

# Cours métabolisme

Sabine Peres

BioRegul - Porquerolles - 25 juin 2019



- 1 Modélisation de réseaux métaboliques
- 2 Modes élémentaires
- 3 Logic-based methods
- 4 Contraintes thermodynamique

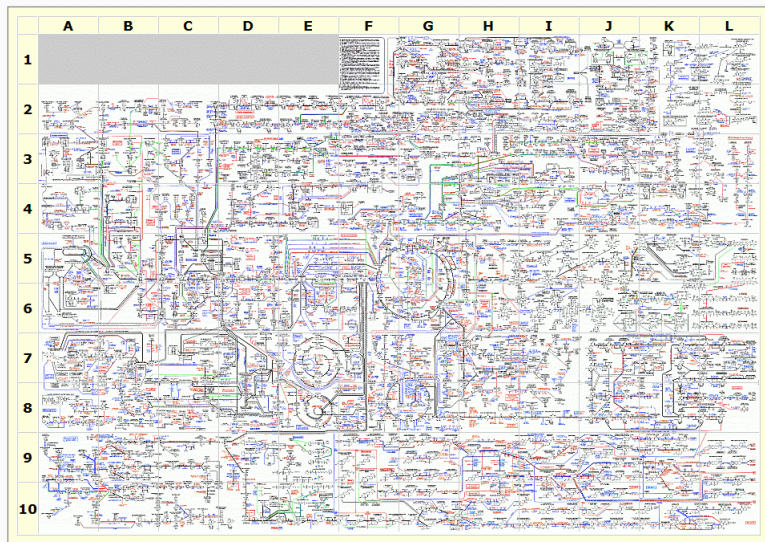
## 1 Modélisation de réseaux métaboliques

## 2 Modes élémentaires

## 3 Logic-based methods

## 4 Contraintes thermodynamique

# Qu'est-ce qu'une voie dans réseau métabolique?



# Le métabolisme est un système complexe

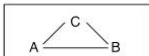
Il est impossible de prévoir son comportement par la seule connaissance de ses composants élémentaires.

Il est nécessaire de le modéliser pour mieux le comprendre.

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

# Modélisation du métabolisme

(a) Interaction-based



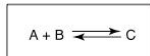
Static models

No stoichiometry

No parameters



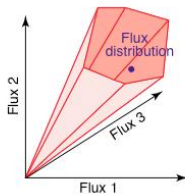
(b) Constraint-based



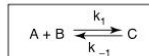
Static models

Stoichiometry

No parameters



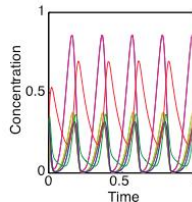
(c) Mechanism-based



Dynamic models

Stoichiometry

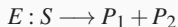
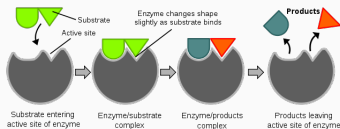
Kinetic parameters



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

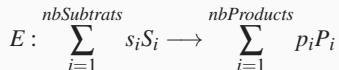
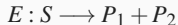
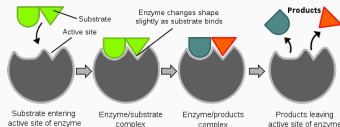
## Formalisation des réactions enzymatiques

- 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 relargage de l'enzyme libre.
- Exemple:  $E + S \longrightarrow ES \longrightarrow E + P_1 + P_2$



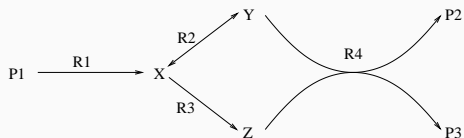
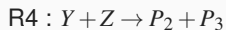
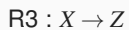
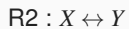
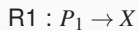
## Formalisation des réactions enzymatiques

- 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 relargage de l'enzyme libre.
- Exemple:  $E + S \longrightarrow ES \longrightarrow E + P_1 + P_2$





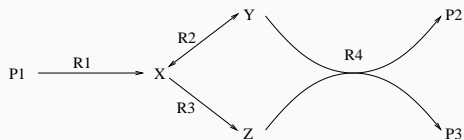
# Représentation des réseaux métaboliques par hypergraphes orientés



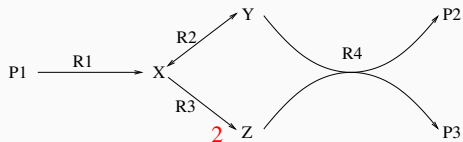
## Matrice de stœchiométrie

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 :

