Constraint-based Modeling of Metabolic Networks

Alexander Bockmayr

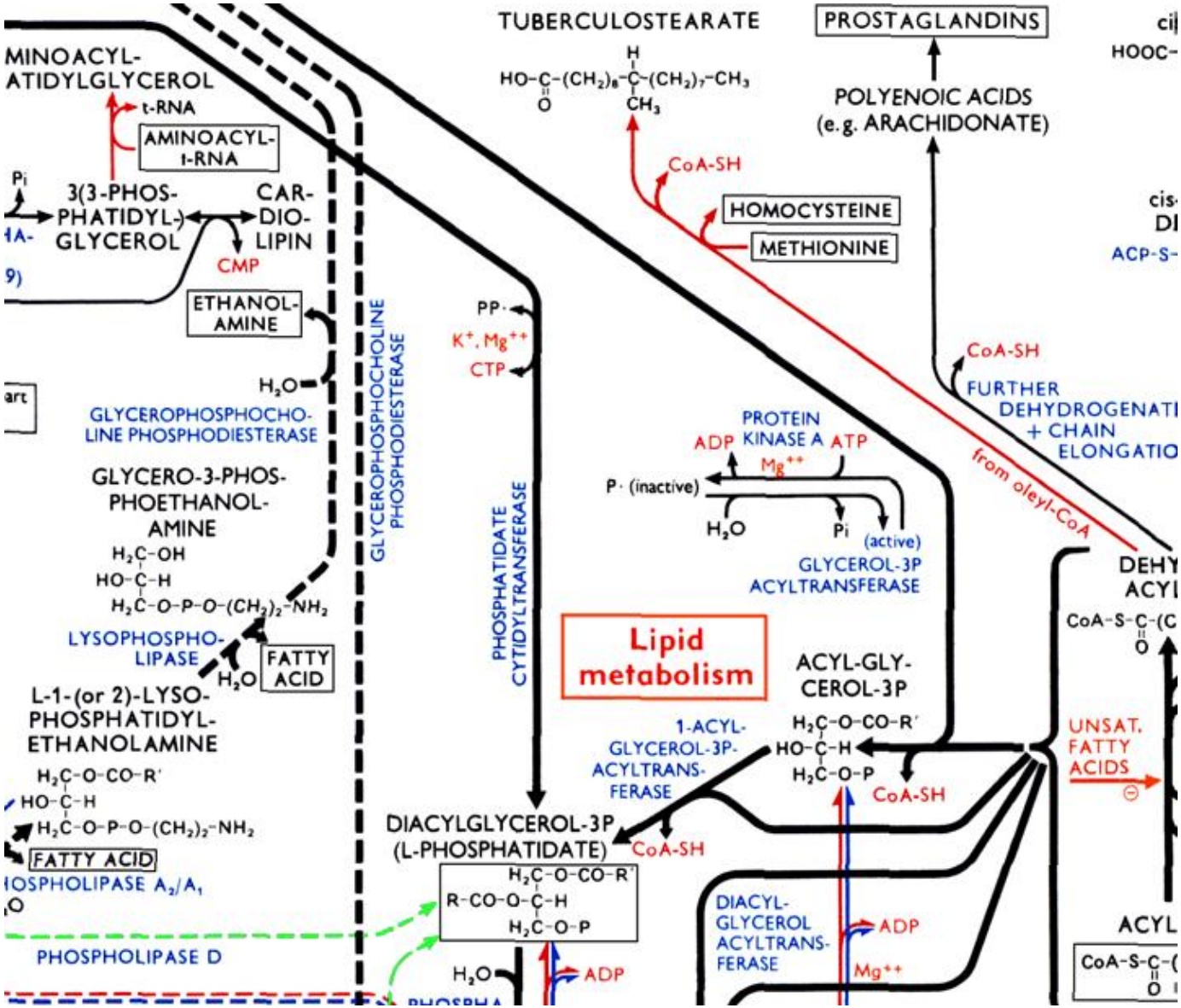
Freie Universität Berlin

Research Center MATHEON

Mathematics for key technologies

Bioinformatics Summer School, Angers, 10 July 2015

Metabolic networks

