent to R*k}
9       Rnew ← []                                   /* empty matrix to store new rays */
10      foreach (i, k) ∈ τadj do
11          p ← R*i
12          q ← R*k
13          r(ik) ← pj q − qj p
14          append column r(ik) to Rnew
15      end
16      R ← [ [R*i] , [R*h] , Rnew ]    with i ∈ τ> and h ∈ τ0
17      ρ ← ρ ∪ {j}
18  end
end
```

EFMs = Extremal ray of the polyhedral convex cone

EFMs enumeration

- Based on **Double Description (DD) method** (Motzkin et al., 1953).
- **nullspace initial matrix** to improve performance (Wagner et al., 2004).
- **binary approach** to reduce memory demands (Gagneur et al., 2004)
- **bit pattern trees** to optimize searching of subsets during elementary testing. (Terzer et al., 2008)

Geometric version

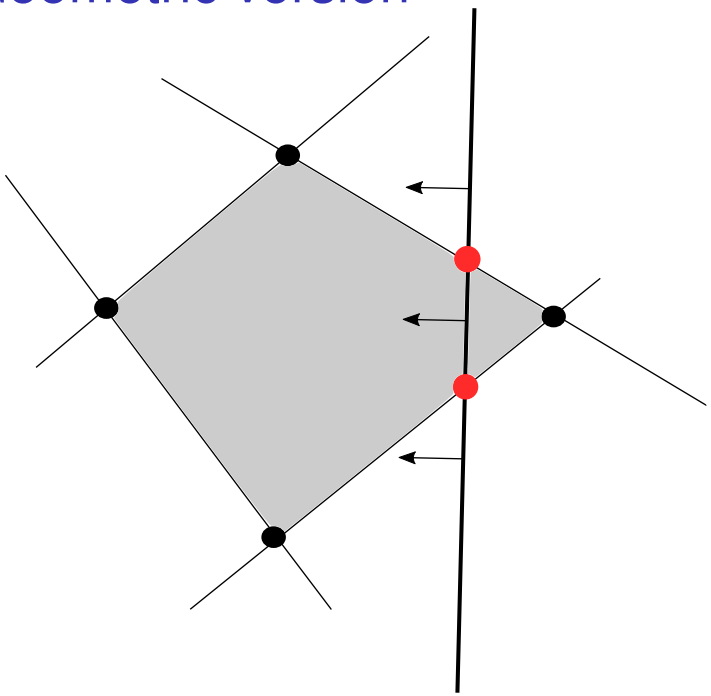


EFMs = Extremal ray of the polyhedral convex cone

EFMs enumeration

- Based on **Double Description (DD) method** (Motzkin et al., 1953).
- **nullspace initial matrix** to improve performance (Wagner et al., 2004).
- **binary approach** to reduce memory demands (Gagneur et al., 2004)
- **bit pattern trees** to optimize searching of subsets during elementary testing. (Terzer et al., 2008)

Geometric version

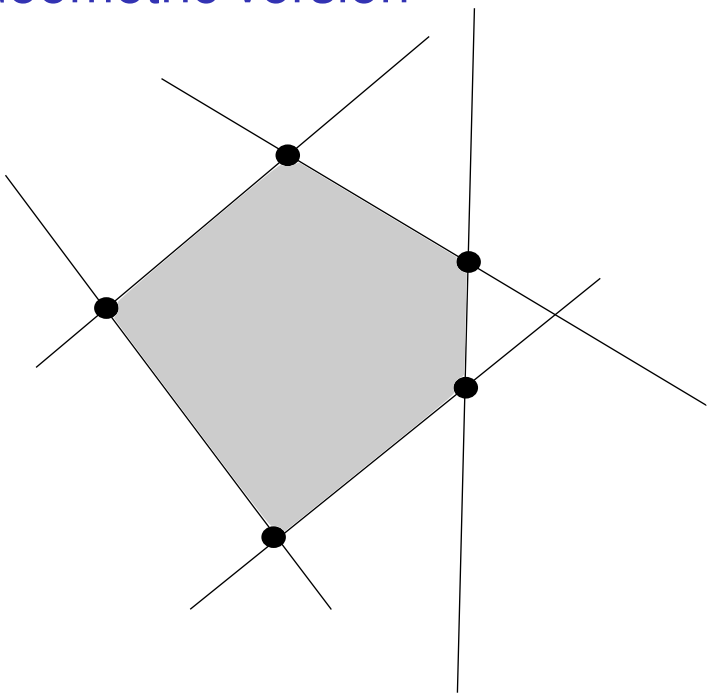


EFMs = Extremal ray of the polyhedral convex cone

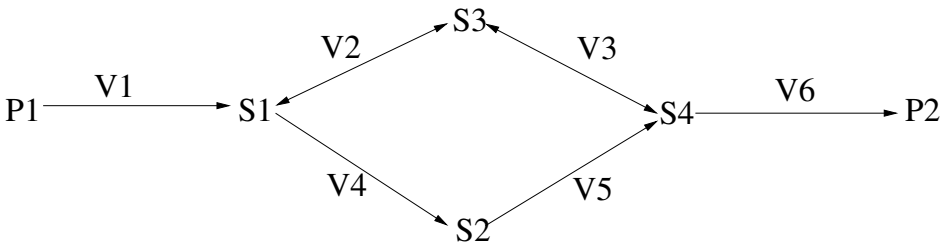
EFMs enumeration

- Based on **Double Description (DD) method**
  (Motzkin et al., 1953).
- **nullspace initial matrix** to improve performance
  (Wagner et al., 2004).
- **binary approach** to reduce memory demands
  (Gagneur et al., 2004)
- **bit pattern trees** to optimize searching of subsets during elementary testing.
  (Terzer et al., 2008)

Geometric version



Exemple



1: S4 + P1 => S1 + P2
2: S1 => S4
3: S1 = S4

métabolites/enzymes	R1	R2	R3	$R3_{rev}$
S1	1	-1	-1	+1
S4	-1	1	1	-1

Exemple

$$A = \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & -1 & -1 & +1 \\ -1 & 1 & 1 & -1 \\ \hline -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & +1 \end{array}$$

Exemple

Initialisation (K):

on part d'une sous matrice maximale A_K de A dont les lignes sont linéairement indépendantes (donc $|K| = 4$ puisque A est de rang maximal) et d'une matrice R tel que $A_K R = I$, soit l'inverse de A_K. Un choix naturel est par exemple $A_{1..4}$, qui est la matrice identité, d'où $R = A_{1..4}$.

K+1: *newLine* est donc la 5^{eme} ligne. A_K devient donc :

$A_{1..5} =$

c1	c2	c3	c4
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	-1	-1	1
-1	1	1	-1
-1	1	1	-1
1	-1	-1	1

$R_5 =$

c1	c2	c3	c4
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

On cherche toutes les combinaisons linéaires de façon à ce les nouvelles arêtes soient dans l'hyperplan: c1+c2, c1+c3, c2+c4, c3+c4

On supprime c2 et c3 qui sont dans la partie négative de l'espace coupé par l'hyperplan et on ajoute les nouvelles

Mise à jour A et R

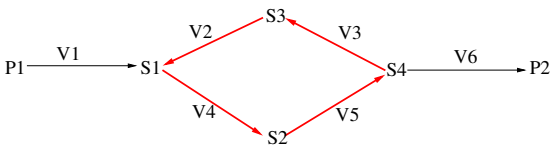
$A_{1..6}$ =

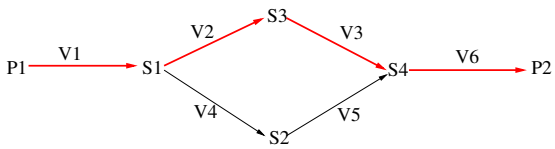
c1	c4	c1+c2	c1+c3	c2+c4	c3+c4
1	+1	0	0	0	0
-1	-1	0	0	0	0
-1	-1	0	0	0	0
1	+1	0	0	0	0

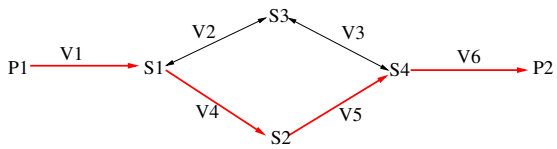
R_6 =

c1	c2	c3	c4	c1+c2	c1+c3	c2+c4	c3+c4
1	0	0	0	1	1	0	0
0	1	0	0	1	0	1	0
0	0	1	0	0	1	1	1
0	0	0	1	0	0	0	1
	*	*					*
*			*				

3 *EFMs*: R1-R2, R1-R3, R2-R3





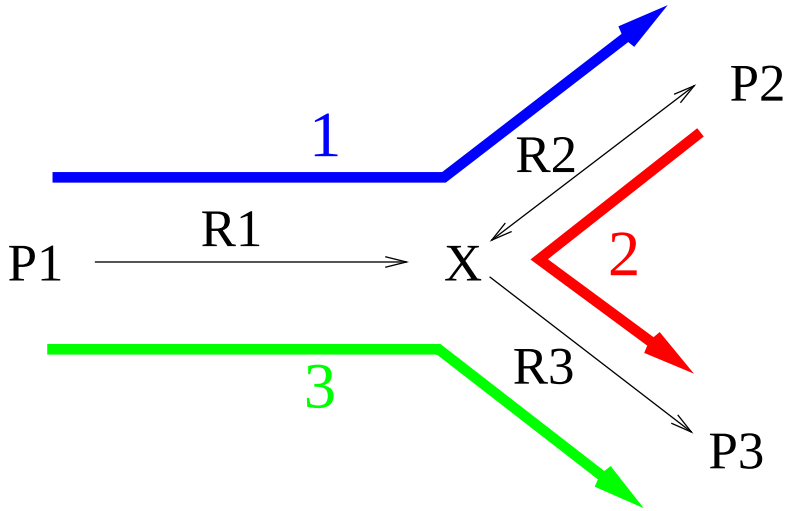


Logiciels pour déterminer les modes élémentaires

- METATOOL (in C/Matlab/Octave) - Th. Pfeiffer, A. von Kamp, S. Schuster
- EFMtool (in Java/MATLAB) - M. Terzer, J. Stelling
- CellNetAnalyzer (in MATLAB) - S. Klamt
- FluxmodeCalculator (in MATLAB) - J.B. van Klinken
- Elmo-Comp (in C++) - D. Jevremovic, D. Boley
- regEfmTool/ tEFMA (in java) - C. Jungreuthmayer, J. Zanghellini
- SMTtool (in C++) - M. Morterol, P. Dague, S. Peres, L. Simon
- aspefm (in ASP/python) - M. Mahout, R. P. Carlson, S. Peres

These algorithms still connot cope with genome-scale metabolic networks reconstructly recently ($\geq$ 1 000 reactions).

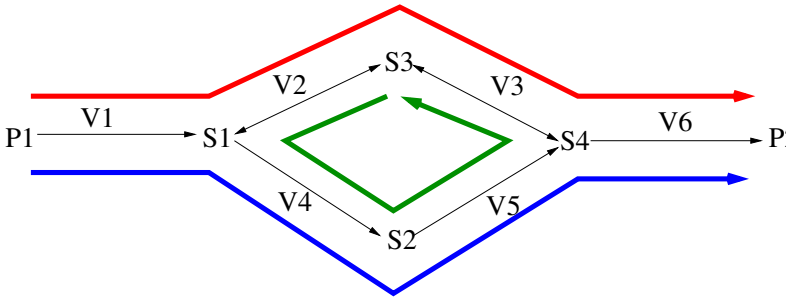
Exemple



- 3 *EFMs* : (R1,R2); (-R2,R3); (R1,R3).
- Tout flux du réseau peut être écrit comme combinaison linéaire positive de ces modes.
- R1 est nécessaire pour la production de P2 mais sa suppression n'empêche pas la production de P3.

Minimal Cut Sets

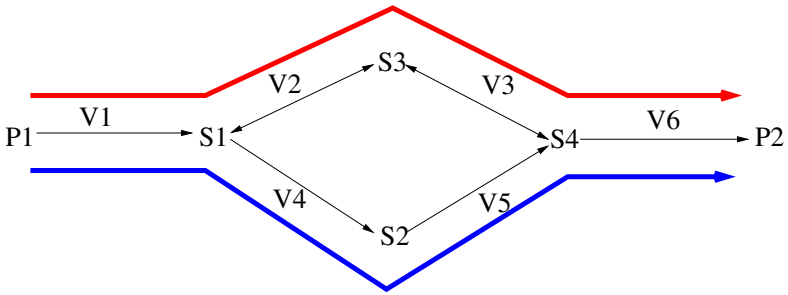
The concept of a minimal cut set has been introduced to determine the minimal set of reactions whose deletion completely blocks a target.