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

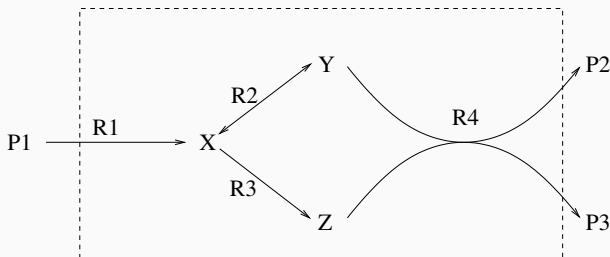


$$N = \begin{array}{cccc|l} & R1 & R2 & R3 & R4 & \\ \left( \begin{array}{cccc} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right) & & & & \begin{array}{l} X \\ Y \\ Z \\ P1 \\ P2 \\ P3 \end{array} \end{array}$$



$$N = \begin{pmatrix} & R1 & R2 & R3 & R4 \\ \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 2 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} & X \\ & Y \\ & Z \\ & P1 \\ & P2 \\ & P3 \end{pmatrix}$$

## Bornes du système



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

$$\begin{pmatrix} N_{int} \\ \hline N_{ext} \end{pmatrix} = \begin{pmatrix} R1 & R2 & R3 & R4 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ \hline -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} X \\ Y \\ Z \\ P1 \\ P2 \\ P3 \end{matrix}$$

## Dynamique d'un réseau métabolique

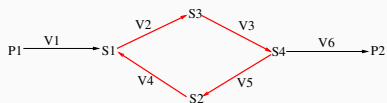
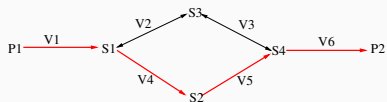
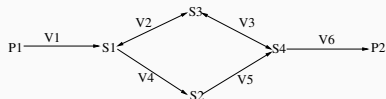
- Description des **transformations biochimiques** du réseau.

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

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

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

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

Exemple  $\ker N$ 

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

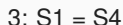
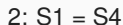
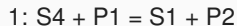
# Reaction correlations

## 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.

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

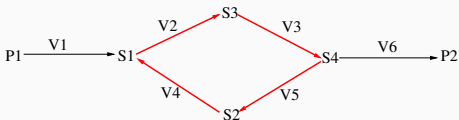
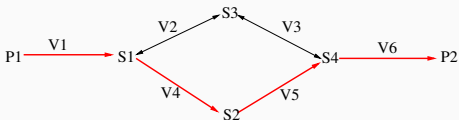
### overall reaction





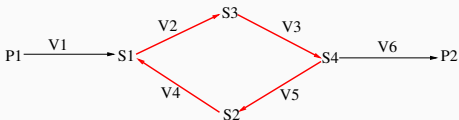
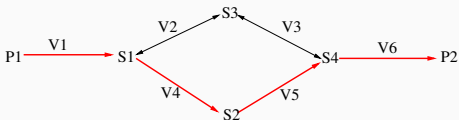
## Contraintes sur l'espace de solutions

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



## Contraintes sur l'espace de solutions

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



→ Il faut tenir compte de l'irréversibilité de certaines réactions.

## Constraint-based modeling (CBM)

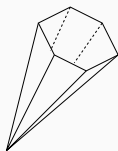
- Contraintes stœchiométriques :

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

- Constraints thermodynamiques :

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

⇒ Espace de solutions :



Cône convexe polyédrique

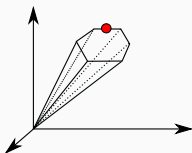


# CBM: deux approches

## FBA

### Flux Balance Analysis

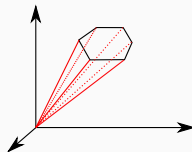
$$\begin{array}{ll}
 \max & C^T \cdot v \quad (\text{Objective fct}) \\
 \text{s. t.} & S \cdot v = 0 \quad (\text{steady state}) \\
 & l \leq v \leq u \quad (\text{flux constraints})
 \end{array}$$



## EFMA

### Elementary Flux Mode Analysis

EFMs = minimal supports (for inclusion) of  $P$   
 For  $v \in P$ ,  $\text{supp}(v) = \{j \in \mathbb{N} : v_j \neq 0\}$



- 1 Modélisation de réseaux métaboliques
- 2 Modes élémentaires**
- 3 Logic-based methods
- 4 Contraintes thermodynamique

## Définition des modes élémentaires

Le cône peut être décrit par un ensemble de vecteurs "non décomposables" que l'on appelle **Elementary Flux Modes (EFM)**.