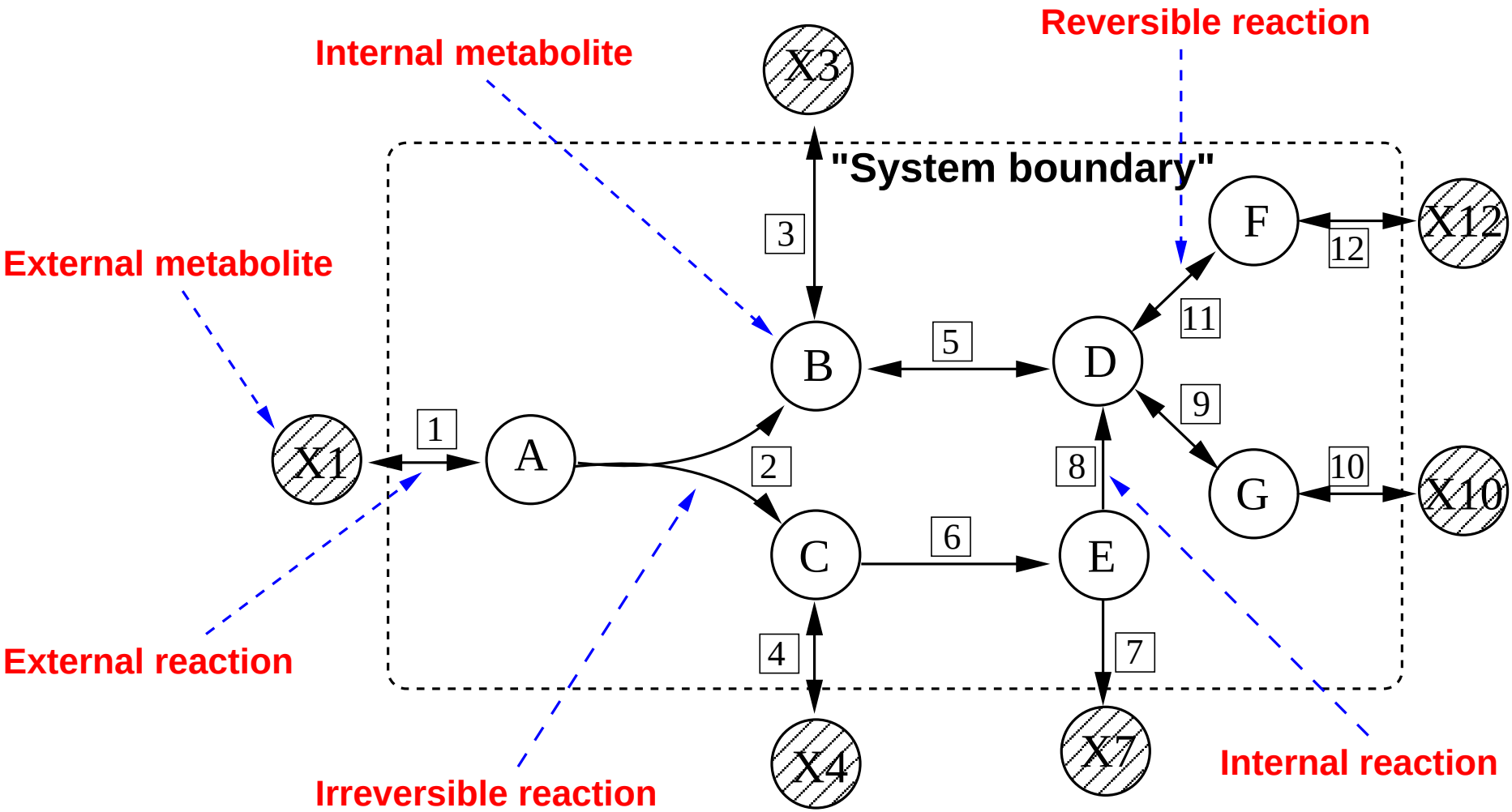


http://web.expasy.org/pathways/

Importance

- Biology
  - Cell metabolism
  - Catabolism, anabolism
- Medicine
  - Metabolic disorders
  - Cancer
- Biotechnology
  - Biofuel, e.g. phototrophic organims
  - Bioleaching

