Metabolic Network Analysis

Christoph Flamm

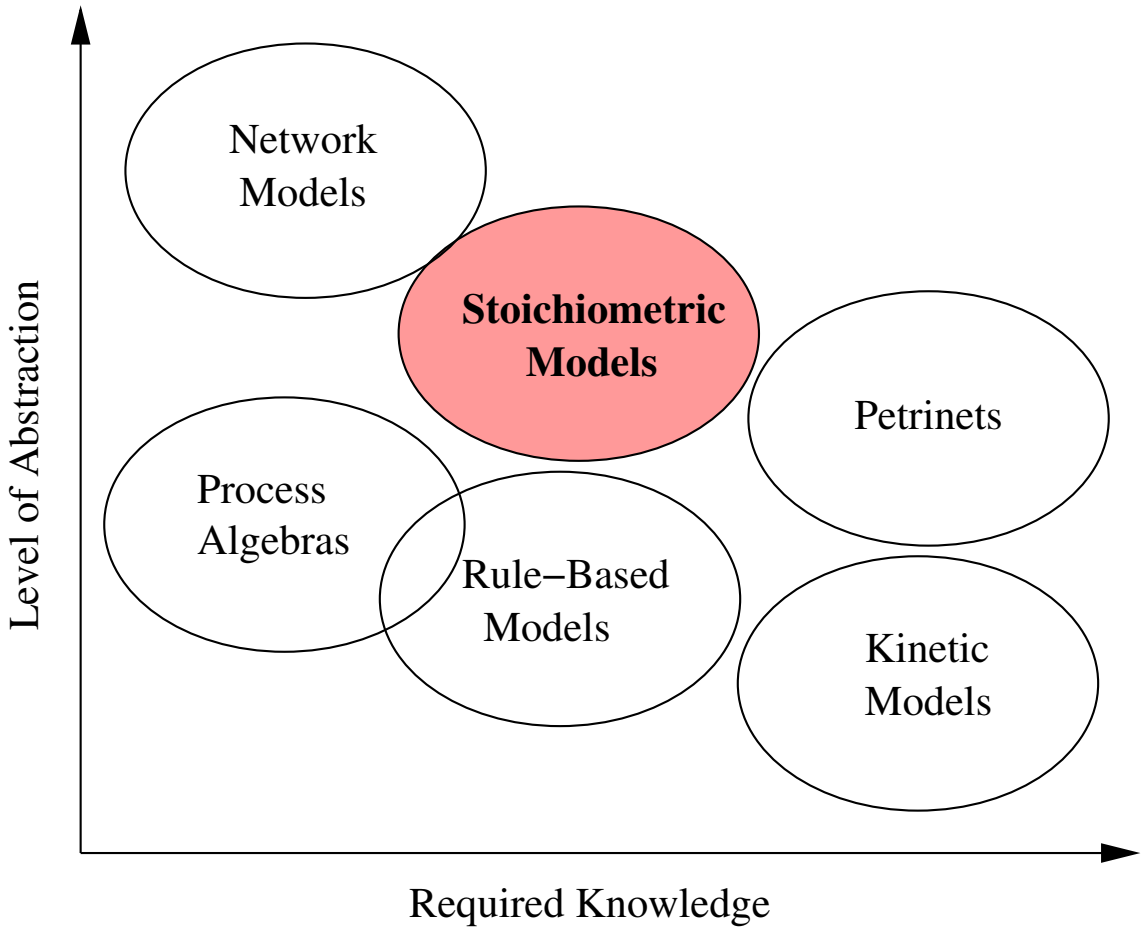
xtof@tbi.univie.ac.at

Institute for Theoretical Chemistry
University of Vienna

VS 791328 Modern Bioinformatics
Wien, December 14, 2016

universität wien

The Systems Biology Modelling Space



Genome-scale Reconstruction of Metabolic Networks

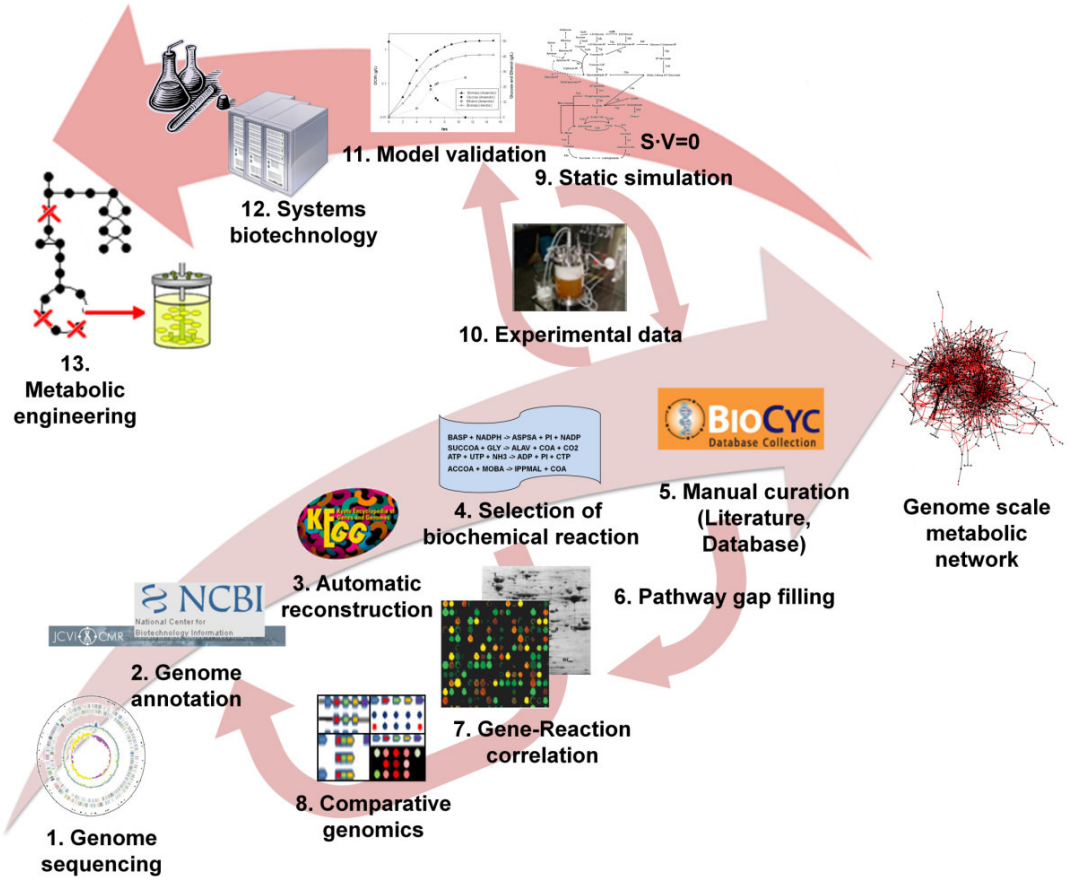
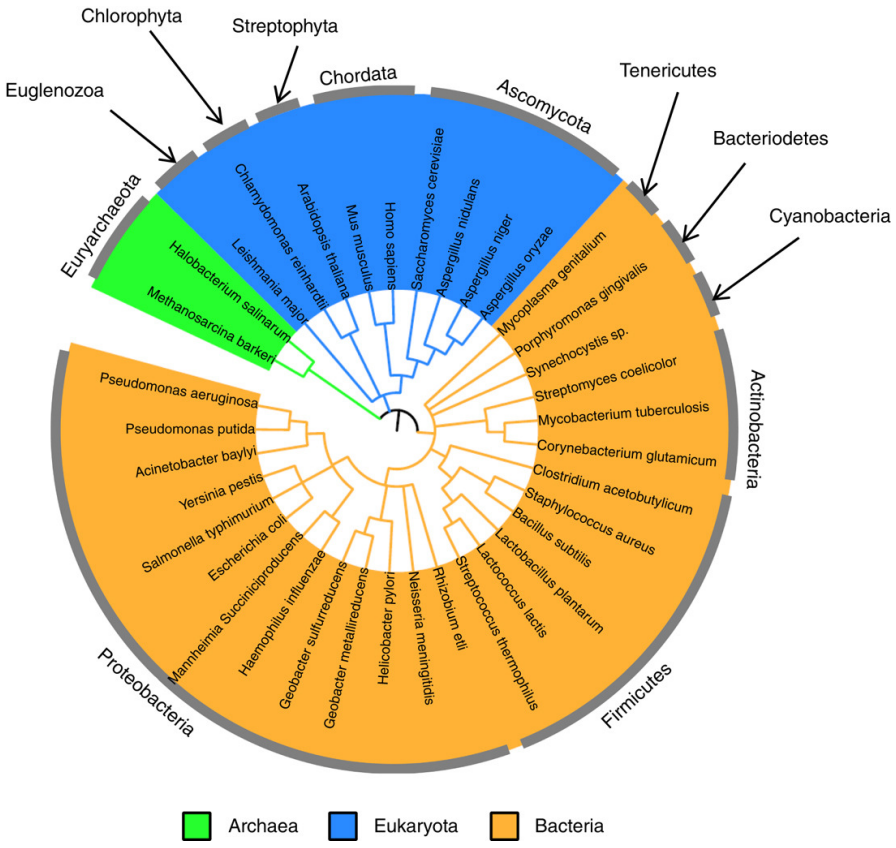


Figure from Lee KY et al (2010), Microbial Cell Factories 9:94 | doi:10.1186/1475-2859-9-94

Phylogenetic tree of metabolic reconstractions



Our metabolic knowledge is srongly biased towards cultivatable bacteria.