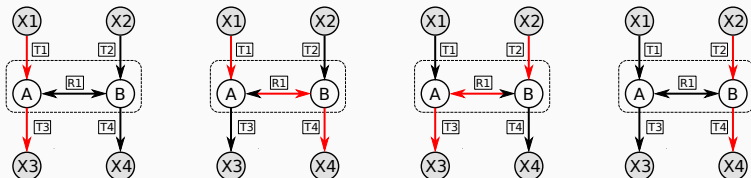
Un vecteur  $e = (e_1, \dots, e_r)^t \in \mathbb{R}^r$  est un *mode élémentaire* (EFM) s'il vérifie les conditions suivantes :

1. *état stationnaire* :  $Ne = 0$ .
2. *faisabilité* : Pour chaque indice  $j$  d'une réaction irréversible  $e_j \geq 0$ .
3. *minimalité* : Soit  $\text{supp}(v) = \{j \in \mathbb{N} : v_j \neq 0\}$ . Pour chaque *efm*  $e'$  de  $N$ ,  $\text{supp}(e') \subseteq \text{supp}(e) \Rightarrow \exists \alpha \in \mathbb{R}$  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.

## Exemple

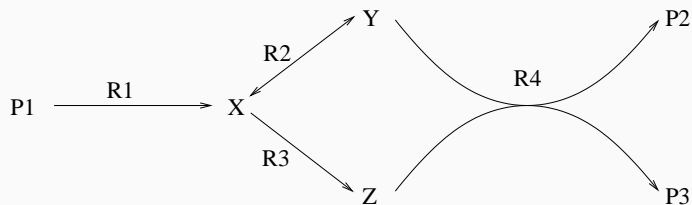
**EFMs** = **support minimal vectors of the flux cone** that contains all feasible steady-state flux vectors of a given metabolic network



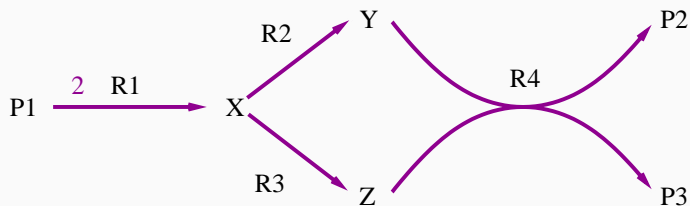
All other possible flux distributions:  $r = \sum_{j=1}^4 \alpha_j \cdot e^j$  with  $\alpha_j \geq 0$



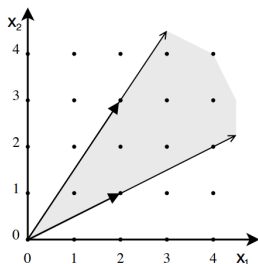
## Exemple



## Exemple



Consider the following cone  $\mathcal{C}$  and its two representations:



- $\mathcal{C} = \{x \in \mathbb{R}^2 : 3x_1 - 2x_2 \geq 0, -x_1 + 2x_2 \geq 0\}$ .
- $\mathcal{C} = \{\lambda_1(2, 1) + \lambda_2(2, 3) : \lambda_1, \lambda_2 \in \mathbb{R}_+\}$ .

## 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  $P$  :  $P = \{x \in \mathbb{R}^d | x = Ry \text{ avec } y \geq 0\}$ .

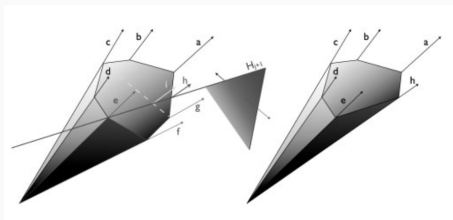
- On appelle la paire  $(A, R)$  une « paire de double description » ou « **DD paire** ». La réciproque de ce théorème est également vraie (théorème de Weyl).

## Double Description (DD) method

- Soit  $A \in \mathbb{R}^{m \times d}$  et  $P(A) = \{x : A \cdot x \geq 0\}$
- 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:

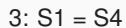
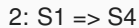
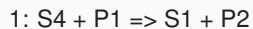
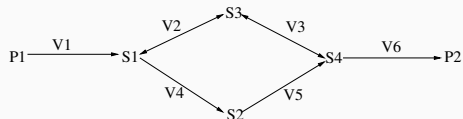
$$N.V = 0 \Rightarrow N.V \geq 0 \ \& \ -N.V \geq 0$$

- Contraintes irréversibilités :

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

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

# Exemple



métabolites/enzymes	R1	R2	R3	R3 <sub>rev</sub>
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