Mathematical representation



Algebraic description

- Stoichiometric matrix $S \in \mathbb{R}^{m \times n}$
  - Rows $\rightsquigarrow$ internal metabolites $i = 1, \ldots, m$
  - Columns $\rightsquigarrow$ internal and external reactions $j = 1, \ldots, n$
  - S_{ij}: stoichiometric coefficient of reactant i in reaction j
- Set of irreversible reactions Irr
- Metabolic model $\mathscr{M} = (S, Irr)$

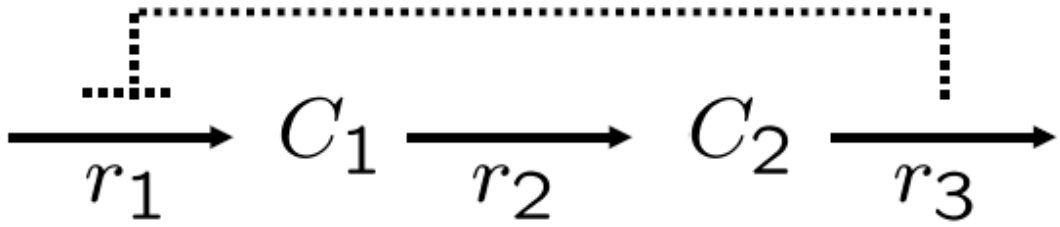
1. Kinetic modeling

- Metabolites i and reactions j
- $C_i(t)$: metabolite concentrations at time t
- $v_j = v_j(C, k)$: reaction rates, depending on kinetic law and kinetic parameters k
- S_{ij}: stoichiometric coefficient

$$\frac{dC_i}{dt} = \sum_{j=1}^{n} S_{ij} v_j \quad \text{or} \quad \frac{dC}{dt} = S \cdot v(C, k)$$

- System of ordinary differential equations (ODEs)

Example



$$\begin{pmatrix} dC_1/dt \\ dC_2/dt \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1(C,k) \\ v_2(C,k) \\ v_3(C,k) \end{pmatrix}$$

$$\begin{aligned} v_1(C,k) &= v_{m1}/(1+(C_2/k_i)^p) \\ v_2(C,k) &= v_{m2} \cdot C_1/(k_1+C_1) \\ v_3(C,k) &= v_{m3} \cdot C_2/(k_2+C_2) \end{aligned}$$

Which kinetic laws?

Which kinetic parameters?

2. Constraint-based modeling

- Steady-state assumption:
  Assume metabolite concentrations C_i and reaction rates v_j are constant $\rightsquigarrow$ flux vector $v \in \mathbb{R}^n$
- Stoichiometric constraints (mass balance):

$$\sum_{j=1}^{n} S_{ij} v_j = 0, \text{ for all } i = 1, \ldots, m$$

- Thermodynamics constraints (reaction directionality):

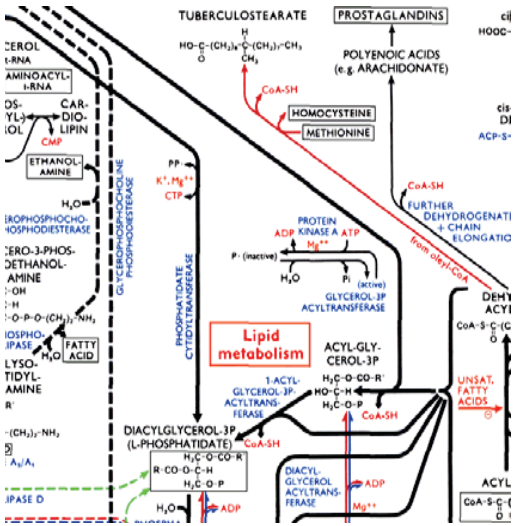
$$v_j \geq 0, \text{ if } j \text{ is irreversible}$$

$\rightsquigarrow$ system of linear equations and inequalities in $\mathbb{R}^n$

Set of all possible steady-state flux distributions

$$C = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_i \geq 0,\ i \in Irr\}$$