The identification of a minimal cut set is useful in metabolic engineering and in the identification of drug targets (Klamt and Gilles 2004).

Minimal Cut Sets

The concept of a minimal cut set has been introduced to determine the minimal set of reactions whose deletion completely blocks a target.

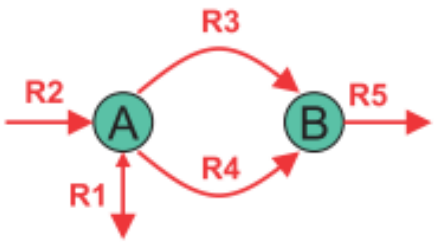


$$MCS = \begin{pmatrix} V_1 & V_2 & V_3 & V_4 & V_5 & V_6 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

The identification of a minimal cut set is useful in metabolic engineering and in the identification of drug targets (Klamt and Gilles 2004).

Minimal Cut Sets

Primal Network



$$\mathbf{N} = \begin{array}{c} \begin{matrix} R1 & R2 & R3 & R4 & R5 \end{matrix} \\ \begin{bmatrix} -1 & 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \end{matrix}$$

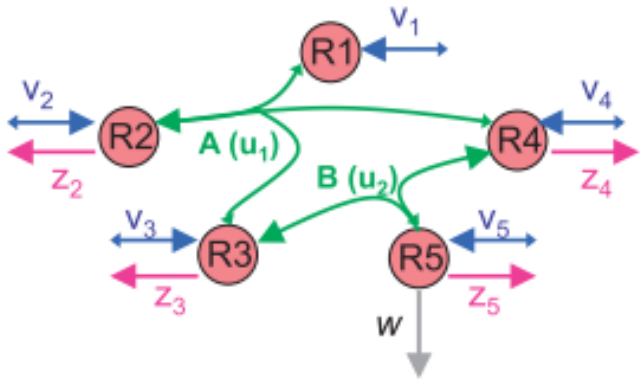
Reversible reaction: R1

Target reaction: R5

Target EMs (with R5):
- EM1={R1,R3,R5}
- EM2={R1,R4,R5}
- EM3={R2,R3,R5}
- EM4={R2,R4,R5}

MCSs for target reaction R5
- MCS1={R1,R2}
- MCS2={R3,R4}
- MCS3={R5}

Dual Network



$$\mathbf{N}_{dual} = (\mathbf{N}^T \ \mathbf{I} \ -\bar{\mathbf{I}}_{irrev} \ -\mathbf{t}) = \begin{array}{c} \begin{matrix} N_A^T & N_B^T & & \mathbf{I} & & & & & -\bar{\mathbf{I}}_{irrev} & & & & -\mathbf{t} \\ u_1 & u_2 & v_1 & v_2 & v_3 & v_4 & v_5 & z_2 & z_3 & z_4 & z_5 & w & \end{matrix} \\ \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} \end{array} \begin{matrix} R1 \\ R2 \\ R3 \\ R4 \\ R5 \end{matrix}$$

Reversible reactions: $v_1 \ldots v_5$ and u_1, u_2 (originating from metabolites A and B in $\mathbf{N}^T$)

EMs with support in w and minimal support with respect to reactions $v_1 \ldots v_5$:
- EM1={v_1,v_2,u_1,u_2,w}
- EM2={v_3,v_4,u_2,w}
- EM3={v_5,w}

MCSs hitting the three EMs above by cutting only $v_1,\ldots,v_5$:
- MCS1={v_1,v_3,v_5}
- MCS2={v_1,v_4,v_5}
- MCS3={v_2,v_3,v_5}
- MCS4={v_2,v_4,v_5}

Ballerstein et al. Bioinformatics, 2012

Limitations

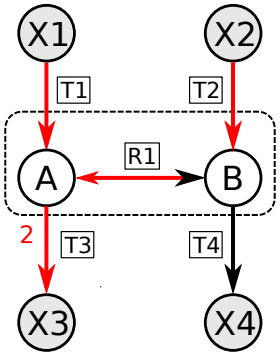
- Combinatorial explosion of the number of *EFMs*.
  - E. Coli model (Orth *et al.*, EcoSal 2009): 94 metabolites, 95 reactions, $226,3.10^6$ ***EFMs***.
  - Estimation in Human reconstructed network (Yeung *et al.*, BMC Bioinfo, 8:363, 2007): 3311 reactions, 10^{29} ***EFMs***.

- *EFMs* are not all feasible (kinetics, thermodynamics, regulations constraints).

FBA vs EFMA

FBA
Flux Balance Analysis

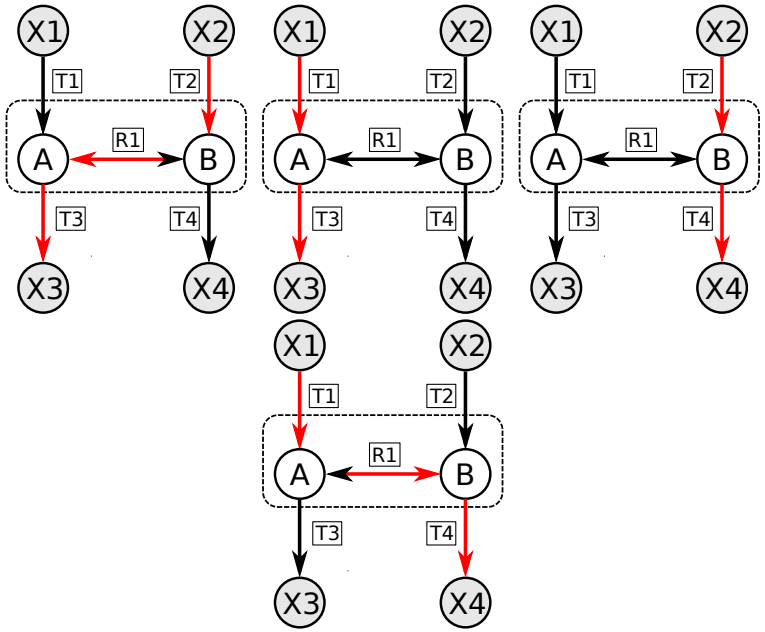
$$\begin{array}{lll} \max & C^T.v & \text{(Objective fct)} \\ \text{s. t.} & S.v = 0 & \text{(steady state)} \\ & l \leq v \leq u & \text{(flux constraints)} \end{array}$$



Solution is not unique

EFMA
Elementary Flux Mode Analysis

$P = \{v \in \mathbb{R}^r \mid S.v = 0 \text{ and } v_{irrev} \geq 0\}$
EFMs = minimal supports of P ($\subseteq$)
For $v \in P, supp(v) = \{j \in \mathbb{N} : v_j \neq 0\}$



Combinatorial explosion

EFMs and FBA are not all feasible !

Adding biological constraints to characterize more relevant solutions and to reduce the solution space

Few constraints can be integrated in the computation

- For EFMs (Double Description method), constraints must be support-monotone for the inclusion otherwise they must be treated in post-processing.
- For FBA (convex optimization), constraints must be linear

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.
Thus, if a flux distribution violates the condition C, so it is for any flux distribution with a larger support.

Biological constraints

- **Thermodynamic** : Gibbs free energy (ΔG) or equilibrium constant (K_{eq})
  Not linear(Dinh et al. 2017) but (ΔG) support-monotone for the inclusion
- **Kinetic** : $v = E.k(X, k_m, k_{cat}, ...)$
  Not linear, not support-monotone for the inclusion but optimal solutions of the enzyme allocation problem with general kinetics are *EFMs*. (Müller et al. 2014)
- **Transcriptomic regulations** : Boolean rules
  Only conjunction of negative clause are support monotone for the inclusion
- **Operating cost** : C mol glucose consumed / C mol biomass produced
  Linear constraints

Thermodynamic constraints

Binary distinction

In the traditional method for calculating *EFMs*, a binary distinction is made between reversible and irreversible reactions.

Gibbs free energy

$$\Delta_r G_j = \Delta_r G_j^0 + RT \ln \prod_i X_i^{s_{ij}}$$

where:
ΔG_j^0: the standard free energy change
X_i: the metabolite concentration
s_{ij}: stoichiometric coefficients
R: molar gas constant
T: the absolute temperature

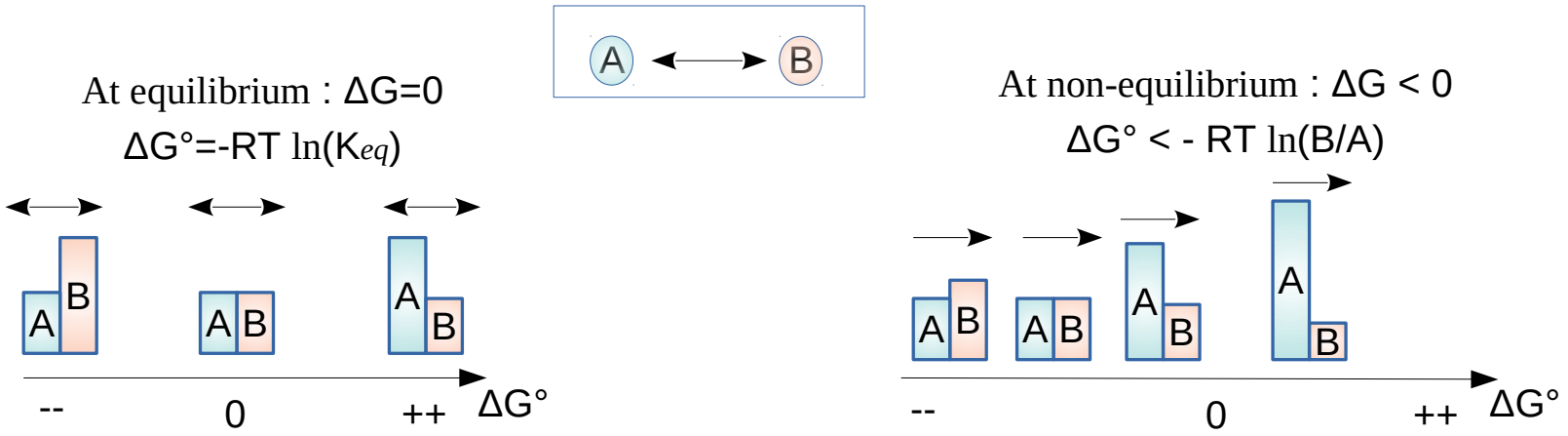
The Gibbs energy of a reaction j can be calculated from the Gibbs energies of formation of the participating reactants i:

$$\Delta_r G_j = \sum_{i=1}^{m} S_{ij} \Delta_f G_i$$

Gibbs free energy

Assuming constant pressure and a closed system, according to the second law of thermodynamics a reaction occurs only in the direction of negative Gibbs energy of reaction :

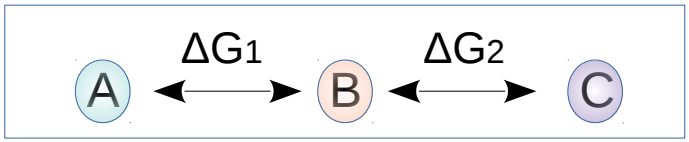
$$\Delta_r G_j = \Delta_r G_j^0 + RT \ln \prod_i X_i^{s_{ij}} < 0$$

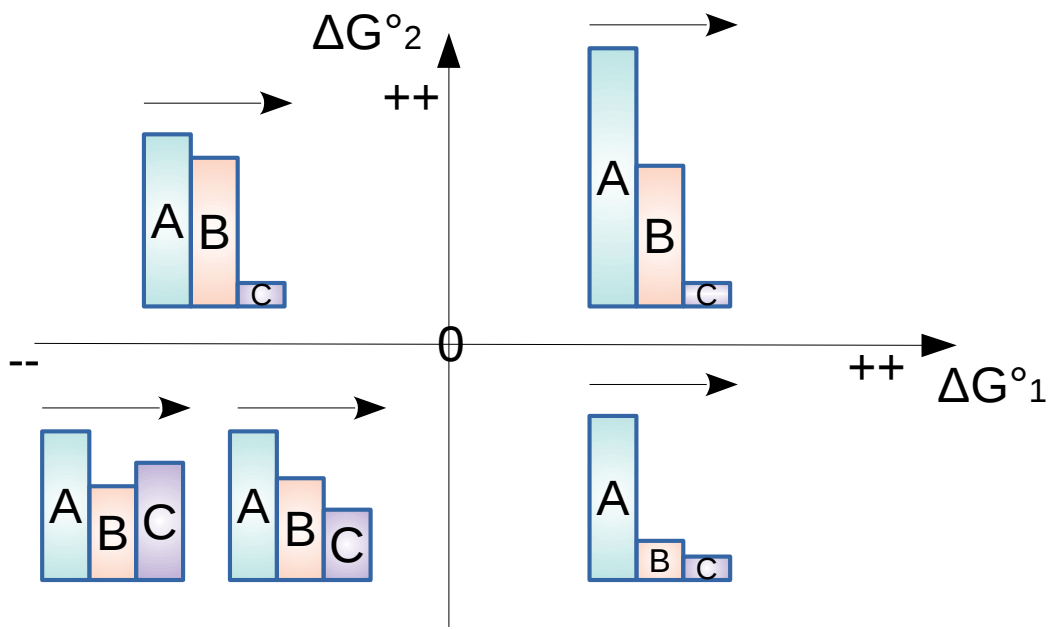


Thermodynamic constraints on a pathway (*EFM*)

The Gibbs energies of reactions are constrained by the mutual thermodynamic interdependencies of reactions in a network/pathway (e):

$$\Delta_r G_j < 0, \forall j \in supp(e)$$





Non-convexity of the general problem but support-monotone for the inclusion

In NET analysis, the Gibbs energies of reaction are constrained by the mutual thermodynamic interdependencies of reactions in a network/pathway (e):

$$\Delta_r G_j < 0, \forall j \in supp(e)$$

The Gibbs energy of a reaction j can be calculated from the Gibbs energies of formation of the participating reactants i:

$$\Delta_r G_j = \sum_{i=1}^{m} S_{ij} \Delta_f G_i$$

$$\Delta_f G_i = \Delta_f G_i^0 + RTln(X_i)$$

$\Rightarrow$ limits the feasible ranges of Gibbs energies of reaction.