Figure from Oberhardt MA et al (2009) Mol Syst Biol 5:320 | doi:10.1038/msb.2009.77

Chemical Reactions and Stoichiometry

Educt(s) are found left and **product(s)** right of the reaction arrow.

$$
\begin{aligned}
2\ \mathrm{Ag} + 1\ \mathrm{S} &\longrightarrow 1\ \mathrm{Ag_2S} \\
4\ \mathrm{Fe} + 3\ \mathrm{O_2} &\longrightarrow 2\ \mathrm{Fe_2O_3} \\
1\ \mathrm{C_6H_6O_2} &\longrightarrow 2\ \mathrm{C_6H_5OH} + 2\ \mathrm{CO_2} \\
\sum_{k=1}^{i} |\nu_k| \cdot X_k &\longrightarrow \sum_{k=i+1}^{m} |\nu_k| \cdot X_k
\end{aligned}
$$

The ν_k's are *integer values* called stoichiometric coefficients.

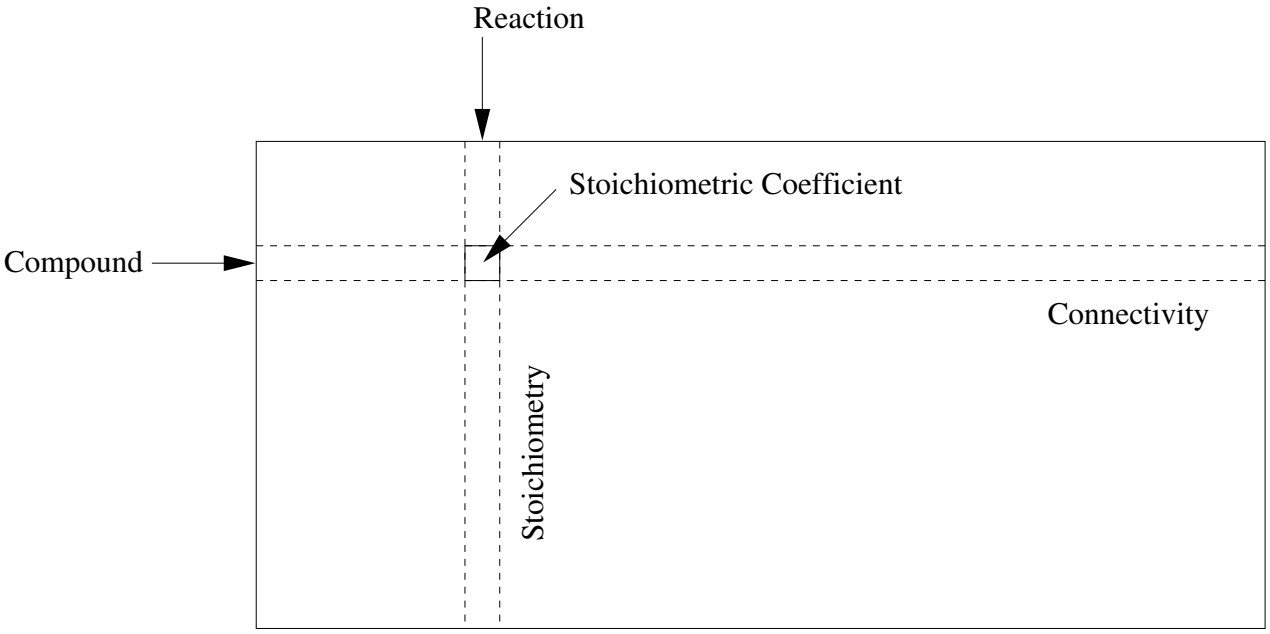
By convention:

$$
\nu_k = \begin{cases} > 0 & \text{if compound is formed by reaction} \\ < 0 & \text{if compound is consumed by reaction} \\ = 0 & \text{otherwise} \end{cases}
$$

Note that the stochiometric coeffitionts are **constants** and do not depend on reaction conditions such as temperature, pressure, pH, . . .

Stoichiometric Matrix **S**

It is convenient to handle sets of chemical reactions, composing a reaction network, in matrix form.

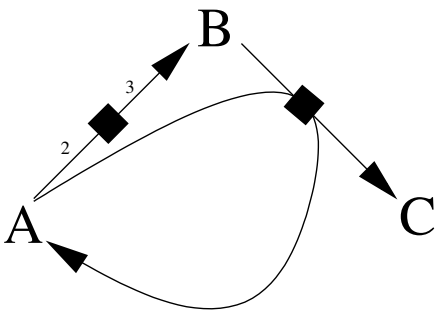


The **rows** correspond to *chemical compounds*,
and **columns** to reactions.

Examples: Stoichiometric Matrix

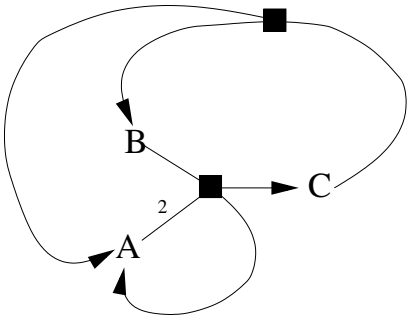
$$2\,\mathrm{A} \longrightarrow 3\,\mathrm{B}$$
$$1\,\mathrm{A} + 1\,\mathrm{B} \longrightarrow 1\,\mathrm{A} + 1\,\mathrm{C}$$

$$\mathbf{S} = \begin{pmatrix} -2 & 0 \\ 3 & -1 \\ 0 & 1 \end{pmatrix}$$



$$2\,\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{A} + \mathrm{C}$$
$$\mathrm{C} \longrightarrow \mathrm{A} + \mathrm{B}$$

$$\mathbf{S} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$



The three representations are equivalent apart from catalysts.

Genome-scale Stoichiometric Matrix

are sparse objects.

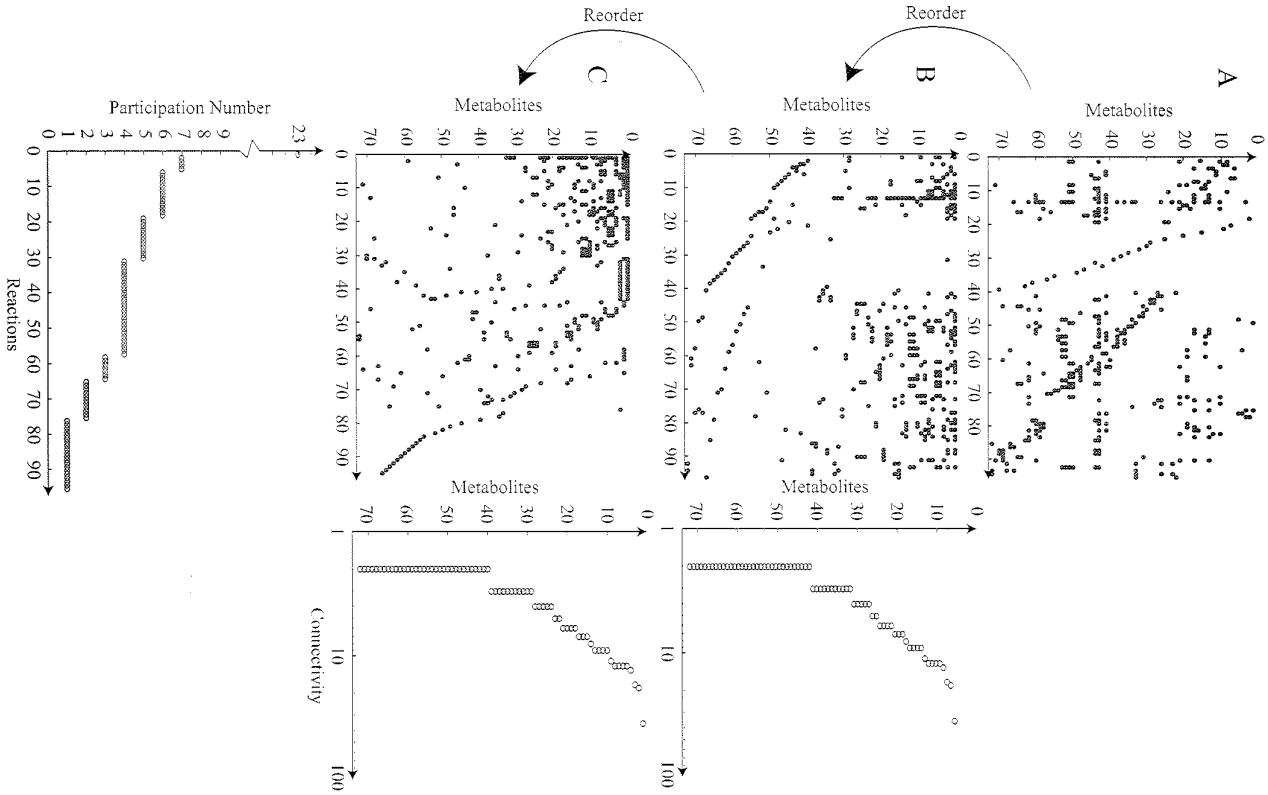
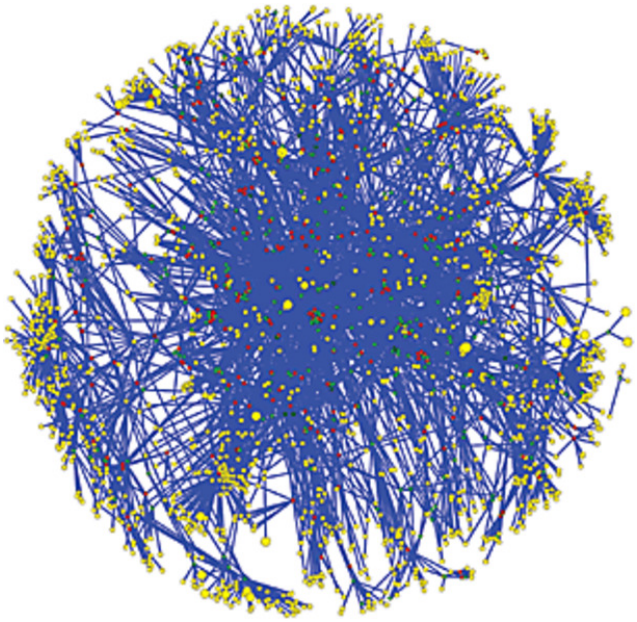