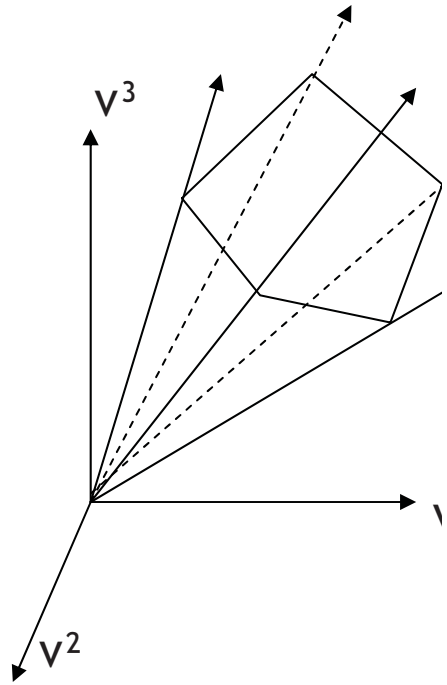
$\rightsquigarrow$ polyhedral cone



Reactions

$$\begin{pmatrix} 1 & 0 & \dots & -1 \\ 0 & 1 & \dots & -2 \\ \dots & \dots & \dots & \dots \\ 2 & 0 & \dots & 1 \end{pmatrix}$$

Metabolites



3. Flux balance analysis (FBA)

- Assume cellular behavior is determined by a certain biological objective.
- Determine a corresponding “best” flux distribution.
- Use mathematical optimization to predict phenotype.

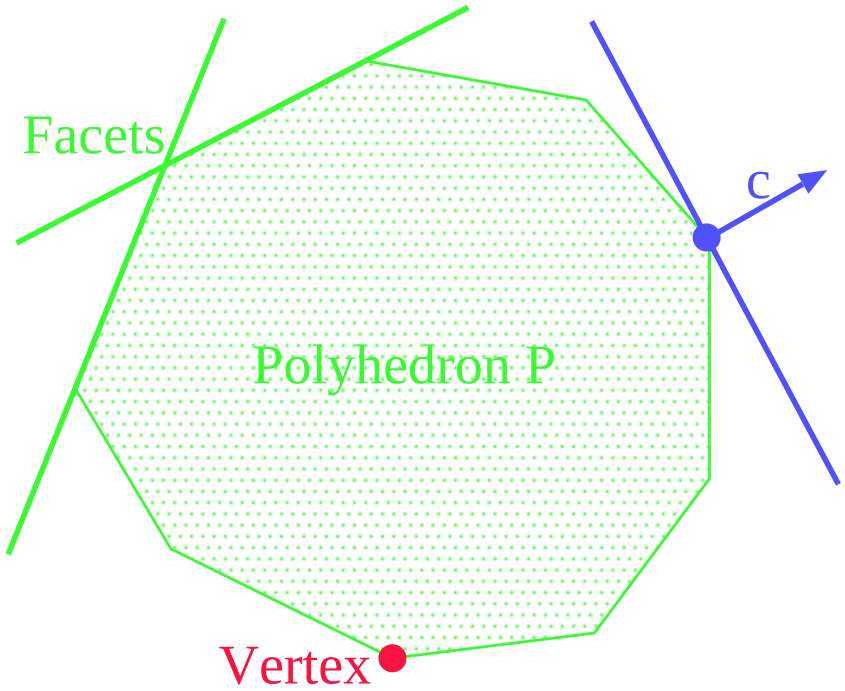
- Simplest case: Linear programming (LP)

$$\max\{c^T x \mid Ax \leq b, x \in \mathbb{R}^n\}$$

- Flux balance problem (FBA)

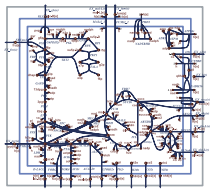
$$\max\{c^T v \mid Sv = 0,\ l \leq v \leq u\} \qquad \text{(FBA)}$$

Reminder: Linear programming



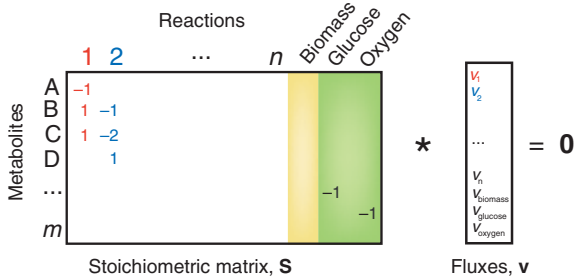
Example

a Genome-scale metabolic reconstruction



$A \leftrightarrow B + C$ Reaction 1
$B + 2C \rightarrow D$ Reaction 2
...
Reaction *n*

b Mathematically represent metabolic reactions and constraints



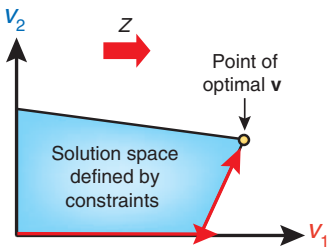
c Mass balance defines a system of linear equations

$$-v_1 + \ldots = 0$$
$$v_1 - v_2 + \ldots = 0$$
$$v_1 - 2v_2 + \ldots = 0$$
$$v_2 + \ldots = 0$$
etc.