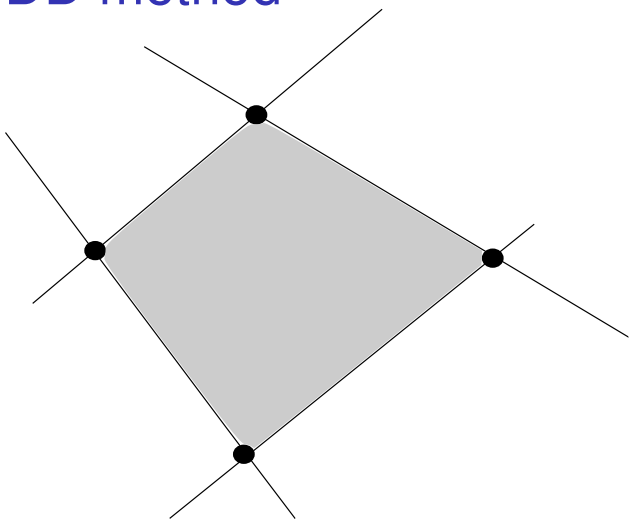
NET analysis (Kümmel *et al.*, 2006)

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.

Adding monotonic constraint in DD method

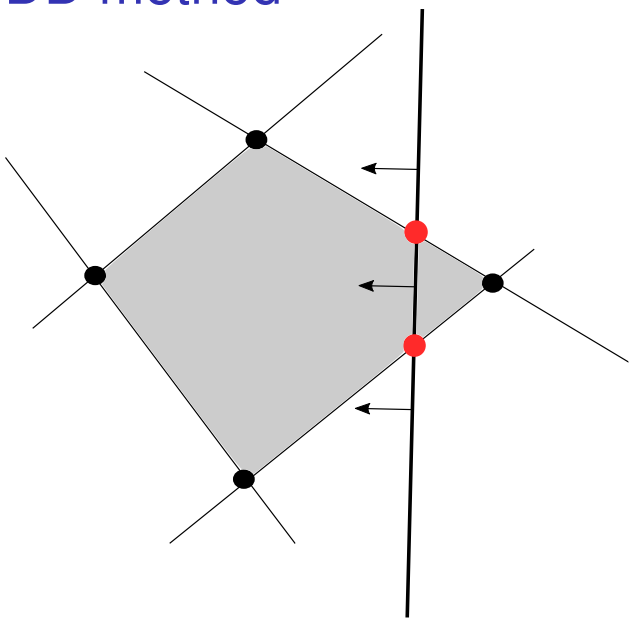


New extreme ray is a positive linear combination of 2 adjacents rays: its support is union of their 2 supports and thus larger.

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.

Adding monotonic constraint in DD method

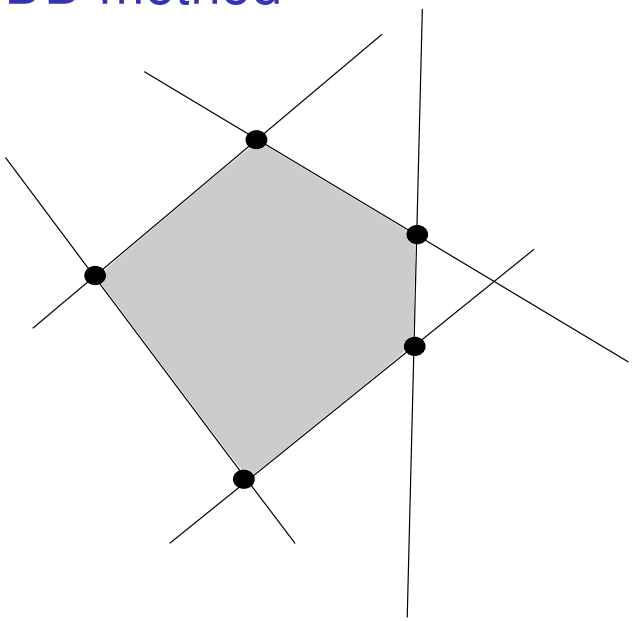


New extreme ray is a positive linear combination of 2 adjacents rays: its support is union of their 2 supports and thus larger.

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.

Adding monotonic constraint in DD method

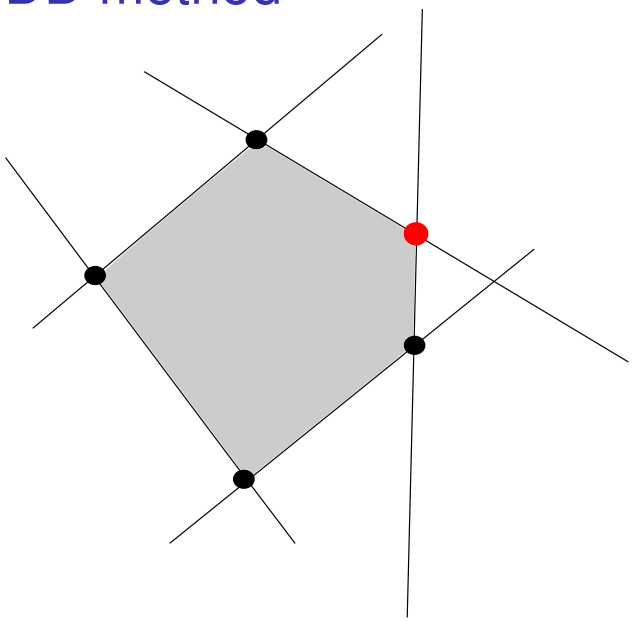


New extreme ray is a positive linear combination of 2 adjacents rays: its support is union of their 2 supports and thus larger.

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.

Adding monotonic constraint in DD method

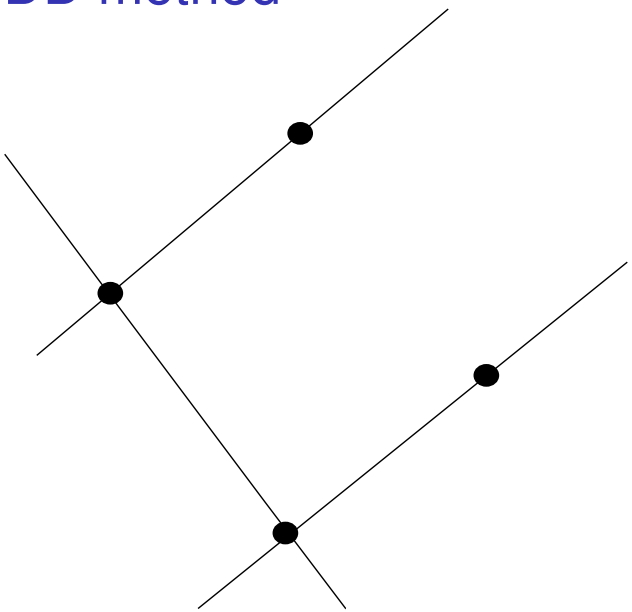


If it violated monotone constraint, it has no use for the rest of the algo and can be discarded.

Thermodynamic constraint is support-monotone for the inclusion and linear in $ln(X_i)$

A constraint C is **support-monotone for the inclusion** if:
for a given flux distribution v verifying the constraint C, then for all flux distribution v' such that $supp(v') \subset supp(v)$, v' verifies C.

Adding monotonic constraint in DD method



Thermodynamic satisfiability can be checked by an **LP program** at each iteration of the DD method (`tEFMA`).

(Gerstl et al. 2015)

Thermodynamic description

At equilibrium:

$$\frac{\prod_{i=} X_i^{s_{ij}^+}}{\prod_{i=} X_i^{s_{ij}^-}} = K_j$$

K_j: equilibrium constant

For non-equilibrium states

where the reaction j proceeds in the forward direction ($\Delta G_j < 0$) we have:

$$\frac{\prod_i X_i^{s_{ij}^+}}{\prod_i X_i^{s_{ij}^-}} < \hat{K}_j$$

where $\hat{K}_j = \frac{K_j}{\prod_i \bar{X}_i^{s_{ij}}}$ (apparent equilibrium)
X_i : internal metabolite concentrations
$\bar{X}_i$: external metabolite concentrations
s_{ij}: stoichiometric coefficients

EFMs consistent with equilibrium constants

Under the plausible assumption that $X_i > 0$ and concentrations are assumed to be dimensionless quantities after division by the unit concentration X_0: $y_i = \log \frac{X_i}{X_0}$, then

$$\sum_{i=1}^{m} s_{ij} y_i < \log \hat{K}_j$$

By using one form of Farkas duality lemma (Kuhn and Fourier theorem):
$Ax < b$ has a solution in x iff $z^t A = 0$ avec $z \geq 0 \Rightarrow z^t b > 0$

An *EFM e* is thermodynamically feasible if:

$$e^t \log \hat{K} > 0 \tag{1}$$

Compute thermodynamically feasible *EFM* without knowledge of internal metabolite concentration.

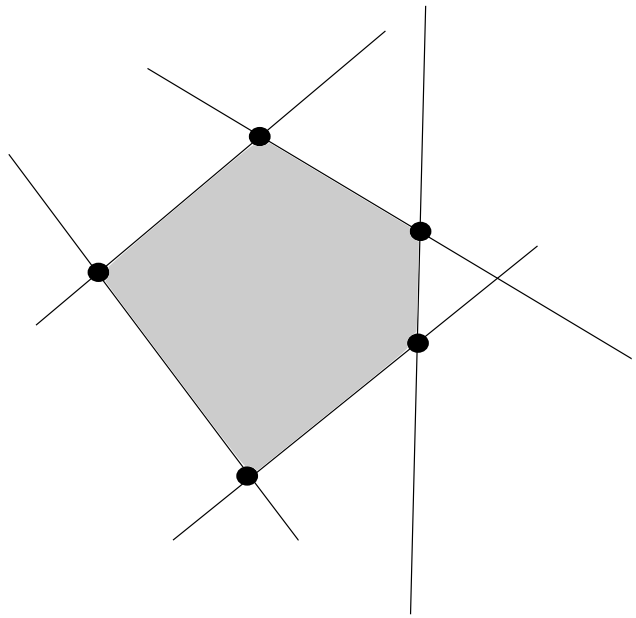
Schuster *al.* J. math. biol, 1991

Peres *al.* Plos one, 2017

Adding the thermodynamic constraint in DD method as a supplementary linear inequality

- Check the scalar product in post-processing
- the formula (1) can be directly added as a new linear inequality constraint in the DD algorithm.