Figure modified from Palsson, BØ (2015), Systems Biology: Constraint-based Reconstruction and Analysis, ISBN 978-1-107-03885-1

What is a Biological Pathway?



$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix}$$

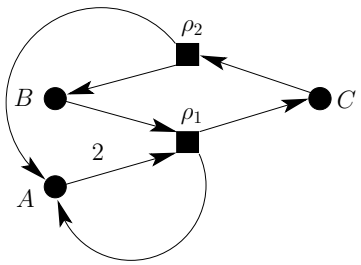
1. Mathematical description of what a pathway is.
2. Formalism to describe mass flows in the network.

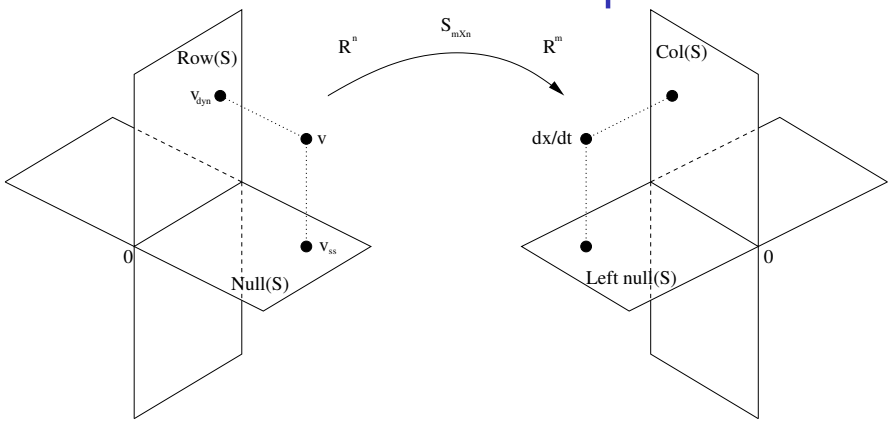
Wanted: All **minimal sets of reaction**, which
can operate at steady state and *respect* reaction directions.

Such a set is call an elementary pathway.

The stoichiometric Matrix is a linear Operator

$$2A + B \rightarrow C + A$$
$$C \rightarrow A + B$$



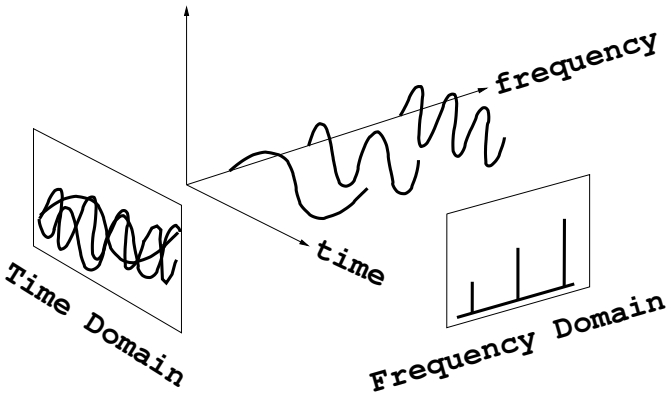


Reaction velocity Domain Concentration Domain

$$\frac{\mathrm{d}}{\mathrm{dt}}[\vec{X}] = S \cdot \vec{J} = \frac{\mathrm{d}}{\mathrm{dt}} \begin{pmatrix} [A] \\ [B] \\ [C] \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} k_{\rho_1}[A]^2[B] \\ k_{\rho_2}[C] \end{pmatrix}$$

$$[\dot{A}] = k_{\rho_2} \cdot [C] - k_{\rho_1} \cdot [A]^2 \cdot [B]$$
$$[\dot{B}] = k_{\rho_2} \cdot [C] - k_{\rho_1} \cdot [A]^2 \cdot [B]$$
$$[\dot{C}] = k_{\rho_1} \cdot [A]^2 \cdot [B] - k_{\rho_2} \cdot [C]$$

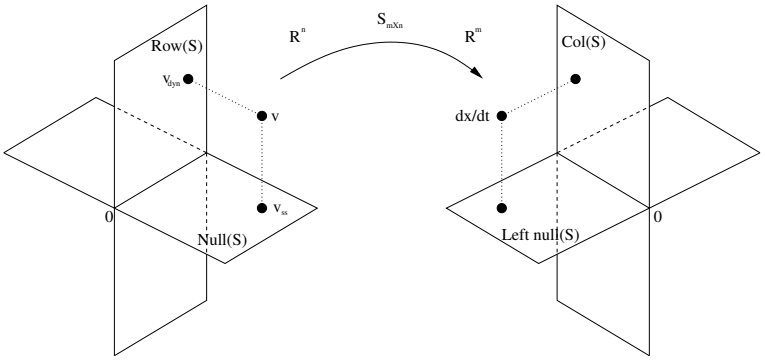
Linear transformations between (vector) spaces



$$\frac{d}{dt}\vec{x} = \mathbf{S} \cdot \vec{v}$$

S as linear transformation from $\vec{v}$ to $d\vec{x}/dt$.

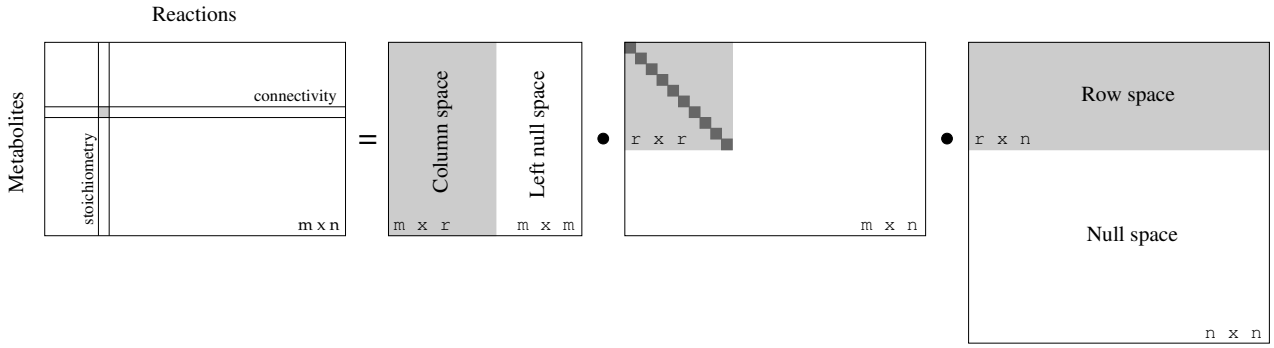
- **Flux domain**
  Row space
  Null space.
- **Conc. domain**
  Column space
  Left null space.



A probably familiar linear operator is the Fourier transform switching between time and frequency domain of a signal.