d Define objective function ($Z = c_1 * v_1 + c_2 * v_2 \ldots$)

To predict growth, $Z = v_{biomass}$

e Calculate fluxes that maximize *Z*



- *E. coli* metabolism
- Genome-scale reconstruction (*i*JO1366)
- 1336 metabolites, 2251 reactions
- Objective function: biomass
- Glucose and oxygen uptake reactions
- Aerobic and anaerobic growth
- Software: e.g. COBRA Toolbox 2.0

Orth/Thiele/Palsson 10

4. Flux variability analysis (FVA)

- Optimal solutions to FBA problems need not be unique.
- Enumerating all optimal solutions is computationally expensive.
- Alternative: Analyse flux variability

$$z_{opt} = \max\{z = c^T v \mid Sv = 0,\ l \leq v \leq u\} \qquad \text{(FBA)}$$

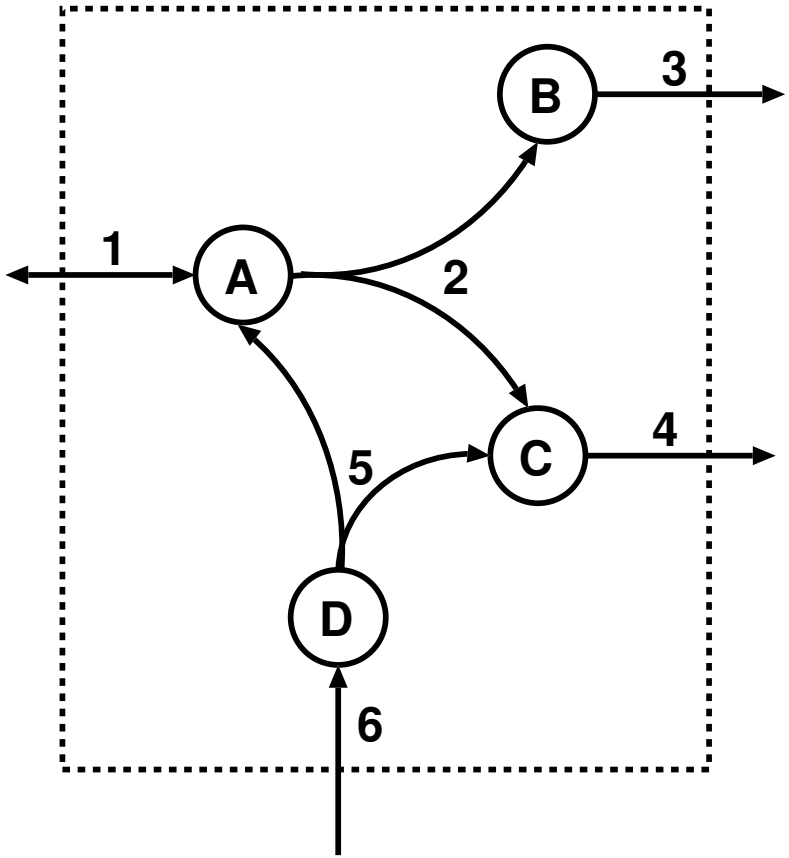
For all $j = 1, \ldots, n$:

$$\max\{\pm v_j \mid Sv = 0,\ l \leq v \leq u,\ c^T v = z_{opt}\} \qquad \text{(FVA)}$$