Peres *al.* Bioch. Trans. Soc., 2018.

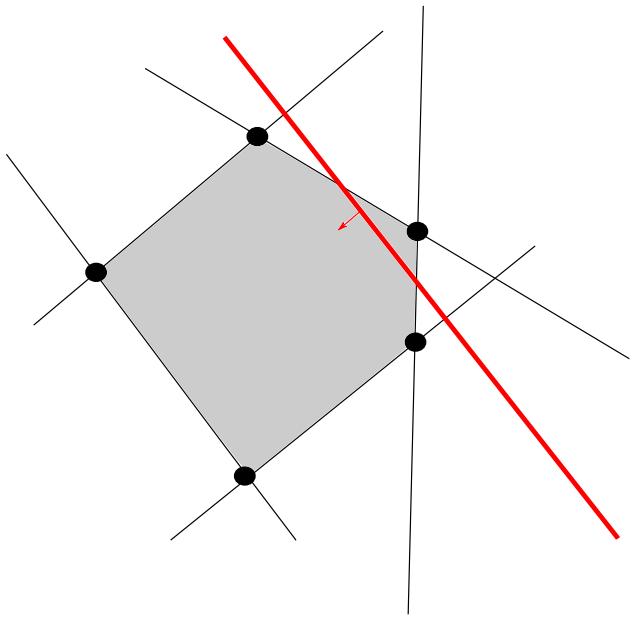


It would be necessary to be able to control the order in which the inequality constraints are processed in order to apply the thermodynamic one in first, before the irreversibility ones, to compare the performances.

Adding the thermodynamic constraint in DD method as a supplementary linear inequality

- Check the scalar product in post-processing
- the formula (1) can be directly added as a new linear inequality constraint in the DD algorithm.

Peres *al.* Bioch. Trans. Soc., 2018.

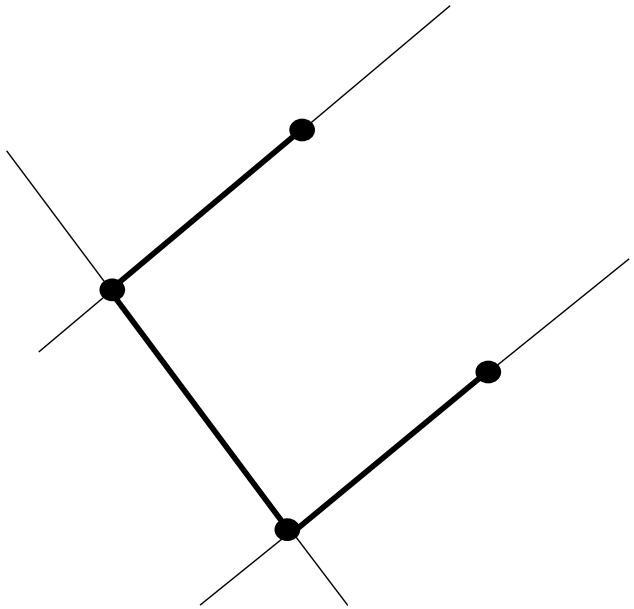


It would be necessary to be able to control the order in which the inequality constraints are processed in order to apply the thermodynamic one in first, before the irreversibility ones, to compare the performances.

Adding the thermodynamic constraint in DD method as a supplementary linear inequality

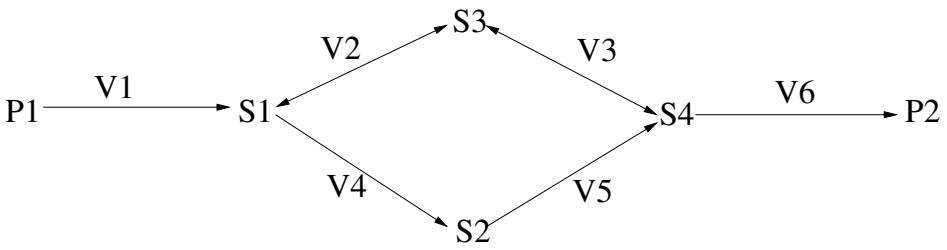
- Check the scalar product in post-processing
- the formula (1) can be directly added as a new linear inequality constraint in the DD algorithm.

Peres *al.* Bioch. Trans. Soc., 2018.



It would be necessary to be able to control the order in which the inequality constraints are processed in order to apply the thermodynamic one in first, before the irreversibility ones, to compare the performances.

Exemple



$\log\left(k'_1\right) = 5$
$\log\left(k'_2\right) = -2$
$\log\left(k'_3\right) = -1$

Add pseudo reaction v_{r+1} to obtain $\sum \log\left(\hat{K}\right) > 0$:

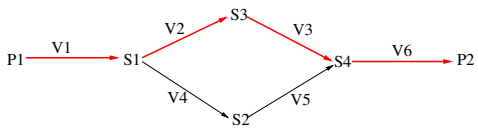
- $\sum \log\left(\hat{K}\right) - v_{r+1} \geq 0$
- $v_{r+1} > 0$

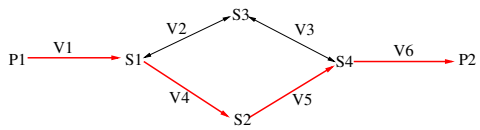
Keep solutions where v_{r+1} is present.

$A =$

R'$_1$	R'$_2$	R'$_{3f}$	R'$_{3b}$	R'$_4$
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
1	-1	-1	+1	0
-1	1	1	-1	0
-1	1	1	-1	0
1	-1	-1	+1	0
5	-2	-1	1	-1

$$R_{final} = \begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 3 & 4 \\ 3 & 4 & 0 & 0 \\ & & * & * \end{matrix}$$





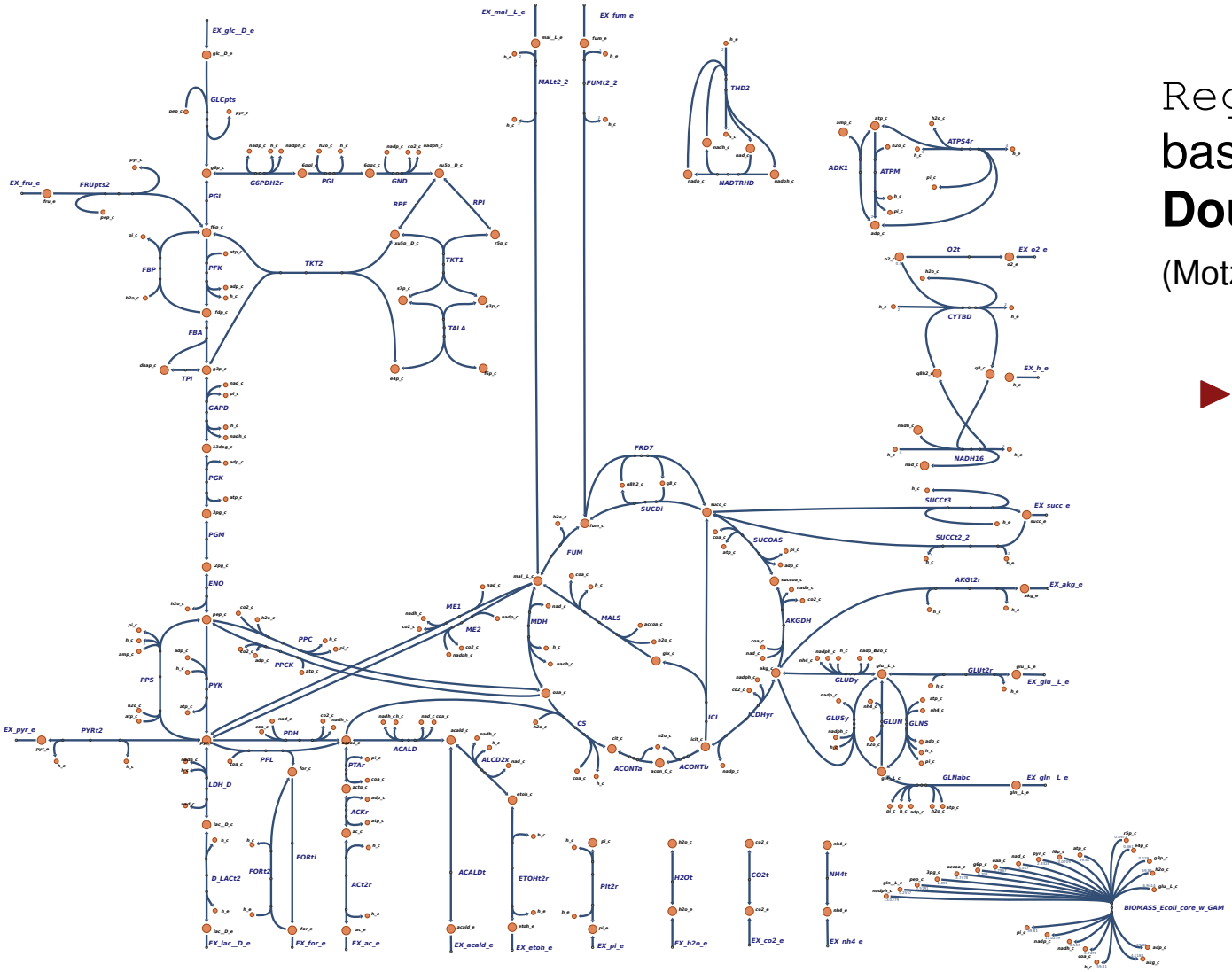
How important is thermodynamics for identifying elementary flux modes ?