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<sup>eme</sup> ligne.  $A_K$  devient donc :

$$A_{1..5} = \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & -1 & -1 & 1 \\ \hline -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{array}$$

$$R_5 = \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}$$

On cherche toutes les combinaisons linéaires de façon à ce les nouvelles arêtes soient dans l'hyperplan:  $c_1+c_2$ ,  $c_1+c_3$ ,  $c_2+c_4$ ,  $c_3+c_4$

On supprime  $c_2$  et  $c_3$  qui sont dans la partie négative de l'espace coupé par l'hyperplan et on ajoute les nouvelles

# Mise à jour A et R

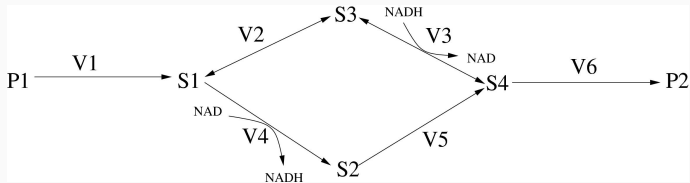
$$A_{1..6} = \begin{array}{cc|cccc} 1 & +1 & 0 & 0 & 0 & 0 \\ \hline -1 & -1 & 0 & 0 & 0 & 0 \\ \hline -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & +1 & 0 & 0 & 0 & 0 \end{array}$$

$$R_6 = \begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ & * & * & & & & & * \\ * & & & * & & & & \end{array}$$

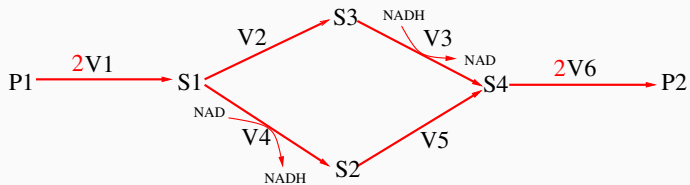
3 *EFMs*: R1-R2, R1-R3, R2-R3



## Exemple 2



## Exemple 2



## 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)



## 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

These algorithms still cannot cope with genome-scale metabolic networks reconstructly recently ( $\geq 1\ 000$  reactions).

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

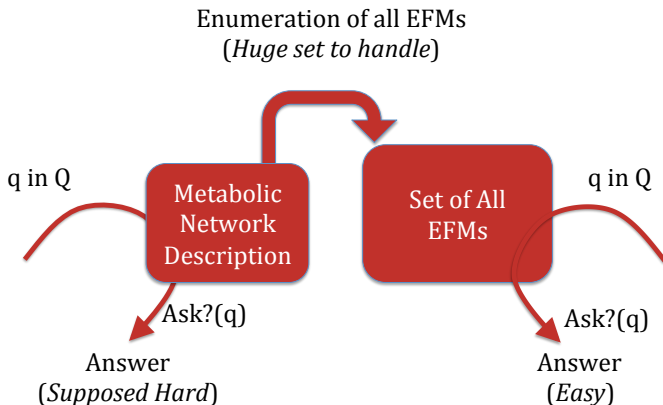
- Identification de voies métaboliques.
- Robustesse du réseau (si la production d'un métabolite est toujours possible après suppression de réactions).
- Importance des réactions (fréquence des réactions).
- Corrélation des réactions.
- Délétions létales minimales (minimal cut set).
- Hybride cybernetic modelling (HCM).

## Limitations

- Combinatorial explosion of the number of *EFMs*.
  - E. Coli model (Orth *et al.*, EcoSal 2009): 94 metabolites, 95 reactions,  $226,3 \cdot 10^6$  *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).

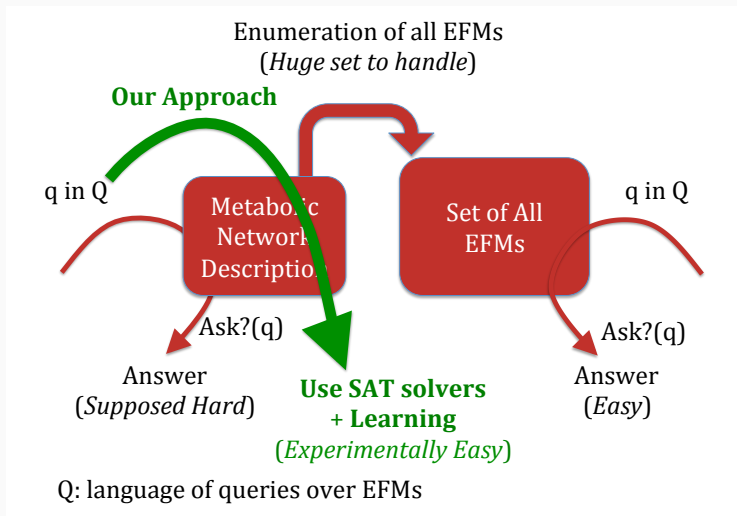
- 1 Modélisation de réseaux métaboliques
- 2 Modes élémentaires
- 3 Logic-based methods**
- 4 Contraintes thermodynamique