5. Flux coupling analysis (FCA)

Burgard et al. 04

- $C = \{v \mid Sv = 0,\ v_k \geq 0, k \in Irr\}$ flux cone
- A reaction i is blocked if $v_i = 0$, for all $v \in C$.
- Let i and j be two unblocked reactions.
  - i is directionally coupled to j, $i \overset{=0}{\rightarrow} j$, if for all $v \in C$, $v_i = 0$ implies $v_j = 0$.
  - i and j are partially coupled, $i \overset{=0}{\leftrightarrow} j$, if for all $v \in C$, $v_i = 0$ is equivalent to $v_j = 0$.
  - i and j are fully coupled, $i \backsim^{\lambda} j$, if there exists $\lambda \in \mathbb{R} \setminus \{0\}$ such that for all $v \in C, v_j = \lambda v_i$.
- $i \backsim^{\lambda} j$ implies $i \overset{=0}{\leftrightarrow} j$, which is equivalent to $i \overset{=0}{\rightarrow} j$ and $j \overset{=0}{\rightarrow} i$.



$$2 \xrightarrow{=0} 3$$
$$3 \xrightarrow{=0} 2$$
$$4 \xrightarrow{=0} 1, 2, 3, 5, 6$$
$$5 \xrightarrow{=0} 6$$
$$6 \xrightarrow{=0} 5$$

LP-based flux coupling analysis

- Reaction i is blocked iff

$$\max\{\pm v_i \mid Sv = 0, v_k \geq 0, k \in Irr\} = 0$$

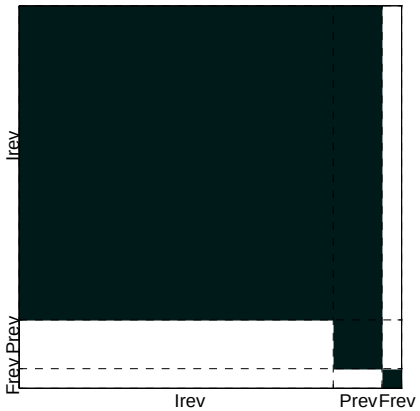
- Two unblocked reactions i and j are directionally coupled, i.e., $i \overset{=0}{\rightarrow} j$ iff

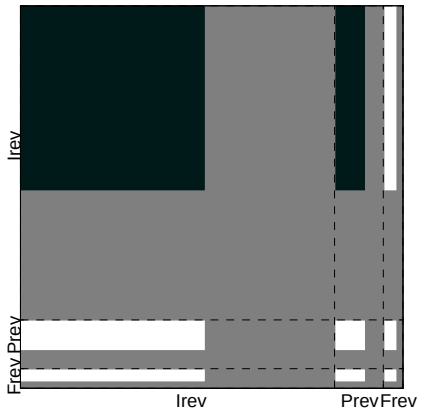
$$\max\{\pm v_j \mid Sv = 0, v_k \geq 0, k \in Irr, v_i = 0\} = 0$$