Simple example: *n* monomolecular reactions

$$\bar{X}_1 \leftrightarrow X_2 \leftrightarrow \ldots \leftrightarrow \bar{X}_{n+1}$$

- Thermodynamic condition: $\frac{\bar{x}_{n+1}}{\bar{x}_1} < \prod_{j=1}^{n} \hat{K}_j$
- Non steady-state: 2^n configurations
  + thermodynamic : 2^{n-1} configurations
- steady state : 2 configurations
  + thermodynamic : 1 configuration
- Same result can be actually obtained by adding a reliable irreversibility condition for an arbitrary reaction among the *n* ones

EFMs of *E. coli* core

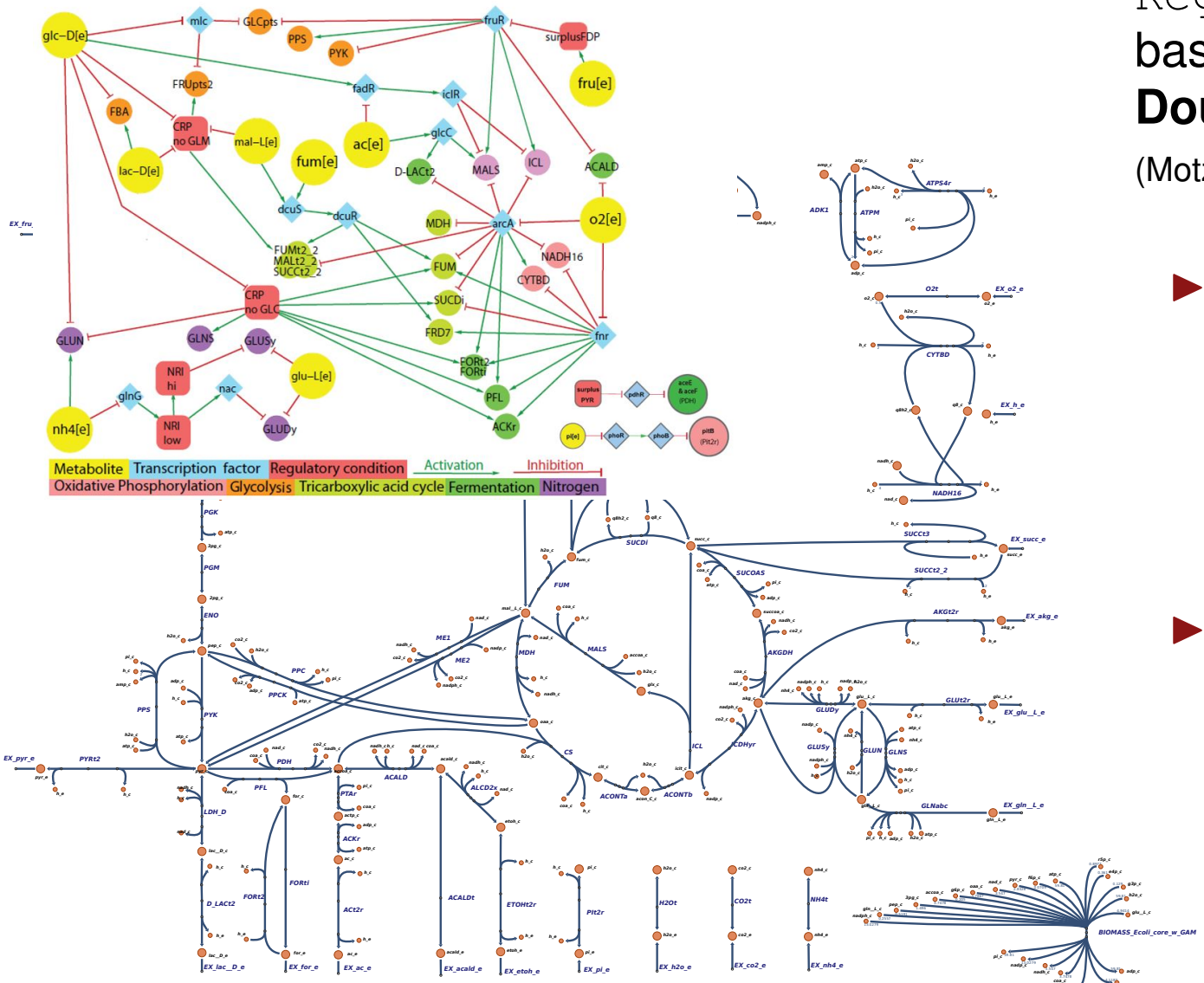


`RegEFMTool` software based on **Double Description**

(Motzkin et al., 1953)

- 94 metabolites, 95 reactions:
  EFMs: **226,3** $\times$ **10**^{6}
  Storage: 251 GB
  Times: **34.1h**

EFMs of *E. coli* core

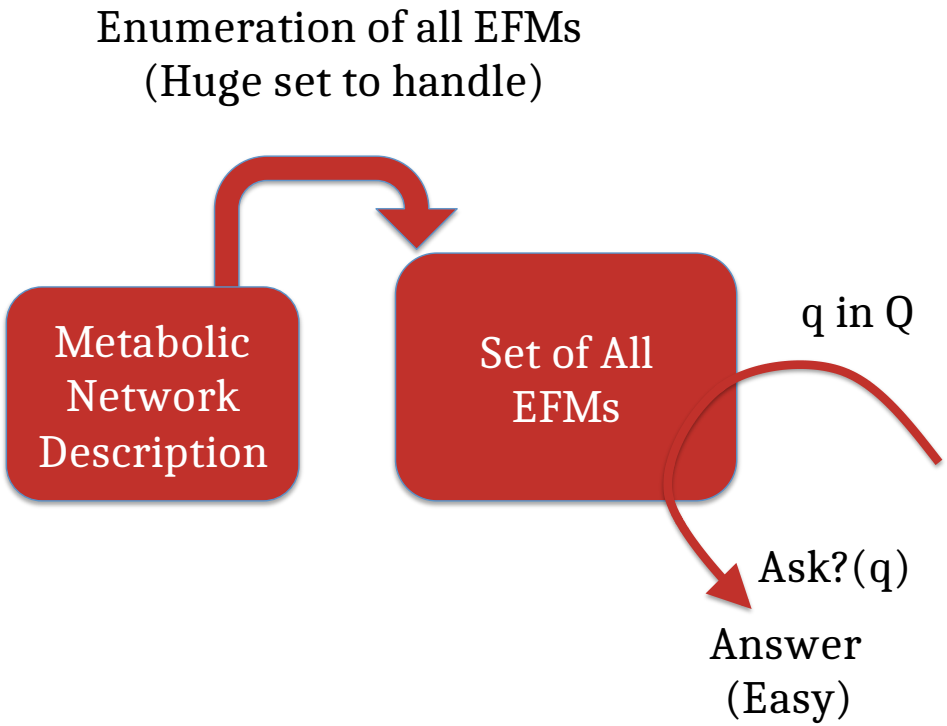


`RegEFMTool` software based on **Double Description**

(Motzkin et al., 1953)

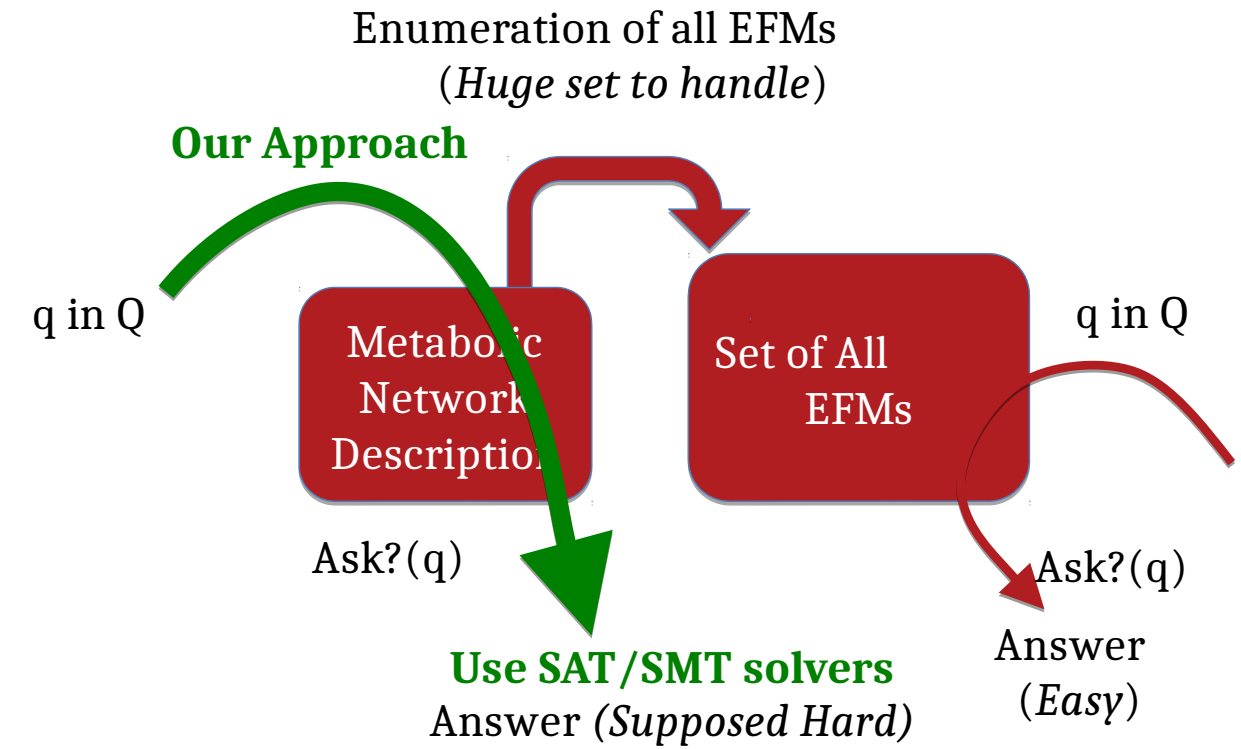
- 94 metabolites, 95 reactions:
  EFMs: **226,3** $\times$ **10**^{6}
  Storage: 251 GB
  Times: **34.1h**
- + 78 regulation rules:
  (only negative rules)
  EFMs: **2,1** $\times$ **10**^{6}
  Storage: 2.3 GB
  Times: **7.1h**

Logic-based method to compute EFMs



Q: langage of queries over EFMs

Logic-based method to compute EFMs



Q: language of queries over EFMs

SAT-based method (for *SATisfiability*) demonstrates that **new approaches for manipulating *EFMs* are possible.**

Peres *et al.*, LNBI, 8859:20-31, 2014.

Propositional Logic

Knowledge Representation

The **facts** :
are true, or false.

The **variables** :
are $\top$, or $\bot$.

Reasoning by calculus

If we know that :

- A implies B
- B implies C

The we can *deduce* that :

- A implies C

If we have :

- $\neg A \vee B$
- $\neg B \vee C$

Then, by **resolution** :

- $\neg A \vee C$

A very simple logic...

The facts are propositional variables
The knowledge is a propositional formula

$$
\begin{array}{cccccc}
 & \neg x_1 & \vee & \neg x_2 & \vee & x_3 \\
\wedge & & & & & \neg x_3 \\
\wedge & x_1 & \vee & x_2 & & \\
\wedge & & & x_2 & \vee & x_3
\end{array}
$$

- Variables : $x_1 \dots x_3$;
- Literals : x_1, $\neg x_1$;
- Clauses : $\neg x_1 \vee \neg x_2 \vee x_3$ (disjunction of literals);
- Formula Σ written in CNF (conjonction of clauses);

A very simple logic...

The facts are propositional variables
The knowledge is a propositional formula

$$
\begin{array}{cccccc}
 & \neg x_1 & \vee & \neg x_2 & \vee & x_3 \\
\wedge & & & & & \neg x_3 \\
\wedge & x_1 & \vee & x_2 & & \\
\wedge & & & x_2 & \vee & x_3
\end{array}
$$

x_1	x_2	x_3
$\bot$	$\bot$	$\bot$

- Variables : $x_1 \dots x_3$;
- Literals : x_1, $\neg x_1$;
- Clauses : $\neg x_1 \vee \neg x_2 \vee x_3$ (disjunction of literals);
- Formula Σ written in CNF (conjonction of clauses);

A very simple logic...

The facts are propositional variables
The knowledge is a propositional formula

$$
\begin{array}{cccccc}
 & \neg x_1 & \vee & \neg x_2 & \vee & x_3 \\
\wedge & & & & & \neg x_3 \\
\wedge & x_1 & \vee & x_2 & & \\
\wedge & & & x_2 & \vee & x_3
\end{array}
$$

x_1	x_2	x_3
$\bot$	$\bot$	$\bot$

- Variables : $x_1 \dots x_3$;
- Literals : x_1, $\neg x_1$;
- Clauses : $\neg x_1 \vee \neg x_2 \vee x_3$ (disjunction of literals);
- Formula Σ written in CNF (conjonction of clauses);

A very simple logic...

The facts are propositional variables
The knowledge is a propositional formula

$$
\begin{array}{cccccc}
 & \neg x_1 & \vee & \neg x_2 & \vee & x_3 \\
\wedge & & & & & \neg x_3 \\
\wedge & x_1 & \vee & x_2 & & \\
\wedge & & & x_2 & \vee & x_3
\end{array}
$$

x_1	x_2	x_3
$\bot$	$\top$	$\bot$

- Variables : $x_1 \dots x_3$;
- Literals : x_1, $\neg x_1$;
- Clauses : $\neg x_1 \vee \neg x_2 \vee x_3$ (disjunction of literals);
- Formula Σ written in CNF (conjonction of clauses);

BIG QUESTIONS

- **SAT** : is there an assignment of variables that makes the formula true?
- **UNSAT**: is the theory contradictory?

A very simple logic...

The facts are propositional variables
The knowledge is a propositional formula

$$\begin{array}{cccccc} & \neg x_1 & \vee & \neg x_2 & \vee & x_3 \\ \wedge & & & & & \neg x_3 \\ \wedge & x_1 & \vee & x_2 & & \\ \wedge & & & x_2 & \vee & x_3 \end{array}$$

x_1	x_2	x_3
$\bot$	$\top$	$\bot$

- Variables : $x_1 \dots x_3$;
- Literals : x_1, $\neg x_1$;
- Clauses : $\neg x_1 \vee \neg x_2 \vee x_3$ (disjunction of literals);
- Formula Σ written in CNF (conjonction of clauses);

BIG QUESTIONS

- **SAT** : is there an assignment of variables that makes the formula true?
- **UNSAT**: is the theory contradictory?

MODERN SOLVER

Conflict-driven clause learning (CDCL) $\rightarrow$ very efficient

SMTtool: SMT approach for *EFM*

A SMT solver (for *Satisfiability modulo theories*): decision problem for logical formulas, with respect of a theory (Integer, Real, linear...)
$\Rightarrow$ improve expressivity and computation performance wrt SAT.