Singular Value Decomposition (SVD)

$$\mathbf{S} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$$



$$\mathbf{S} = \sum_{i=1}^{r} \sigma_i \cdot (\mathbf{u}_i \cdot \mathbf{v}_i^T)$$

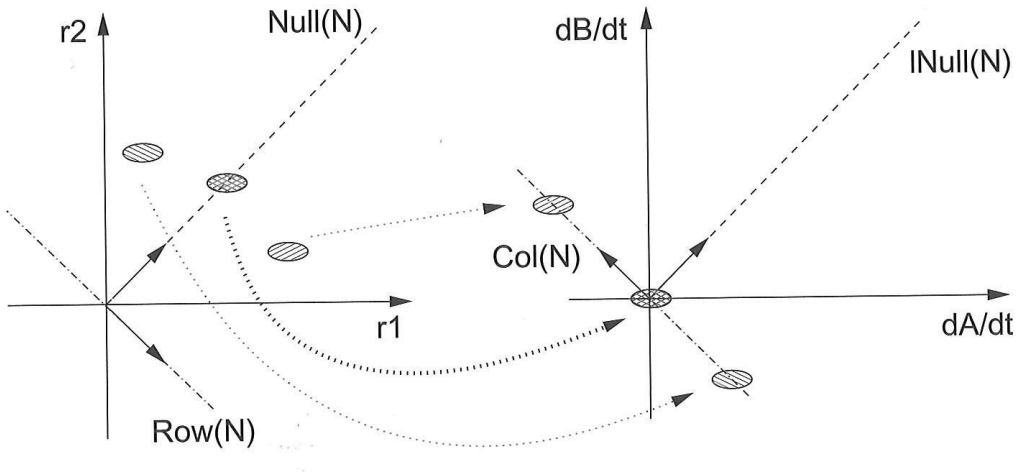
1. Row space contains all dynamic flux distributions.
2. Null space contains all steady-state flux distributions.
3. Column space contains all time derivatives of the conc. vector.
4. Left null space contains all conservation relationships.

r is the rank of S, i.e. the number of linear-independent row or column vectors

Example: SVD of a Reversible Reaction

$$\mathrm{A} \underset{r_2}{\overset{r_1}{\rightleftharpoons}} \mathrm{B}$$

$$N = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \underbrace{\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}}_{U} \cdot \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}}}_{V^T}$$



Relation between the 4 fundamental sub-spaces (Row(N), Null(N), Col(N), lNull(N)) of the stoichiometric matrix N with rank $r = 1$. Figure adapted from Palsson BØ Systems Biology – Properties of reconstructed networks.

The Kernel Matrix **K** is a Null Space Basis

In steady-state the differential change in species concentrations vanishes

$$\frac{\mathrm{d}}{\mathrm{dt}}[\vec{X}] = \vec{0} = \mathbf{S} \cdot \vec{J}$$

Non-trivial solutions for the flux vector $\vec{J}$ exists only if there are linear dependencies between columns of **S**

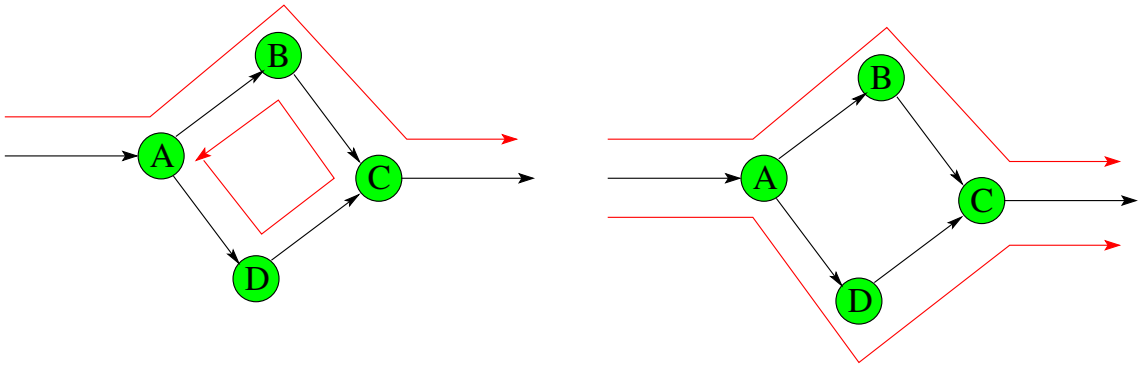
$$\mathrm{Rang}(\mathbf{S}) < \text{number of reactions}$$

These vectors which **span the null space** of **S** are most conveniently organized in a kernel matrix **K**.

- Each column vector $\vec{k}_i$ of **K** solves $\mathbf{S} \cdot \vec{k} = 0$
- Any *admissible flux* in steady-state can be written as a **linear combination** of vectors $\vec{k}_i$

$$\vec{J} = \sum_i \alpha_i \cdot \vec{k}_i$$

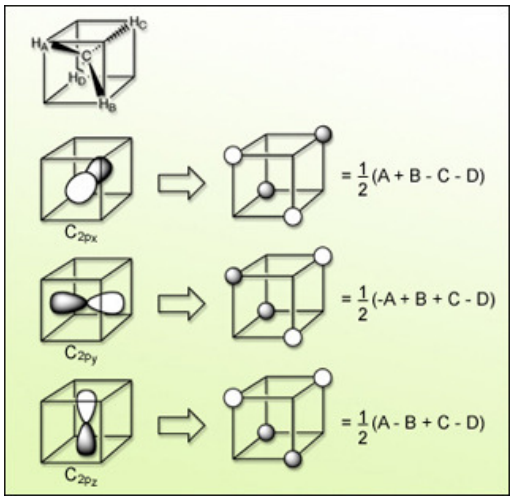
Orthonormal Null Space Basis cannot be Interpreted!

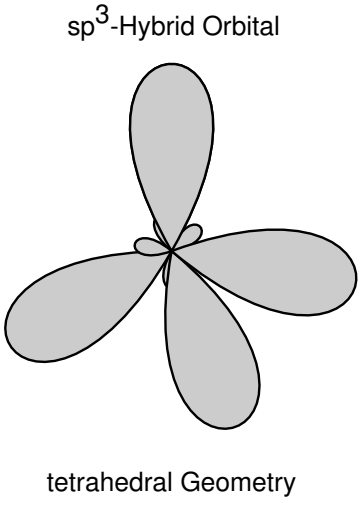


$$\mathbf{S} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad \mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \mathbf{K}' = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

- **K** is not interpretable in chemically meaningful terms.
- Flux through elementary reactions must be positive.
- Non-negative basis vectors are required.
- A convex basis has proven useful for this goal.

Hybrid orbitals: are a basis change!





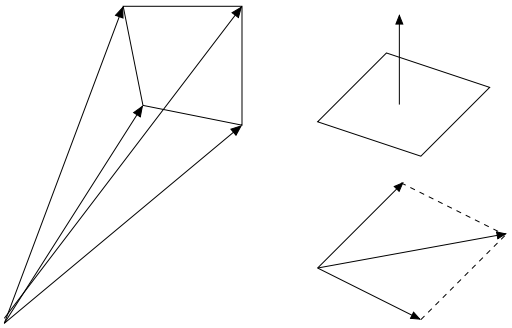
Linear combinations (hybrid orbitals) of H 1s atomic orbitals that match nodal properties of C 2p atomic orbitals to understand the tetrahedral geometry of methane.

Figure by MIT OpenCourseWare

Double Description Method for Basis Change

A pair $(\mathbf{A}, \mathbf{R})$ of real matrices $\mathbf{A}$ and $\mathbf{R}$ is said to be a double description pair if the following relationship holds

$$\mathbf{Ax} \geq \mathbf{0} \quad \text{if and only if} \quad \mathbf{x} = \mathbf{R}\lambda \quad \text{for some } \lambda \geq \mathbf{0}$$



$\mathbf{A}$ is the *representation matrix*, (implicit description by constraints) , $\mathbf{R}$ is the *generating matrix* (explicit description by the edges) of a polyhedral cone P.

Geometrically, the columns of a minimal generating matrix $\mathbf{R}$ are in a 1-to-1 correspondence with the extreme rays of P.

An iterative procedure inizializing $\mathbf{R}$ with $\mathbf{K}$ is used to transform the *ortho-normal* to a *convex basis* respecting reaction directions.

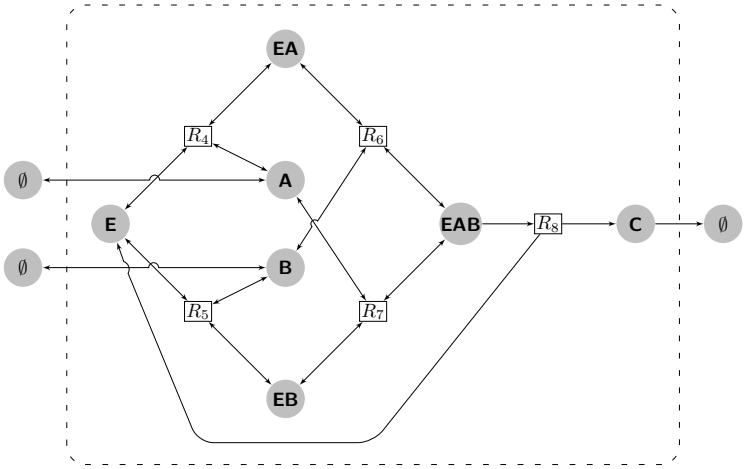
Gagneur J and Klamt S, (2004) Computation of elementary modes: a unifying framework and the new binary approach, *BMC Bioinformatics* **5**:175 | doi:10.1186/1471-2105-5-175

Properties of linear and convex bases

Linear Space	**Convex Space**
• $\mathbf{A} \cdot \vec{x} = \vec{0}$.	• $\mathbf{A} \cdot \vec{x} = \vec{0}$ with $\vec{x} \geq \vec{0}$.
• Set of linearly independent basis vectors $(\vec{e}_i)$.	• Set of conically independent generating vectors $(\vec{p}_i)$.
$\vec{v} = \sum w_i \vec{e}_i$ with $w_i \in [-\infty, +\infty]$	$\vec{v} = \sum \alpha_i \vec{p}_i$ with $\alpha_i \in [0, +\text{inf}]$
• Unique representation of every point.	• Nonunique representation of every point.
• $\lvert\vec{e}_i\rvert = \dim(\text{null}(\mathbf{S}))$.	• $\lvert\vec{p}_i\rvert \geq \dim(\text{null}(\mathbf{S}))$.
• Infinit number of spanning bases.	• Unique set of generating vectors.