## SAT-based approaches



Q: language of queries over EFMs

## SAT-based approaches



# SMTtool: SMT approach for *EFM*

## Logic-based method to compute *EFMs*

$$\begin{array}{l}
 \neg x_1 \quad \vee \quad \neg x_2 \quad \vee \quad x_3 \\
 \wedge \\
 x_1 \quad \vee \quad x_2 \\
 \wedge \\
 x_2 \quad \vee \quad x_3 \\
 \wedge \\
 \neg x_3
 \end{array}$$

- 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  $\Sigma$  written in CNF (conjunction of clauses);

**SAT** : is there an assignment of variables that makes the formula true?

Encoding metabolic pathway in SAT (for *SATisfiability*) demonstrates that new approaches for manipulating *EFMs* are possible.

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

A SMT solver (for *Satisfiability modulo theories*): decision problem for logical formulas, with respect of a theory (Linear Real Arithmetic)

$\Rightarrow$  improve expressivity and computation performance wrt SAT.

# SMTtool: SMT approach for *EFM*

## Logic-based method to compute *EFMs*

$$\begin{array}{l}
 \neg x_1 \quad \vee \quad \neg x_2 \quad \vee \quad x_3 \\
 \wedge \\
 x_1 \quad \vee \quad x_2 \quad \quad \quad \neg x_3 \\
 \wedge \\
 \quad \quad \quad x_2 \quad \vee \quad x_3
 \end{array}$$

$x_1$	$x_2$	$x_3$
$\perp$	$\perp$	$\perp$

- 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  $\Sigma$  written in CNF (conjunction of clauses);

**SAT** : is there an assignment of variables that makes the formula true?

Encoding metabolic pathway in SAT (for *SATisfiability*) demonstrates that new approaches for manipulating *EFMs* are possible.

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

A SMT solver (for *Satisfiability modulo theories*): decision problem for logical formulas, with respect of a theory (Linear Real Arithmetic)

$\Rightarrow$  improve expressivity and computation performance wrt SAT.



# SMTtool: SMT approach for *EFM*

## Logic-based method to compute *EFMs*

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

$x_1$	$x_2$	$x_3$
$\perp$	$\perp$	$\perp$

- 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  $\Sigma$  written in CNF (conjunction of clauses);

**SAT** : is there an assignment of variables that makes the formula true?

Encoding metabolic pathway in SAT (for *SATisfiability*) demonstrates that new approaches for manipulating *EFMs* are possible.

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

A SMT solver (for *Satisfiability modulo theories*): decision problem for logical formulas, with respect of a theory (Linear Real Arithmetic)

$\Rightarrow$  improve expressivity and computation performance wrt SAT.

# SMTtool: SMT approach for *EFM*

## Logic-based method to compute *EFMs*

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

$x_1$	$x_2$	$x_3$
$\perp$	$\top$	$\perp$

- 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  $\Sigma$  written in CNF (conjunction of clauses);

**SAT** : is there an assignment of variables that makes the formula true?

Encoding metabolic pathway in SAT (for *SATisfiability*) demonstrates that new approaches for manipulating *EFMs* are possible.

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

A SMT solver (for *Satisfiability modulo theories*): decision problem for logical formulas, with respect of a theory (Linear Real Arithmetic)

$\Rightarrow$  improve expressivity and computation performance wrt SAT.

## SMT approach for *EFM*

A SMT solver (for « Satisfiability modulo theories »): decision problem for logical formulas, with respect of a theory (Integer, Real, linear...)

We use :

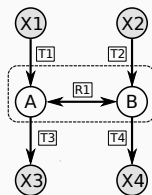
- CVC4: <http://cvc4.cs.nyu.edu>
- **LRA** (Linear Real Arithmetic) theory that allows us to use linear operations on real numbers

Morterol, Dague, Peres & Simon. WCB, 12:65-81, 2016.

## Encoding of *EFM*

Only irreversible reactions:

- So they are all positive:  $\forall r \in R, r \geq 0$ .
- A reaction and its reverse cannot be active at the same time:  
 $\forall r \in rev(R), (r > 0) \Rightarrow (rev(r) = 0)$



In our example:

- $T_1 \geq 0, T_2 \geq 0, T_3 \geq 0, T_4 \geq 0, R_1 \geq 0, R_{1rev} \geq 0$ .
- $(R_1 > 0) \Rightarrow (R_{1rev} = 0), (R_{1rev} > 0) \Rightarrow (R_1 = 0)$ .