Principles of Satisfiability Modulo Theory solvers (SMT)

- The SAT solver works on an **abstraction** of the initial problem
- It sends candidates solutions S_0 to a **more powerful reasoner (SMT)** that checks their validity at the theory level
- Then, theory solver **enriches** the SAT solver with new abstract knowledge

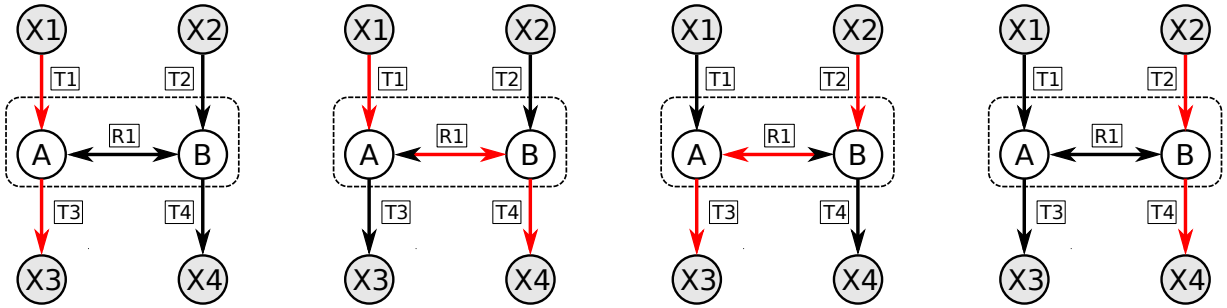
$\Rightarrow$ Use **LRA** (Linear Real Arithmetic) theory that allows us to use linear operations on real numbers with solver CVC4 to compute **EFM with Boolean constraints**.

Morterol *et al.*, WCB, 12:65-81, 2016.

Minimal Constrained Flux Modes (MCFM)

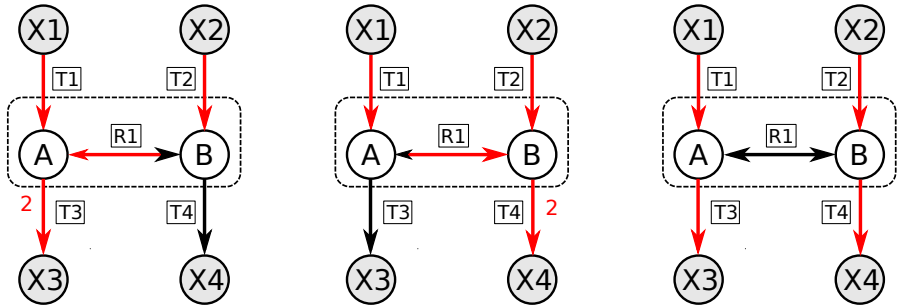
EFMs = minimal supports (for inclusion) of P

$$supp(v) = \{j \in \mathbb{N} : v_j \neq 0\}$$



A valid pathway is not always a combination of "valid" EFMs.

Example: EFMs which contains T1 and T2 ($T1 \wedge T2$)



$\Rightarrow$ Can be discard by a kernel test ($dim(Ker(N^{Supp(Sol)})) = 1$)

Morterol *et al.*, WCB, 12:65-81, 2016.

Answer Set Programming (ASP): simple language to knowledge representation and reasoning

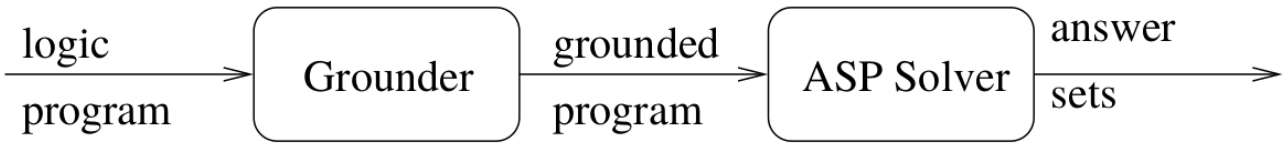
ASP use **first-order logic** (an extension of propositional logic)

i.e. with the addition of quantifiers $\forall$ and $\exists$

- Allows to define problems on sets of values
- Closed World Assumption - What is unknown is false by default

Clingo: ASP Solvers

- **Clingo** is the most famous solver for ASP
  - composed of Gringo (Grounder) et Clasp (Solver)



- Clingo uses the solving approaches of **modern SAT-solvers**
  - Based on conflict-driven clause learning (CDCL) algorithm
- Clingo has been extended to **make theory solving**
  - Clingo[LP] : Linear Programming (Real, Int) - CPLEX

Jahunen *et al.* Theory and Practice of Logic Programming, 2017

ASP logical rules

A logic program in ASP is a finite set of rules of the form:

$$\underbrace{K\{a_1;...;a_r\}L}_{head} \underbrace{:-}_{\Leftarrow} \underbrace{b_1;...;b_m,\neg c_{m+1};...;\neg c_n}_{body}$$

where a_i, b_i, c_i are atoms.
If all terms in body are true then at least K and at most L terms are true on the head.

$:-b_1;...;b_m;\neg c_{m+1};...;\neg c_n.$

If nothing on the head,
then **always false**.

Constraint

$K\{a_1;...;a_r\}L.$

If nothing on the body,
then **always true**.

Fact

Example

- We write a program with constraints and rules in natural language, and the Answer Set Programming solver will find all solutions
- What is unknown is false by default, **Closed World Assumption** (reasoning will work on known and unknown)

```
anaerobic :- not oxygen.

oxygen :- not anaerobic.

respiration.

:- respiration, not oxygen.
```

Example

- We write a program with constraints and rules in natural language, and the Answer Set Programming solver will find all solutions
- What is unknown is false by default, **Closed World Assumption** (reasoning will work on known and unknown)

```
anaerobic :- not oxygen.
```
% anaerobic state is achieved when there is no oxygen present

```
oxygen :- not anaerobic.
```
% oxygen can be present if we are not in the anaerobic state

```
respiration.
```
% respiration is required to be in the returned solutions

```
:- respiration, not oxygen.
```
% not possible to have the respiration without oxygen

Example

- We write a program with constraints and rules in natural language, and the Answer Set Programming solver will find all solutions
- What is unknown is false by default, **Closed World Assumption** (reasoning will work on known and unknown)

```
anaerobic :- not oxygen.
```
% anaerobic state is achieved when there is no oxygen present

```
oxygen :- not anaerobic.
```
% oxygen can be present if we are not in the anaerobic state

```
respiration.
```
% respiration is required to be in the returned solutions

```
:- respiration, not oxygen.
```
% not possible to have the respiration without oxygen

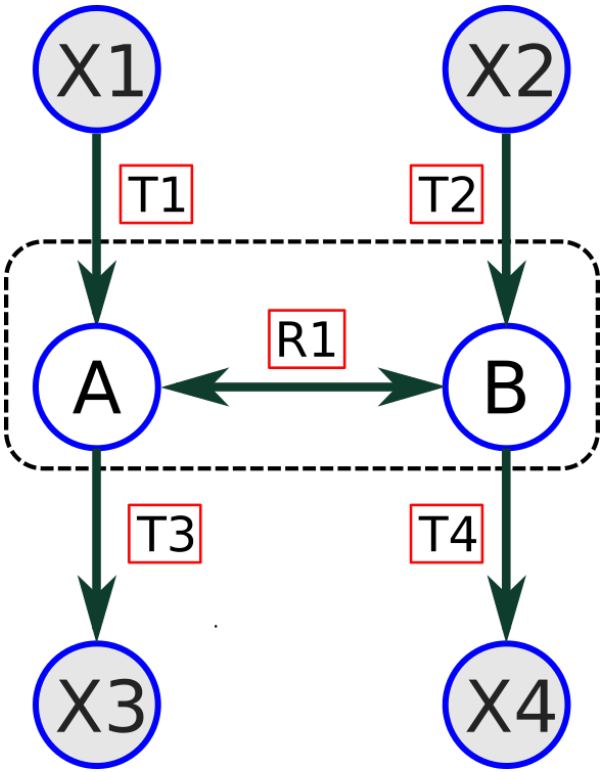
Solution: **respiration and oxygen**

Encoding metabolic network in ASP

Metabolic networks are expressed with ASP in the form of first-order predicates.

$$\begin{aligned}
\mathit{reaction}(r) \quad & \forall r \in R, \\
\mathit{metabolite}(m) \quad & \forall m \in M, \\
\mathit{stoichiometry}(m, r, c) \quad & \forall m, r, c \text{ such that } S_{m,r} = c, \\
c \in \mathbb{R} \text{ with } S \text{ stoichiometric matrix of size } & |R| \times |M|, \\
\mathit{reversible}(r) \quad & \forall r \in \mathit{Rev} \subset R, \\
\mathit{gene}(g) \quad & \forall g \in G.
\end{aligned}$$

Example




```
reaction(t1).
reaction(t2).
reaction(t3).
reaction(t4).
reaction(r1).
metabolite(a).
metabolite(b).
stoichiometry(a, t1, 1).
stoichiometry(b, t2, 1).
stoichiometry(a, t3, -1).
stoichiometry(b, t4, -1).
stoichiometry(a, r1, -1).
stoichiometry(b, r1, 1).
reversible(r1).
```

aspefm: Computing EFMs with ASP

- *aspefm* consists of a declarative logic program in clingo[LP] syntax.
- We represent the **flux of a reaction** r by the variable v_r and if it is active by the **boolean indicator** variable $z_r \in \{0,1\}$

ASP program rules:

- Split reversible reactions are split: $\forall j,\ v_j \geq 0$
- Only one direction for a reaction: $\neg z_r \vee \neg z_{r_{rev}} \quad \forall (r, r_{rev}) \in Rev$
- Exclude trivial solutions: $\bigvee_{r \in R} z_r$
- **Steady-state**: $S.v = 0$
- **support** is 1 when a reaction has a positive flux: $\forall j,\ z_j \leftrightarrow v_j > 0$
- and finally **heuristics** to get answer sets with **subset-minimal** support

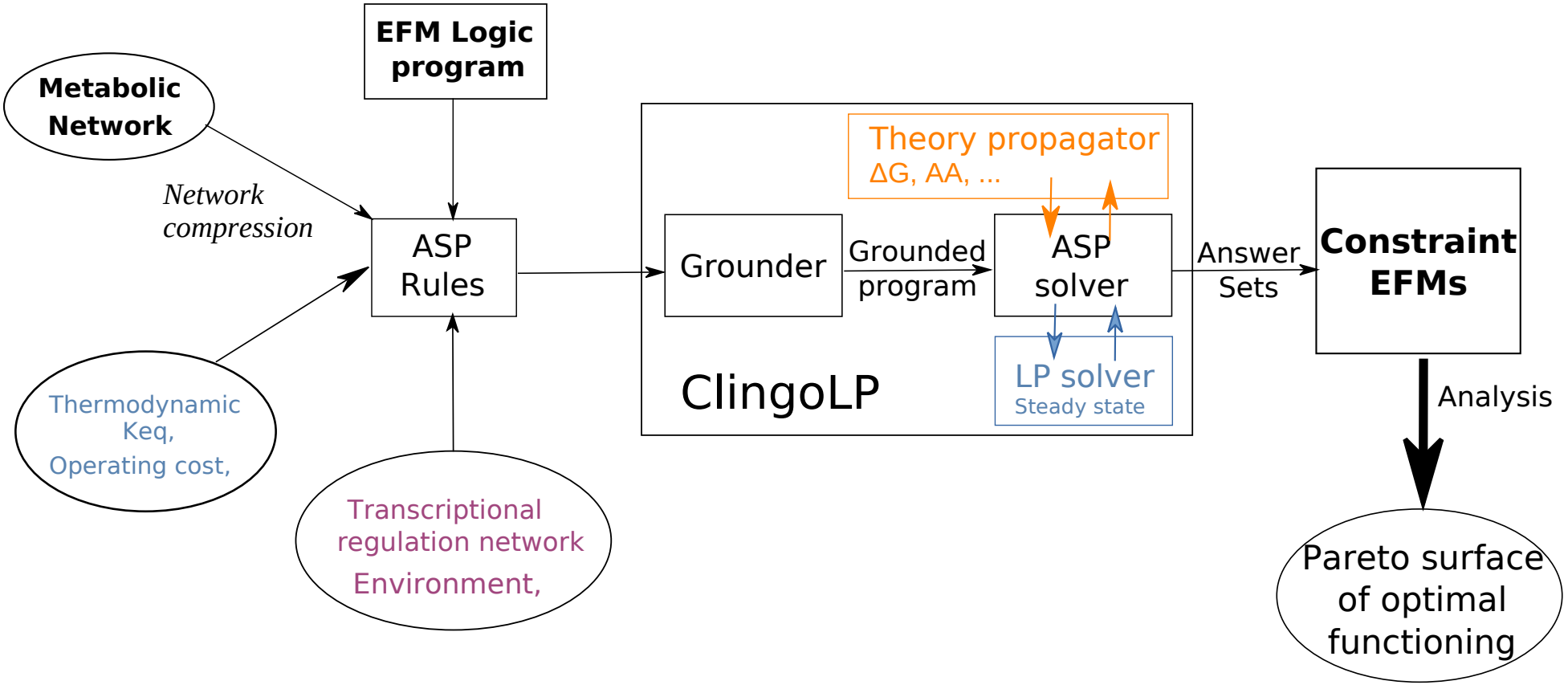
Mahout, Carlson & Peres. *Processes* 2020

Biological constraints

aspefm can compute **subsets** of EFMs respecting any additional constraints on metabolic networks. **logical**, **linear** and **theory propagator**.