Enzyme Mechanism: Unordered Substrate Binding

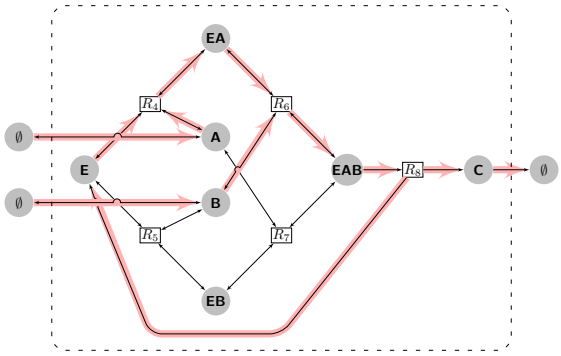
$$
\begin{aligned}
R_1 &: \quad \varnothing \rightleftharpoons A \\
R_2 &: \quad \varnothing \rightleftharpoons B \\
R_3 &: \quad C \rightarrow \varnothing \\
R_4 &: \quad E + A \rightleftharpoons EA \\
R_5 &: \quad E + B \rightleftharpoons EB \\
R_6 &: \quad EA + B \rightleftharpoons EAB \\
R_7 &: \quad EB + A \rightleftharpoons EAB \\
R_8 &: \quad EAB \rightarrow E + C
\end{aligned}
$$

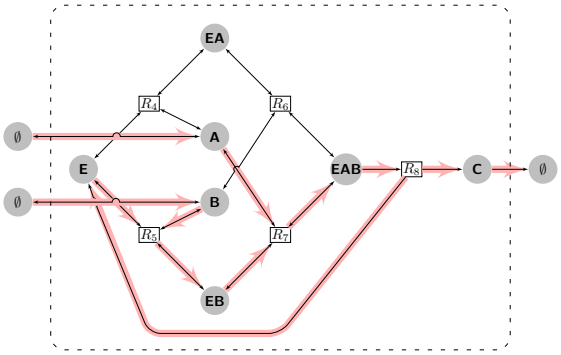


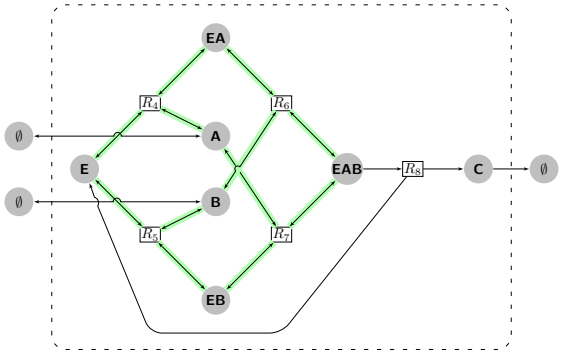
$$
S = \begin{bmatrix}
1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & -1
\end{bmatrix}
$$

Engl HW et al (2009), *Inverse Problems* **25**:123014 | doi:10.1088/0266-5611/25/12/123014

Enzyme Mechanism: Elementary Modes



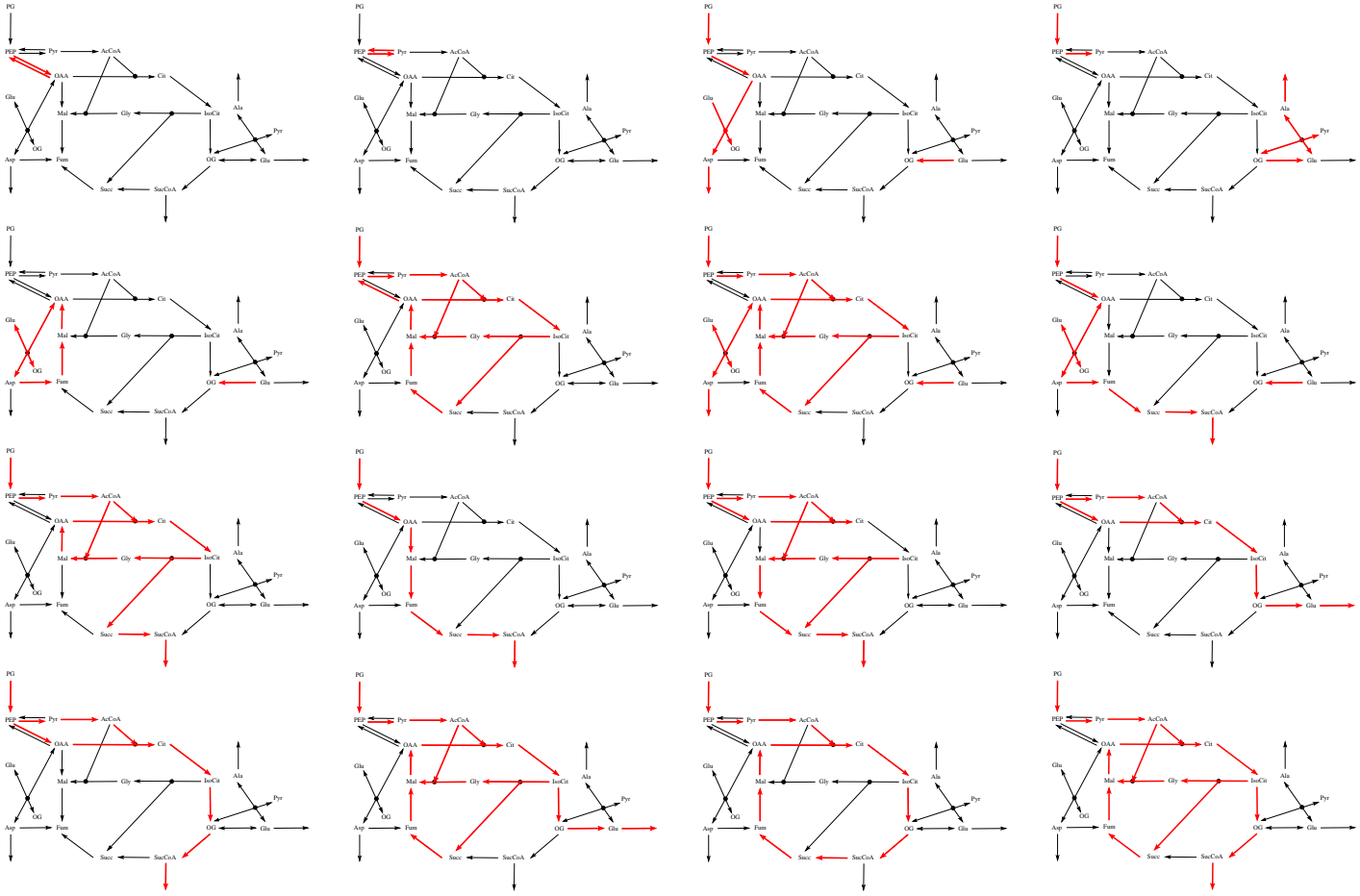




$$
\mathbf{K}' = \begin{bmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & -1 \\
0 & 1 & -1 & 1 \\
1 & 0 & 1 & -1 \\
0 & 1 & -1 & 1 \\
1 & 1 & 0 & 0
\end{bmatrix}
$$

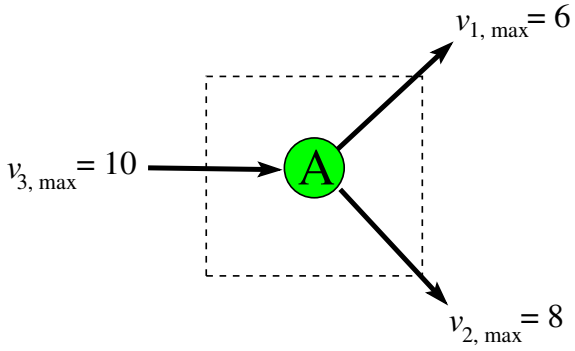
Engl HW et al (2009), *Inverse Problems* **25**:123014 | doi:10.1088/0266-5611/25/12/123014