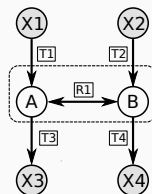
## Encoding of *EFM*

### Constraint on metabolites

- Metabolites cannot be negative:  $\forall m \in M, m \geq 0$
- There must be at least one internal metabolite:  $\sum_{m \in M} m > 0$

### In our example :

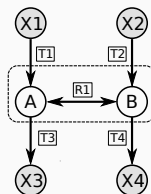
- $A \geq 0, B \geq 0$ .
- $A + B > 0$ .



## Encoding of EFM

### EFMs constraints

- Steady state :  $\forall m \in M, (m > 0) \Rightarrow$   
 $((m = \sum_{r \in \text{cons}(m,R)} S(m,r) \times r) \wedge$   
 $(m = \sum_{r \in \text{prod}(m,R)} S(m,r) \times r))$
- Inactive metabolites cannot be used by reactions:  
 $\forall m \in M, (m = 0) \Rightarrow (\forall r \in \text{cons}(m,R) r = 0, \forall r \in \text{prod}(m,R) r = 0)$



### In our example :

- $(A > 0) \Rightarrow ((A = T1 + R1_{rev}) \wedge (A = T3 + R1))$ .
- $(B > 0) \Rightarrow ((B = T2 + R1) \wedge (B = T4 + R1_{rev}))$ .
- $(A = 0) \Rightarrow (T1 = 0 \wedge T3 = 0 \wedge R1 = 0 \wedge R1_{rev} = 0)$ .
- $(B = 0) \Rightarrow (T2 = 0 \wedge T4 = 0 \wedge R1 = 0 \wedge R1_{rev} = 0)$ .

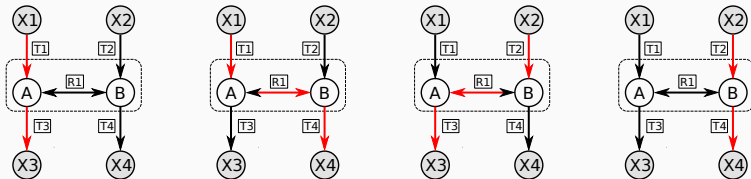
## Minimality

- Now a solution is an *EFM* or a *EFMs* positive combination.
- We need to ensure the minimality.

### Minimality with matrix calculation

- If  $\dim(\text{kernel}(N^{\text{Supp}(S_0)})) = 1 \Rightarrow S_0$  is an *EFM*.
- if not, remove at least one reaction in  $S_0$  to make  $S_1$

EFM =



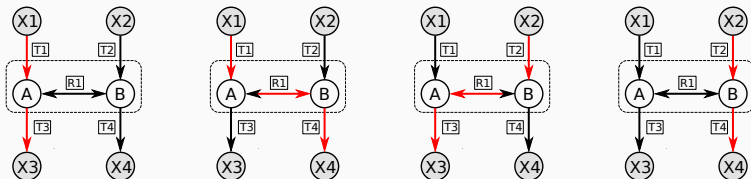
## Transcriptomic regulatory constraints in MPA

Boolean regulatory constraints are considered as propositional variables:

$$v_i \Rightarrow v_i > 0 \text{ and } \neg v_i \Rightarrow v_i = 0$$

- A valid pathway is not always a combination of valid *EFMs*.
- Example: *EFMs* which contains T1 and T2 ( $T1 \wedge T2$ )

EFM =





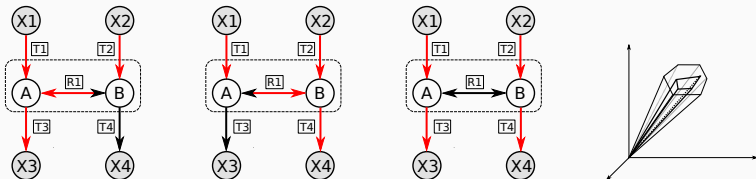
## Transcriptomic regulatory constraints in MPA

Boolean regulatory constraints are considered as propositional variables:

$$v_i \Rightarrow v_i > 0 \text{ and } \neg v_i \Rightarrow v_i = 0$$

- A valid pathway is not always a combination of valid *EFMs*.
- Example: *EFMs* which contains T1 and T2 ( $T1 \wedge T2$ )

MCFM = Minimal Constrained Flux Mode



- 1 Modélisation de réseaux métaboliques
- 2 Modes élémentaires
- 3 Logic-based methods
- 4 Contraintes thermodynamique

## 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:

$\Delta_r G_j^0$ : 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  $i$ :