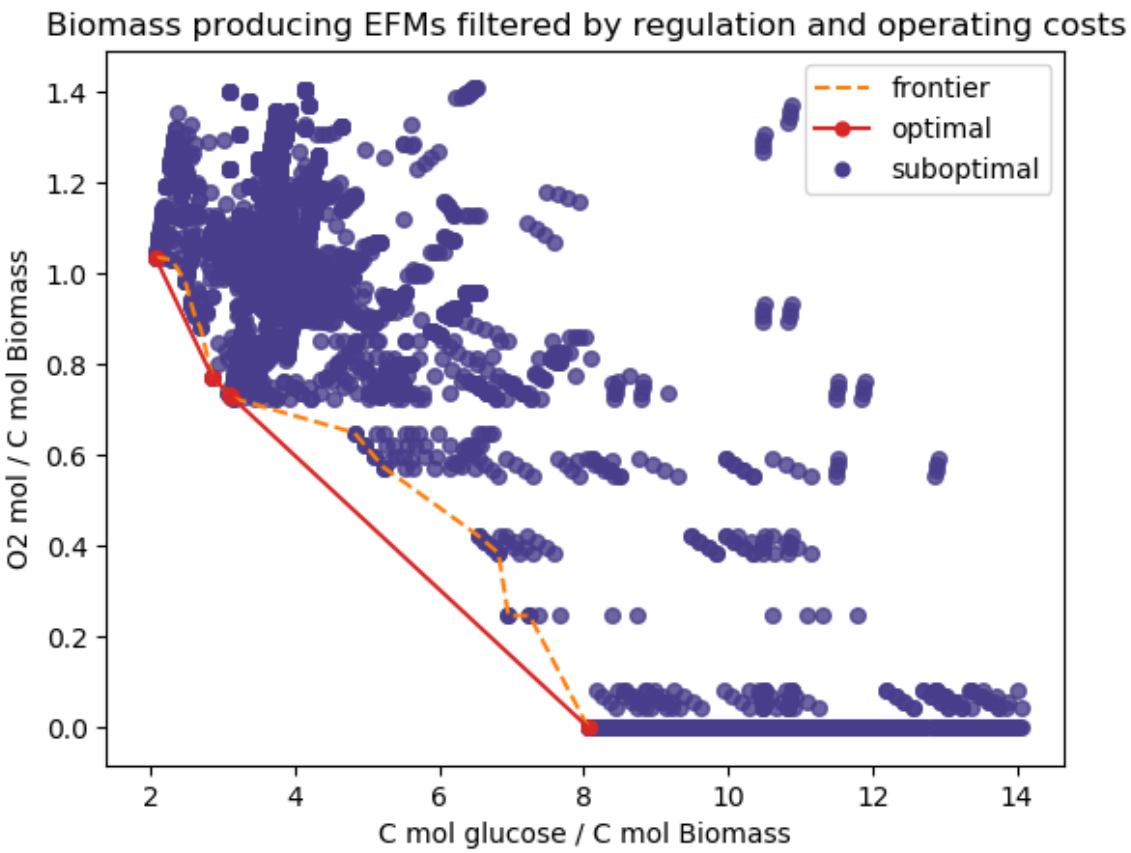
- Logical constraints on reactions, *ex:* $v_{biomass} > 0$
- Regulation rules, *ex:* $z_r \rightarrow b_{regulator}$ $(alt. \neg z_r \vee b_{regulator})$
- Environment, *ex:* $z_{oxy\_transport} \rightarrow b_{oxy\_env}$
- Equilibrium constant, *ex:* $log(k_{eq1})\ v_{oxygen} + ... + log(k_{eq2})\ v_{biomass} > 0$
- Operating cost, *ex:* $v_{oxygen} < K\ v_{biomass}$
- Filter out the cycles, (i.e all EFMs which do not use at least a transporter)
- Investment cost, eliminate EFMs overusing Amino Acid resources.
- Thermodynamics, eliminate EFMs inconsistent with Free Gibbs energies

ASPefm: Boolean and linear constraints with ClingoLP



Mahout, Carlson & Peres. *Processes* 2020

ASPefm to enumerate EFMs
in biologically reasonnable range



- Thermodynamics (Keq)
- Regulatory contraints (78 rules)
- Medium : Glucose
- Growth: biomass production
- Operating cost:
  $v_{oxygen} < 30\ v_{biomass}$
  $6 v_{glucose} < 300\ v_{biomass}$

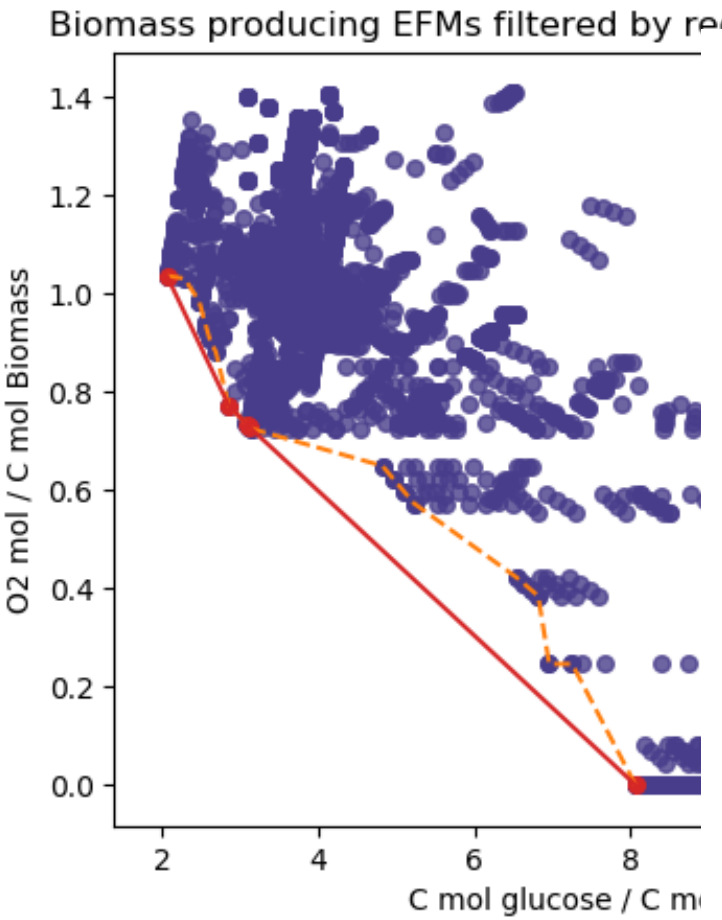
- Aerobic : 1118 EFMs
- Anaerobic : 363 EFMs

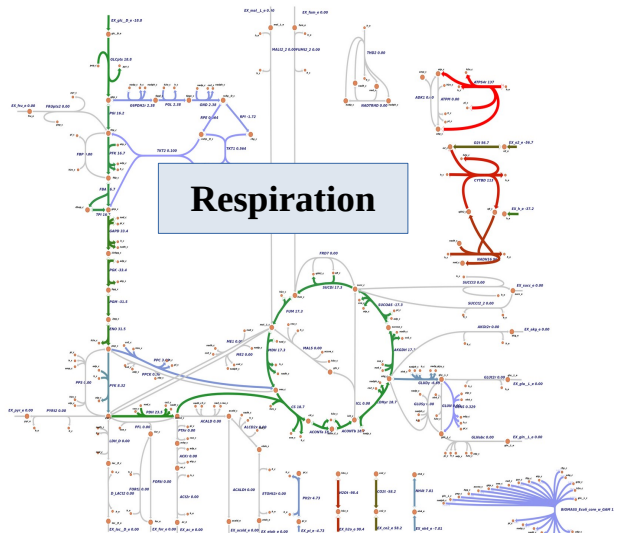
- **Times: 2 min !**

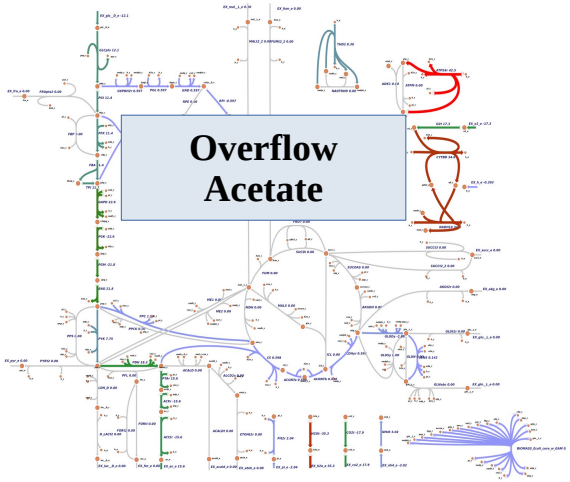
Operating cost identify the most efficient EFMs for converting substrates into biomass

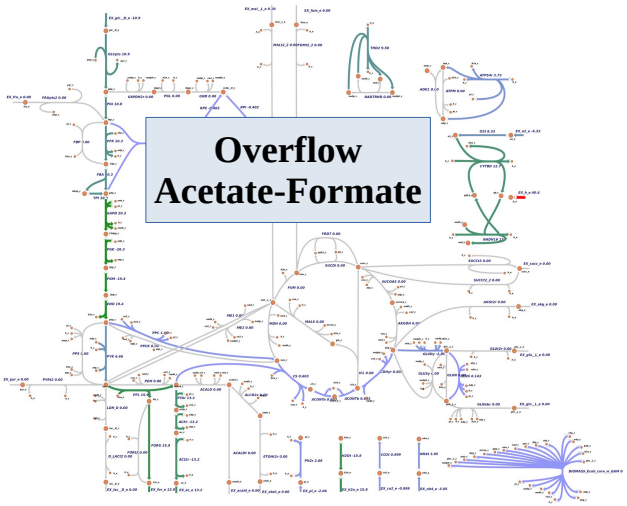
Convex hull $\Rightarrow$ Optimal phenotypes for growth on glucose and of O2 availability

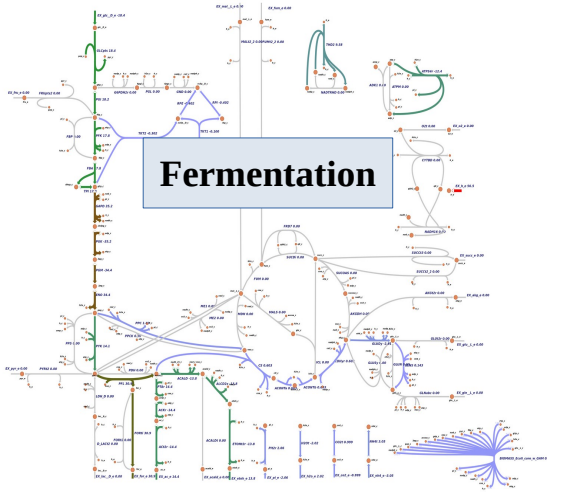
Identification of experimental known efficient phenotypes











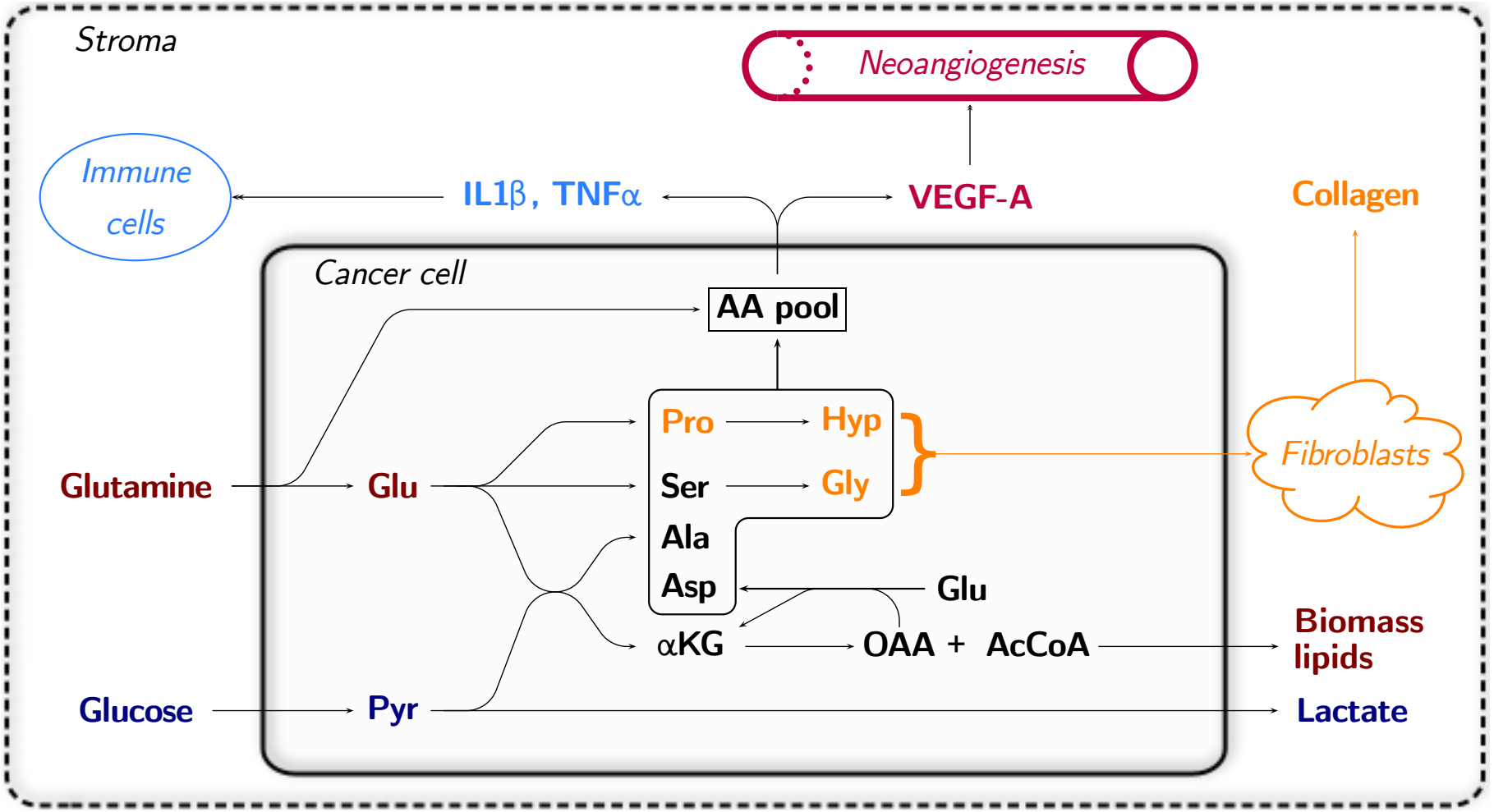
Scaling up to genome-scale model ?