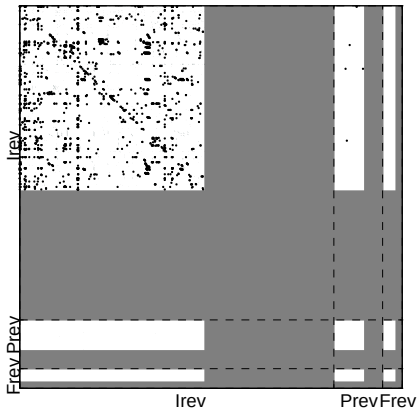
- $O(n^2)$ linear programming problems

Fast Flux Coupling Calculation F2C2

Larhlimi/David/Selbig/Bockmayr 12





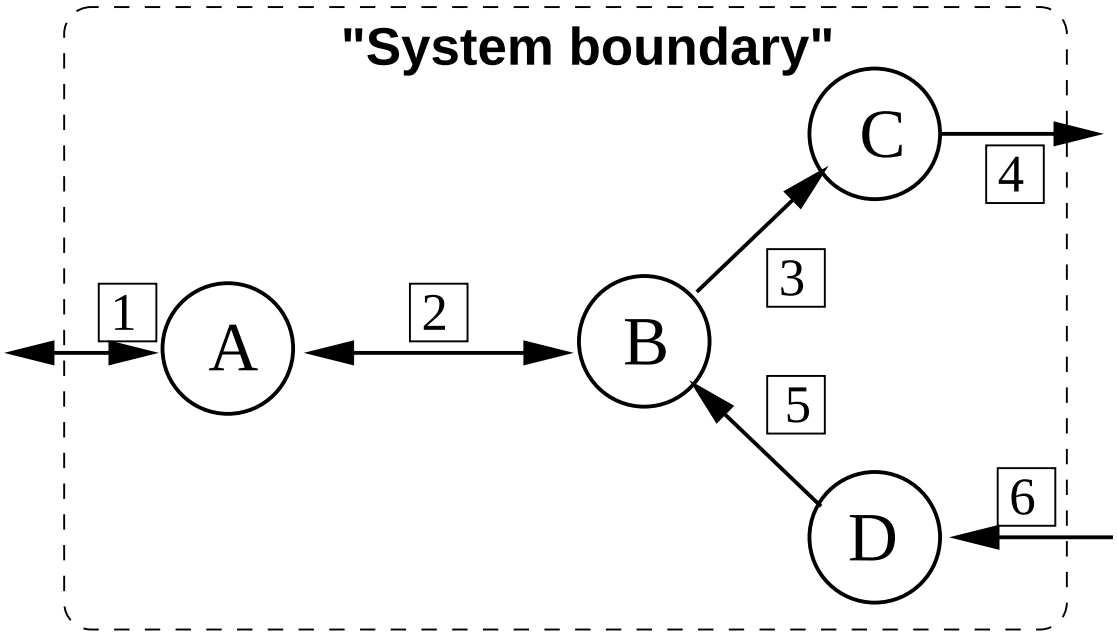


Network	FFCA		F2C2	
	♯LPs	Time	♯LPs	Time
M. barkeri, *iAF*692	301975	59m40s	774	7s
S. cerevisiae, *iND*750	472629	1h50m17s	1280	21s
M. tuberculosis, *iNJ*661	556504	3h5m36s	1506	22s
E. coli, *iJR*904	655437	2h40m33s	1580	26s
E. coli, *iAF*1260	4256786	4d31m26s	3309	2m47s
E. coli, *iJO*1366	4877262	4d5h30m46s	3955	3m55s
H. sapiens, *Recon*1	4566304	4d18h3m37s	3903	5m20s

6. Elementary flux modes

Schuster/Hilgetag'94

- $C = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_i \geq 0, i \in Irr\}$ steady-state flux cone
- Support of $v \in \mathbb{R}^n$: $supp(v) = \{i \in \{1, \ldots, n\} \mid v_i \neq 0\}$.
- Elementary flux mode (EFM):
  Flux vector $v \in C \setminus \{0\}$ with minimal support, i.e., there is no $v' \in C \setminus \{0\}$ with $supp(v') \subsetneqq supp(v)$.



$$e^1 = (1,1,1,1,0,0), \quad e^2 = (-1,-1,0,0,1,1), \quad e^3 = (0,0,1,1,1,1)$$

Computing EFMs

- If all reactions are irreversible, EFMs correspond to extreme rays of the flux cone (Gagneur/Klamt 04).
- EFMs can be computed by algorithms that enumerate the extreme rays of a pointed cone $\rightsquigarrow$ double description method
- Software
  - Metatool (Pfeiffer et al. 99, Univ. Jena)
  - efmtool (Terzer 09, ETH Zurich)
- Enumerating EFMs is computionally hard (Acuña et al. 09 and 10).

MILP to enumerate shortest EFMs

de Figueiredo et al. 09

Assume all reactions are irreversible.