Elementary modes of TCA in *E. coli*

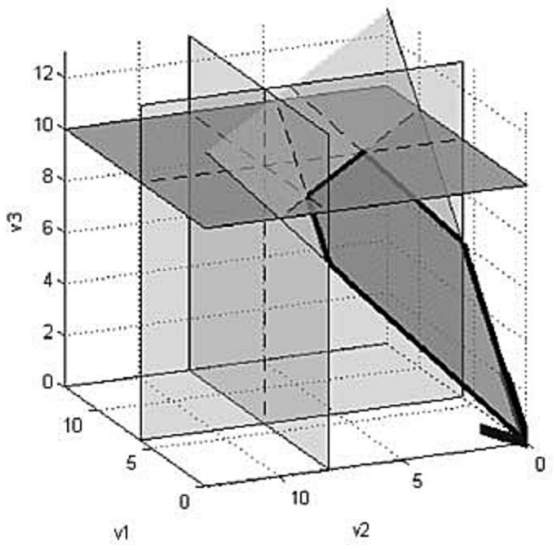


Schuster S et al (1999), Trends Biotechnol 17:53-60 | doi:10.1016/S0167-7799(98)01290-6

Narrowing the Flux Cone: The simple flux split



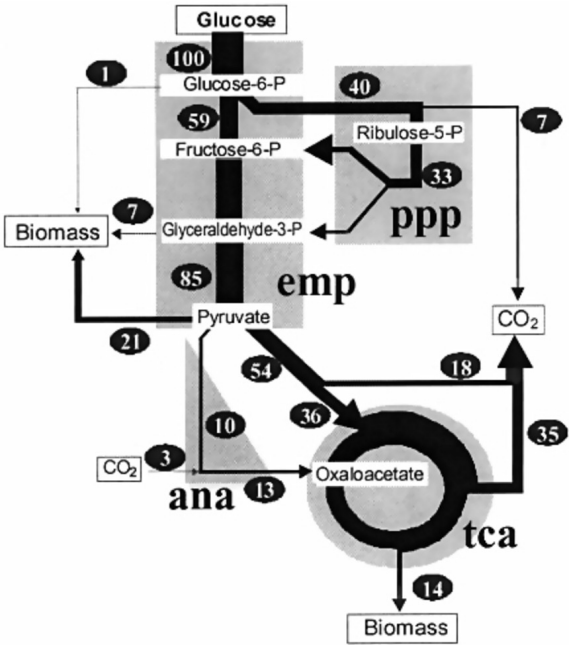
$$\mathbf{S} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \quad \mathbf{K} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$



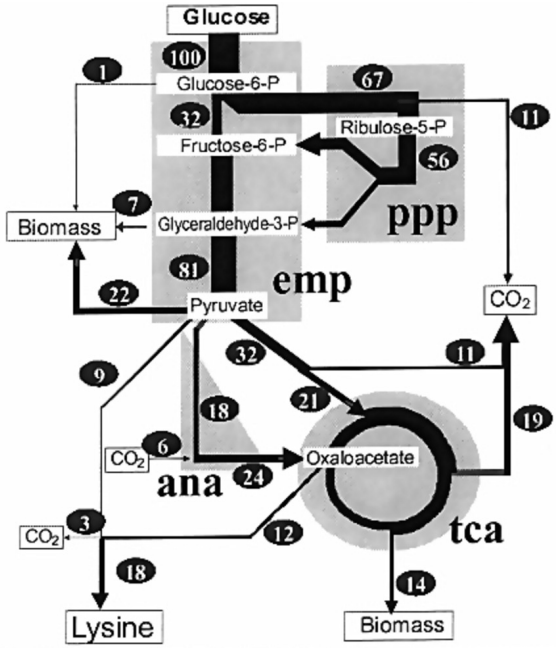
Physico-chemical constrains are required to confine the solution space to functional states the network can achieve.

Price ND et al, (2004) Uniform Sampling of Steady-State Flux Spaces: Means to Design Experiments and to Interpret Enzymopathies, *Biophys J* **87**(4):2172-2186 | 10.1529/biophysj.104.043000

Metabolic Flux Analysis (MFA)



normal growth condition.

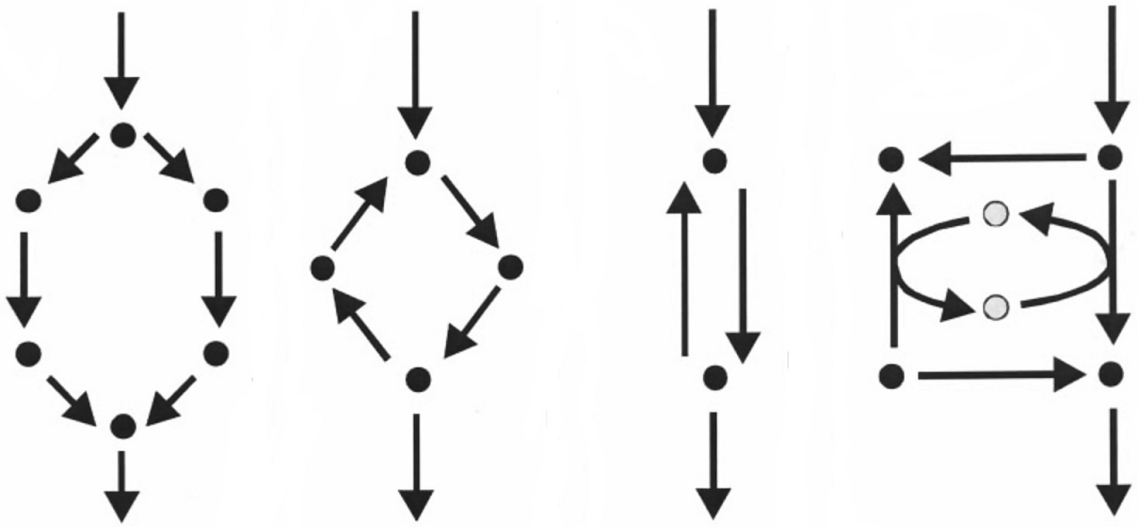


Lys-production condition.

The goal of MFA is the quantitative description of cellular fluxes.

Figure adapted from Marx A, Bestimmung des Kohlenstoffflusses im Zentralstoffwechsel von *Corynebacterium glutamicum* mittels ^{13}C-Isotopenanalyse, PhD-thesis Uni Düsseldorf (1997).

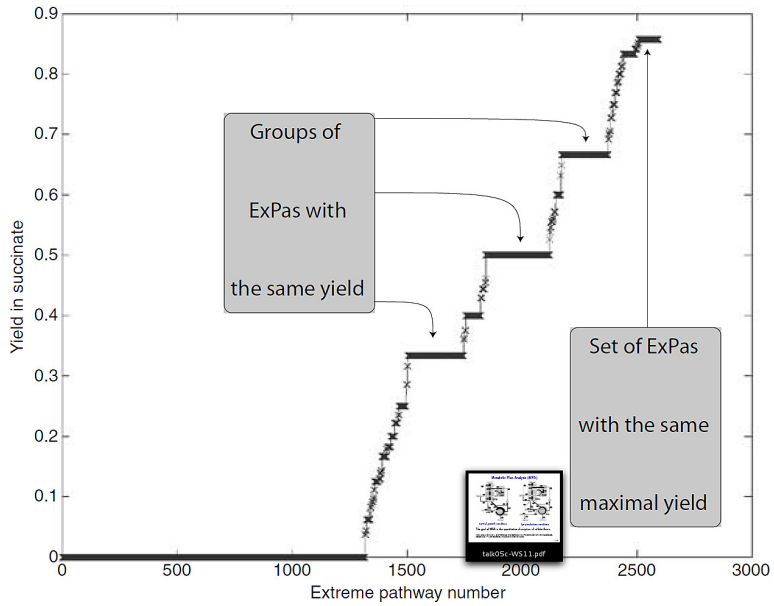
Fail cases for Metabolic Flux Analysis (MFA)



- parallel pathways without any related flux measurement.
- certain metabolic cycles.
- bi-directional reaction steps.
- split pathways when cofactors are not balanced.

Figure adapted from [Wiechert, 2001]

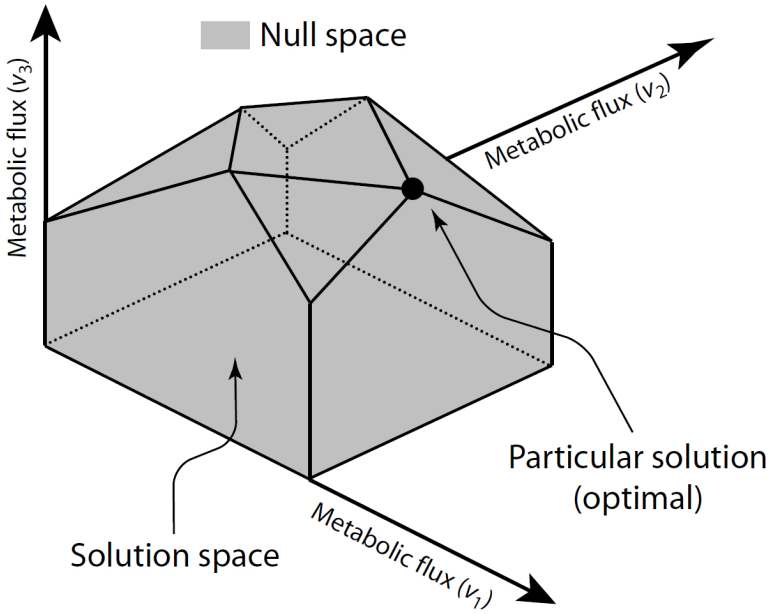
Many Flux Distributions for the same "optimal" yield



Core metabolic network of *E. coli* (56 compounds, 64 reactions, 2598 extreme pathways).