- The model is small scale and contains only 95 reactions.
- In comparison, genome-scale models have thousands of reactions.

*B*ut...

- *aspefm* shows that the analysis of optimal pathways reserved for small models is possible on a larger scale by integrating sufficient constraints.
- Once their metabolic network is compressed for calculating EFMs, these genome-scale models have only $\approx$ 500 reactions, which *aspefm* can handle.

EFMs show a Correlation of neoangiogenesis, collagen production, and inflammation to Warburg effect in cancer



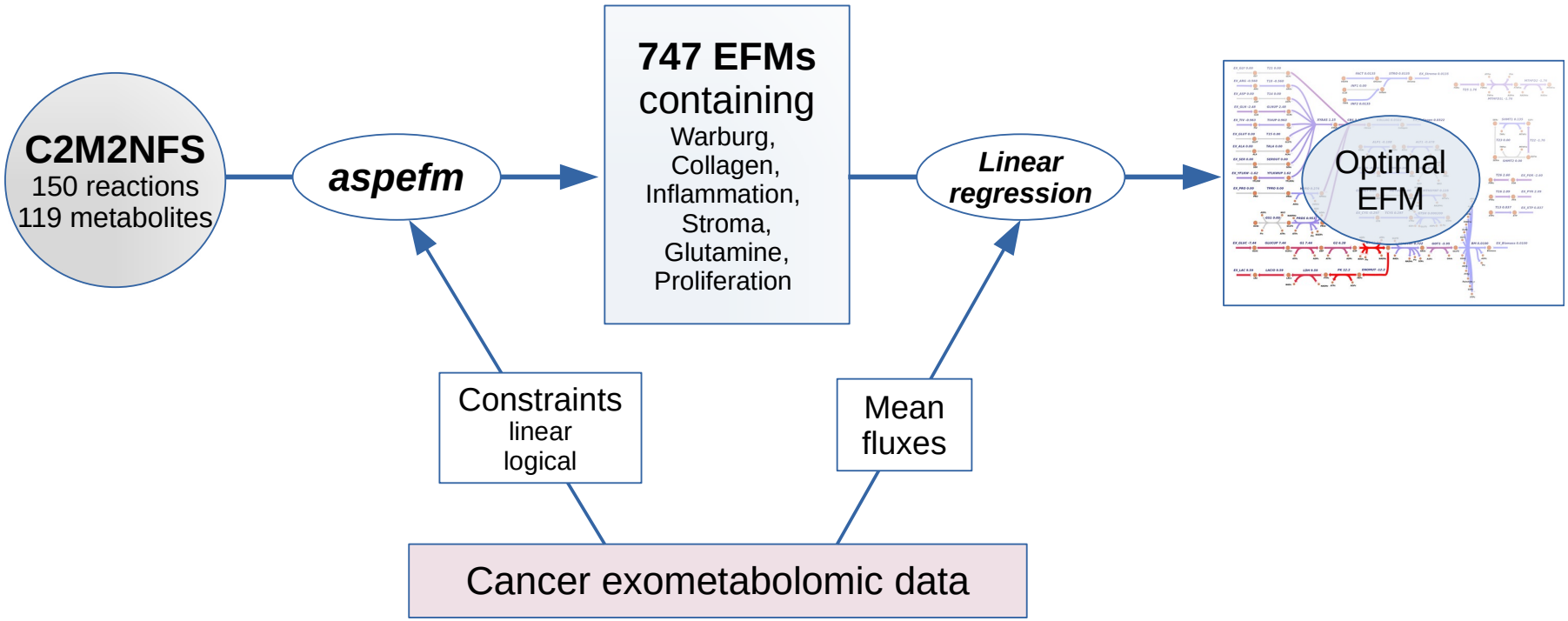
M. Mahout, L. Schwartz, R. Attal, A. Bakkar, S. Peres, submitted 2023.

Exometabolomic data

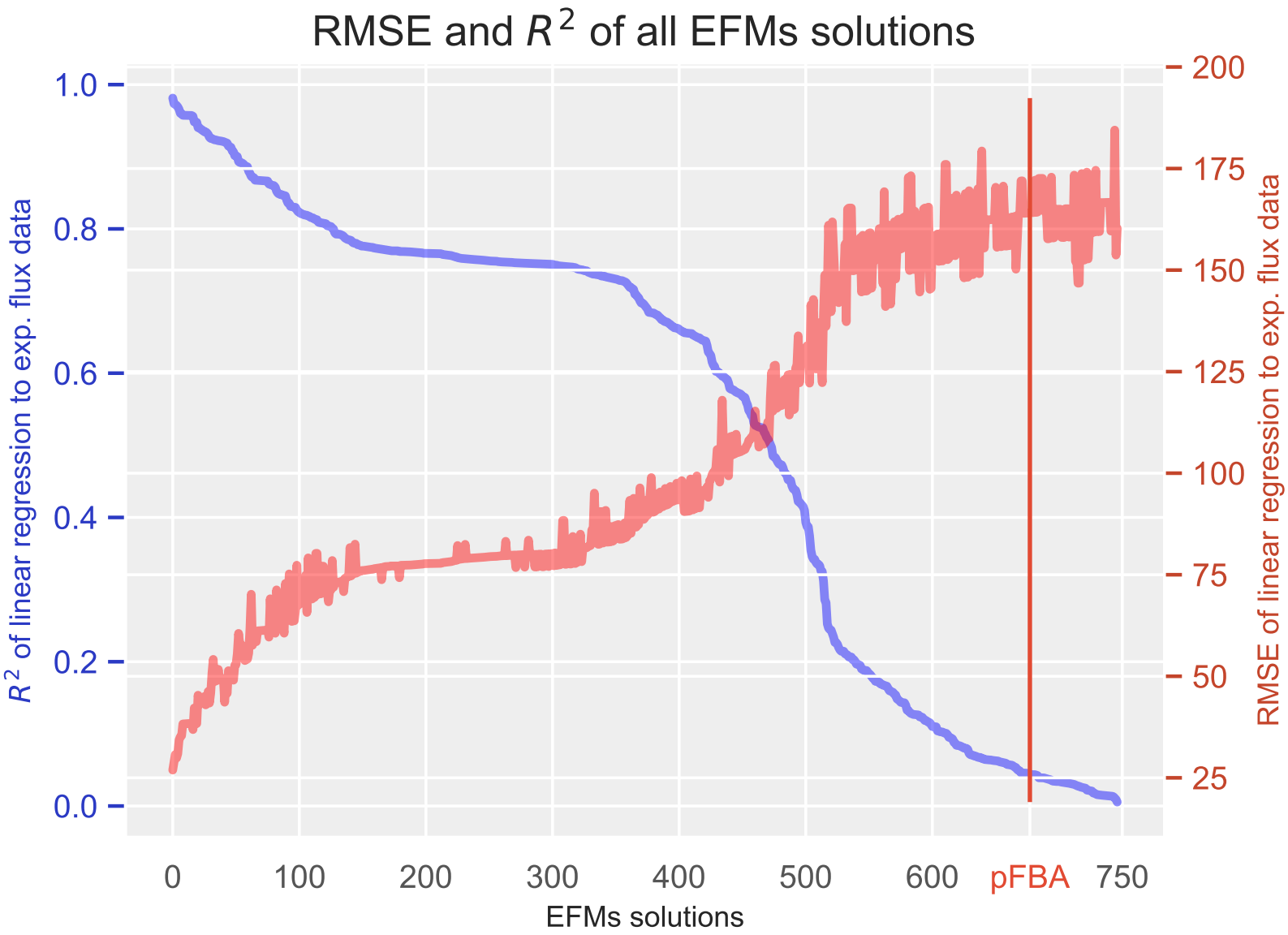
Metabolite	Glucose	Lactate	Glutamine	Glutamate	Serine	Glycine	Alanine	Proline	Asp Asn	Arginine
Mean +/- SD cancer cell exchange flux interval	-326.87 +/- 196.12	442.20 +/- 289.40	-82.48 +/- 56.20	13.54 +/- 16.99	-11.57 +/- 7.05	0.96 +/- 2.97	15.89 +/- 13.34	1.21 +/ 1.49	-3.33 +/- 3.39	-4.90 +/- 4.44
Experimental observations	–	+	–	+	-	+/-	+	+/-	-	-

Metabolite	TIV (Thr, Ile, Val)	YFLKW (Tyr, Phe, Leu, Lys, Trp)	XTP (Nucleotides)	Pyruvate	Formate	Histidine	Cysteine	Methionine
Mean +/- SD cancer cell exchange flux interval	-14.90 +/- 7.99	-19.80 +/- 10.57	0.10 +/- 0.22	Uncalibrated data	Data missing	Data missing	0.05 +/- 0.08	-2.11 +/- 1.23
Expected observations	-	-	+/-	+	+/-	-	+/-	+/-

Méthode



- Hard constraints are encoded as direct logical constraints: "reaction must be active" : $\forall R \in$ Hardconstraints, $v_{R_f} > 0 \Leftrightarrow Z_{R_f}$
- Expected observations (+) and (-) are encoded by forbidding reactions going in the opposite direction : $\forall R \in$ ExpectedObs, $v_{R_b} = 0 \Leftrightarrow \neg Z_{R_b}$


RMSE and R^2 of all EFMs solutions
R^2 of linear regression to exp. flux data
RMSE of linear regression to exp. flux data
1.0
0.8
0.6
0.4
0.2
0.0
200
175
150
125
100
75
50
25
0
100
200
300
400
500
600
pFBA
750
EFMs solutions