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

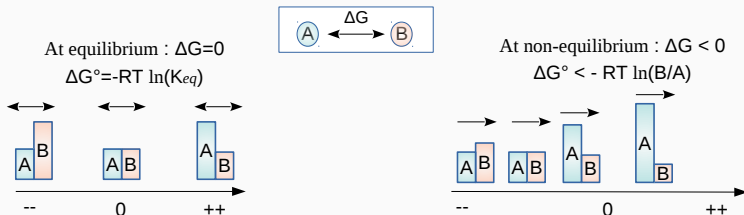
$$\Delta_f G_i = \Delta_f G_i^0 + R.T. \ln(x_i/x_0)$$

NIST database, eQuilibrium etc.

# 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 :

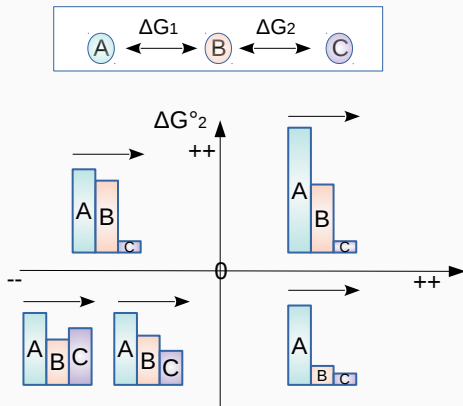
$$\Delta_r G_j = \Delta_r G_j^0 + RT \ln \prod_i X_i^{S_{ij}} < 0$$



## Thermodynamic constraints on a pathway (EFM)

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

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



## Non-convexity of the general problem

- When stoichiometric and thermodynamic constraints are considered together, the set of possible flux configurations does not generally define a convex set  $\Rightarrow$  the problem has multiple local maxima.

(Dinh *et al.*, 2017)

- Verifying if a flux configuration satisfying stoichiometric constraints is thermodynamically feasible is a problem that can be formulated as a convex optimization problem.

(Müller *et al.*, 2014)

## Some methods

### Network embedded Thermodynamic (NET)

In NET analysis, the feasible ranges of Gibbs energies reaction are limited with an optimization problem:

$$\begin{aligned}
 & \min/\max \Delta_r G_j \\
 & \text{subject to} \\
 & d_j \Delta_r G_j < 0 \\
 & \Delta_r G_j = \sum_{i=1}^m S_{ij} \Delta_f G_i \\
 & \Delta_f G_i = \Delta_f G_i^0 + RT \ln(x_i/x_0) \\
 & x_i^- \leq x_i \leq x_i^+
 \end{aligned} \tag{1}$$

(Kümmel *et al.*, 2006)

### Thermodynamics-based flux analysis (TFA)

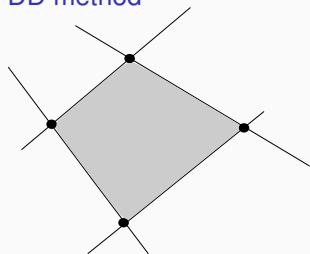
mixed-integer linear programming formulation to compute the flux directionality based on the thermodynamically feasible concentration profiles

(Henry *et al.*, 2007)

## 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  $\text{supp}(v') \subset \text{supp}(v)$ ,  $v'$  verifies  $C$ .

### Adding monotonic constraint in DD method



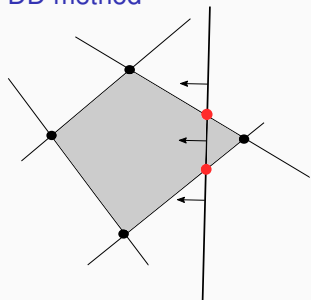
New extreme ray is a positive linear combination of 2 adjacent rays: its support is union of their 2 supports and thus larger.



## 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  $\text{supp}(v') \subset \text{supp}(v)$ ,  $v'$  verifies  $C$ .

### Adding monotonic constraint in DD method

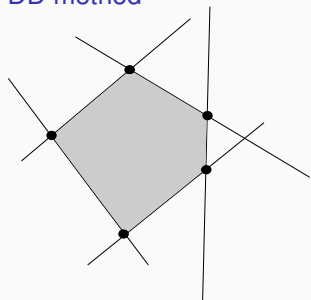


New extreme ray is a positive linear combination of 2 adjacent rays: its support is union of their 2 supports and thus larger.

## 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  $\text{supp}(v') \subset \text{supp}(v)$ ,  $v'$  verifies  $C$ .