The optimal value for `SUC` production (0.86 mol/mol) from `FUM` can be achieved by 88 different flux distributions!
(Note: any non-negative linear combination is optimal as well)

Finding the "Best" Flux Distribution



An objective function is required to pick out particular solutions with desired properties from the bounded null space.

Figure from Palsson, BØ Systems Biology, Properties of Reconstructed Networks, Cambridge University Press, ISBN-13 978-0-521-85903-5

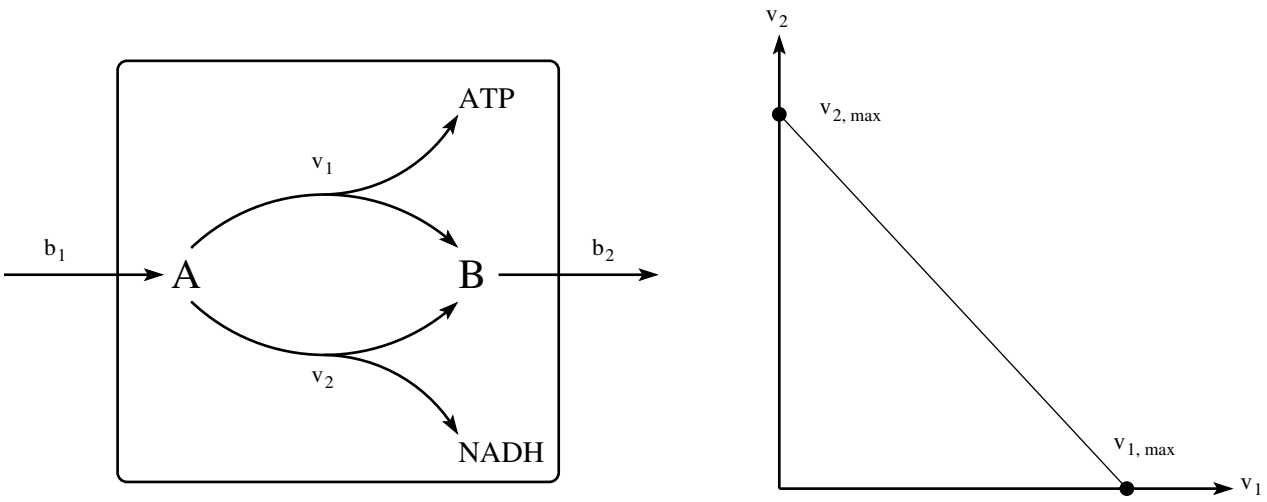
Flux Balance Analysis

Formulate as an linear programing problem with additional constraints (capacity of enzymes, external fluxes, ...) and an appropriate optimization function.

Obj Function	Explanation
$\max \frac{v_{\text{Biomass}}}{v_{\text{Glucose}}}$	biomass yield (same as groth rate)
$\max \frac{v_{\text{ATP}}}{v_{\text{Glucose}}}$	ATP yield
$\min \frac{\sum v_{\text{NADH}}}{v_{\text{Glucose}}}$	redox potential
$\min \sum \delta_i$	reaction steps
$\max \frac{v_{\text{Biomass}}}{\sum v_i^2}$	biomass yield per flux unit

Schuetz R et al (2007), **Systematic evaluation of objective functions for predicting intracellular fluxes in E. coli**, Mol Sys Biol 3:119 | doi:10.1038/msb4100162

A Simple Example



The flux balance for this system is $v_1 + v_2 = b_1 \quad (= b_2)$.

$v_1, v_2 \geq 0$ constrains the solution space to a line segment.

Maximizing `ATP` or `NADH` production lie at the end of the line.

Example modified from Bonarius HPJ et al (1997), Flux analysis of underdetermined metabolic networks: the quest for the missing constraints, *Trends Biotech* **15**(8):308-314 | doi:10.1016/S0167-7799(97)01067-6

Primer: Linear Programming (LP)

Linear programming is an optimization method requiring 2 inputs:

1. A linear objective function.
2. A set of linear constraints.

Example: Production planning problem

Product	machine 1	machine 2	machine 3
A	40	24	0
B	24	48	60

Total machine running time is 8 hours/day.
Profit: 10 €/A and 40 €/B.

Question: How many units of product A and B need to be manifactured in order to maximize profit?

Expressed as LP Problem

1. maximize profit:

$$z = F(x_1, x_2) = 10 \cdot x_1 + 40 \cdot x_2$$

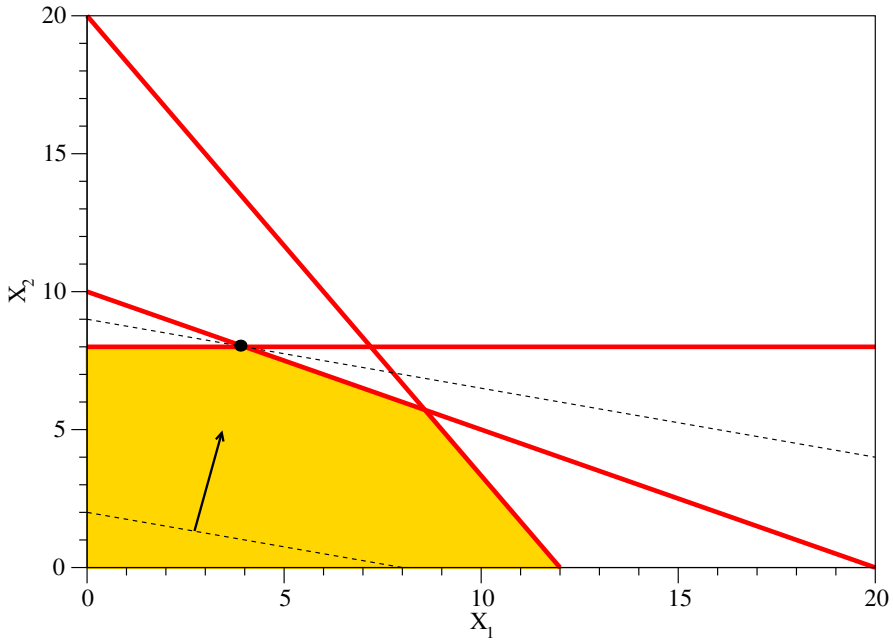
2. subject to the linear constraints:

$$\begin{aligned} 40 \cdot x_1 + 24 \cdot x_2 &\leq 480 \\ 24 \cdot x_1 + 48 \cdot x_2 &\leq 480 \\ 60 \cdot x_2 &\leq 480 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Admissible solutions:

- $x_1 = 0 \wedge x_2 = 0 \implies z = 0$
- $x_2 = 0 \curvearrowright x_1 = 12 \implies z = 120$
- $x_1 = 0 \curvearrowright x_2 = 8 \implies z = 320$

LP Problem: Graphical Solution



$$\begin{aligned} 40 \cdot x_1 + 24 \cdot x_2 \leq 480 &\iff x_2 \leq -\tfrac{5}{3} \cdot x_1 + 20 \\ 24 \cdot x_1 + 48 \cdot x_2 \leq 480 &\iff x_2 \leq -\tfrac{1}{2} \cdot x_1 + 10 \\ 60 \cdot x_2 \leq 480 &\iff x_2 \leq 8 \end{aligned}$$

$$z = 10 \cdot x_1 + 40 \cdot x_2 \iff x_2 = -\tfrac{1}{4} \cdot x_1 + \tfrac{1}{40} \cdot z$$

LP Problem: Formal Formulation

The linear objective function is generally a sum of terms that contain weighted measurable elements from a metabolic model.