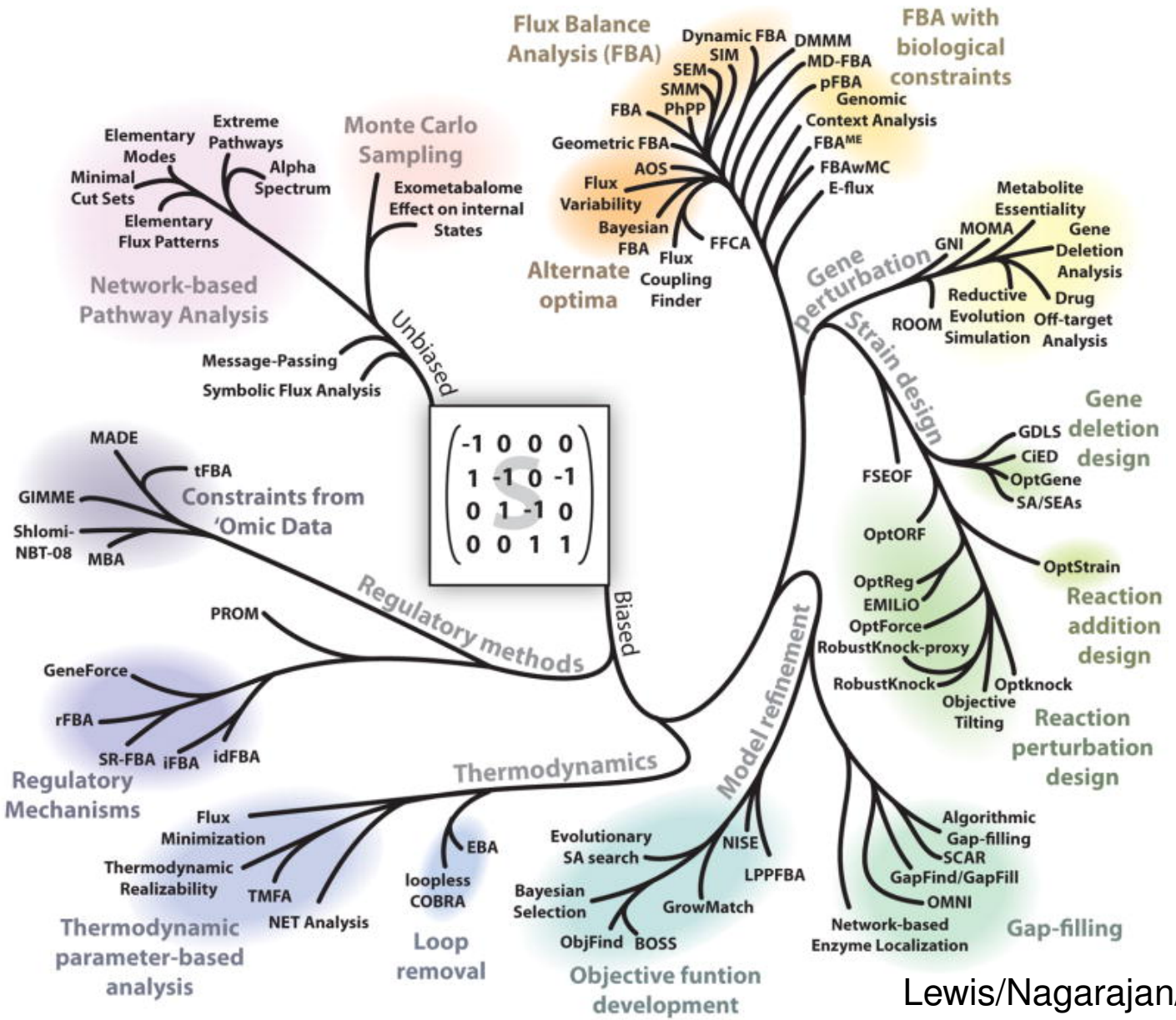
$$\min \sum_{j=1}^{n} a_j$$

$$Sv = 0, v \geq 0,$$
$$a_j \leq v_j \leq M a_j, \quad \text{for } j = 1, \ldots, n, \qquad \text{“BigM”}$$
$$\textstyle\sum_{j=1}^{n} a_j \geq 1,$$
$$v \in \mathbb{R}^n, a \in \{0,1\}^n$$

Forbidding the i-th solution (v^i, a^i):

$$\sum_{j \in supp(v^i)} a_j \leq |supp(v^i)| - 1, \quad \text{for } i = 1, 2, \ldots, k \qquad \text{“no-good cut”}$$

Constraint-based analysis methods



Lewis/Nagarajan/Palsson 12

Summary

- Metabolic networks at steady state
- Steady-state flux cone
- FBA, FVA, FCA
- Elementary flux modes
- Linear programming (LP), integer linear programming (ILP)

References

- Terzer M, Maynard ND, Covert MW, Stelling J. *Genome-scale metabolic networks.* Wiley Interdiscip Rev Syst Biol Med. 2009 Nov-Dec;1(3):285-97.
- Orth JD, Thiele I, Palsson B. *What is flux balance analysis?* Nat Biotechnol. 2010 Mar;28(3):245-8.
- Lewis NE, Nagarajan H, Palsson BO. *Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.* Nat Rev Microbiol. 2012 10(4):291-305.