### Adding monotonic constraint in DD method

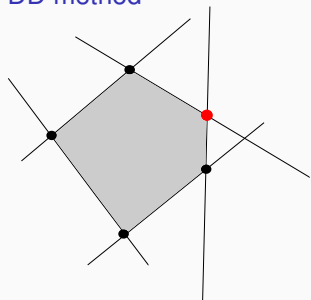


New extreme ray is a positive linear combination of 2 adjacents rays: its support is union of their 2 supports and thus larger.

## 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  $\text{supp}(v') \subset \text{supp}(v)$ ,  $v'$  verifies  $C$ .

### Adding monotonic constraint in DD method

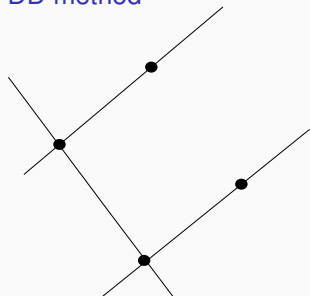


If it violated monotone constraint, it has no use for the rest of the algo and can be discarded.

## 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  $\text{supp}(v') \subset \text{supp}(v)$ ,  $v'$  verifies  $C$ .

### Adding monotonic constraint in DD method



Thermodynamic satisfiability can be checked by an **LP program** at each iteration of the DD method ( $\tau_{\text{EFMA}}$ ).

(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:

$$e^t \log \hat{K} > 0 \tag{2}$$

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

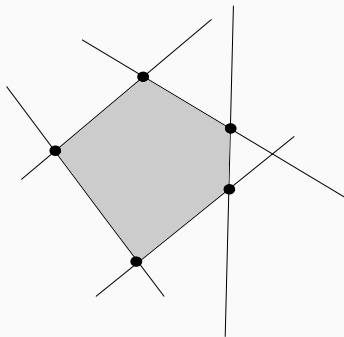
Schuster *al.* J. math. biol, 1991

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 (2) can be directly added as a new linear inequality constraint in the DD algorithm.

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

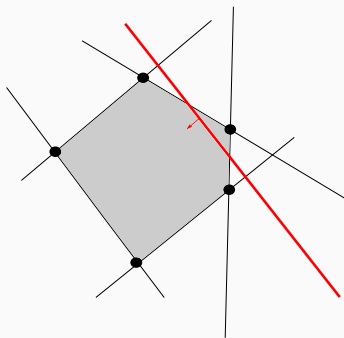


It would be necessary to be able to control the order in which the inequality constraints are processed in order to apply the thermodynamic one in first, before the irreversibility ones, to compare the performances.

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

- Check the scalar product in post-processing
- the formula (2) can be directly added as a new linear inequality constraint in the DD algorithm.

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



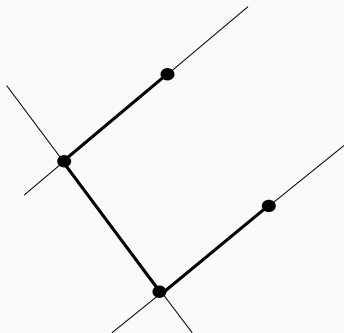
It would be necessary to be able to control the order in which the inequality constraints are processed in order to apply the thermodynamic one in first, before the irreversibility ones, to compare the performances.



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

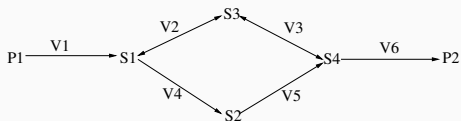
- Check the scalar product in post-processing
- the formula (2) can be directly added as a new linear inequality constraint in the DD algorithm.

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



$$\ln(k'_1) = 5$$

$$\ln(k'_2) = -2$$

$$\ln(k'_3) = -1$$

Add pseudo reaction  $v_{r+1}$

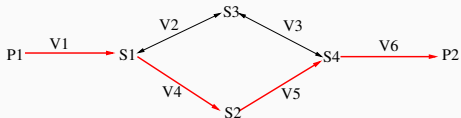
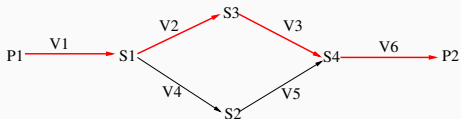
to obtain  $\sum \ln(\hat{K}) > 0$ :

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

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

$$A = \begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \hline 1 & -1 & -1 & +1 & 0 \\ -1 & 1 & 1 & -1 & 0 \\ \hline -1 & 1 & 1 & -1 & 0 \\ 1 & -1 & -1 & +1 & 0 \\ \hline 5 & -2 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{array}$$

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



# How important is thermodynamics for identifying elementary flux modes ?

Simple example:  $n$  monomolecular reactions



- 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

Questions ?