Maximize:

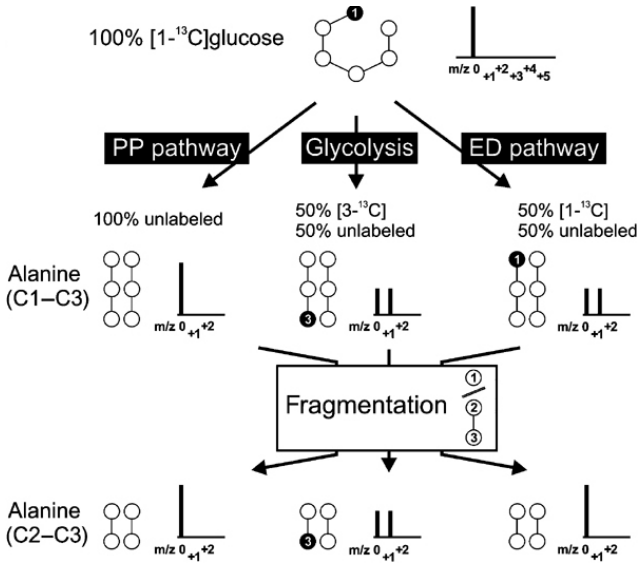
$$Z = c_1 \cdot x_1 + c_2 \cdot x_2 + \cdots = \mathbf{c}^T \cdot \mathbf{x}$$

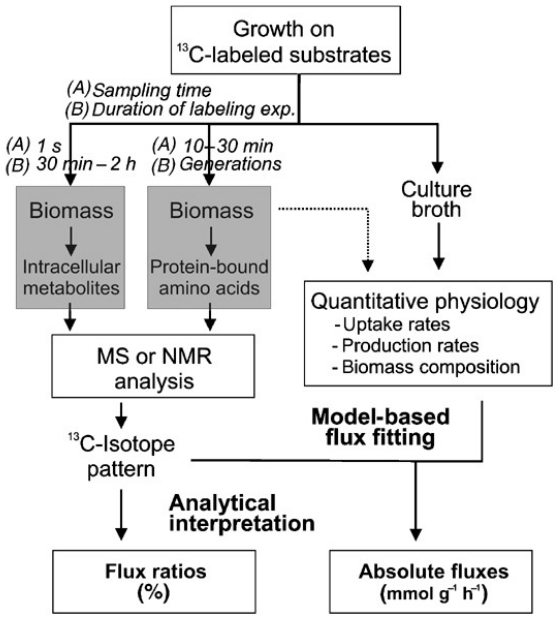
Subject to:

$$\begin{aligned} a_{11} \cdot x_1 + a_{12} \cdot x_2 + &\quad \cdots \quad a_{1n} \cdot x_n \leq b_1 \\ a_{12} \cdot x_1 + a_{22} \cdot x_2 + &\quad \cdots \quad a_{2n} \cdot x_n \leq b_2 \\ &\quad \vdots \qquad\qquad \vdots \\ a_{m1} \cdot x_1 + a_{m2} \cdot x_2 + &\quad \cdots \quad a_{mn} \cdot x_n \leq b_m \end{aligned}$$

$$\mathbf{A} \cdot x \leq b$$

How to measure Fluxes experimentally?

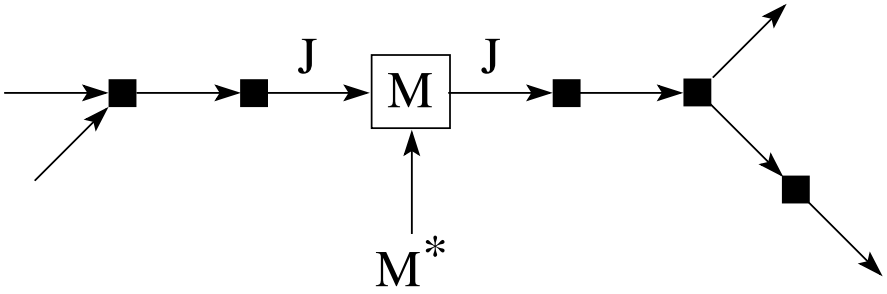




Metabolic fluxes can not be measured directly but must be inferred from isotopomere patterns.

Sauer, U (2006), Metabolic network in motion: ^{13}C-based flux analysis, *Mol Sys Biol* **2**:62 | doi:10.1038/msb4100109

Direct determination of metabolic flux



$$\frac{\mathrm{d}M}{\mathrm{d}t} = -\left(\frac{M^\star}{M}\right) \cdot J$$

J and M are constant due to metabolic steady-state therefore integration gives

$$\log\left(\frac{M^\star(t)}{M^\star(0)}\right) = -\frac{J}{M} \cdot t$$

Hence the unknown flux J can be determined from a semilog plot of "radioactive" counts versus time.

Requirements: (1) $M^\star$ be transported into the cell (2) intensity of $M^\star$ be measurable as a function of time.

Isotopomers

Are defined as isomeres of a metabolit that differ only in the labeling state of their individual atoms

(e.g. carbon [^{12}C, ^{13}C], hydrogen [^{1}H, ^{2}H] or oxygen [^{16}O, ^{17}O, ^{18}O].

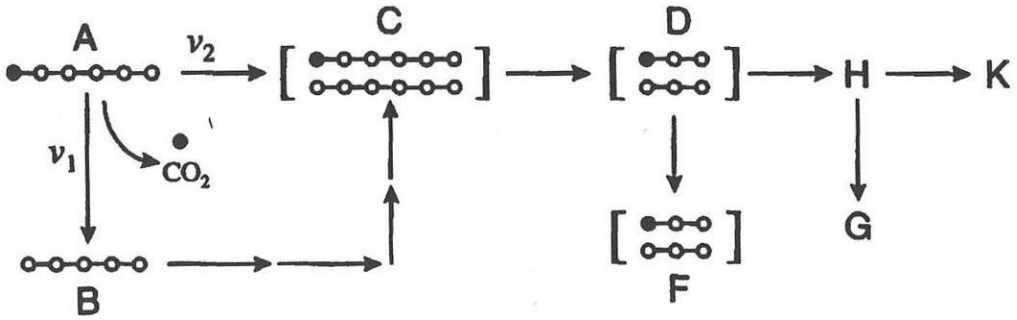
2^N isotopomeres are possible for a metabolite with N atoms that may be in one of two states (unlabeled or labeled).

Example (glucose $C_6H_{12}O_6$)

Atoms	# of Isotopomeres	
C	6.400×10^1	$(2^6 = 64)$
O	7.260×10^2	$(3^6 = 726)$
H	4.096×10^3	$(2^{12} = 4096)$
C, H	2.621×10^5	$(2^6 \times 2^{12} = 262144)$
C, H, O	1.911×10^8	$(2^6 \times 2^{12} \times 3^6 = 191102976)$

Determination of Flux Split Ratios

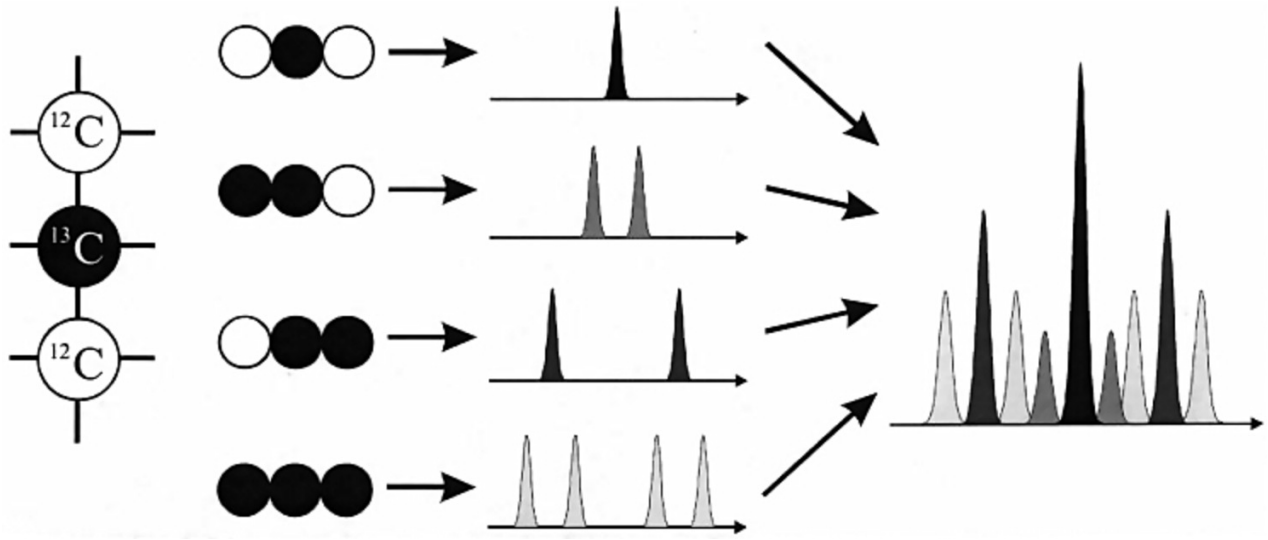
Method works only at both **metabolic** and isotopic steady state. In isotopic steady state the relative population all the isotopomeres of a metabolite is constant.



1. A is a six carbon compound whose first atom is labeled.
2. Competing/alternative pathways must introduce asymmetries.
3. The pathway via intermediate B produces unlabeled C.
4. Via the "direct" pathway the label is retained.
5. The label enrichment in C is directly proportional to the rate of ν_2 relative to the total rate of A consumtion $(\nu_1 + \nu_2)$.

Measuring Isotopomere Distribution

Any experimental technique capable of detecting differences between isotopomeres can be used to measure the labeling state.



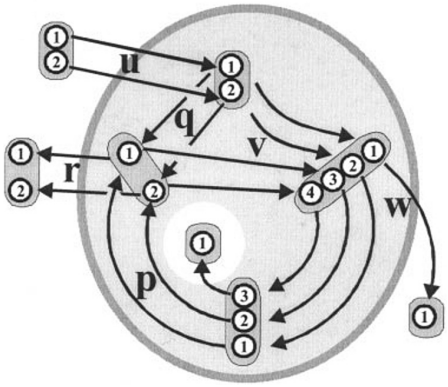
The two dominating technologies are:

1. Nuclear magnetic resonance spectroskopy (NMR).
2. Mass spectrometry (MS).

Figure adapted from [Wiechert, 2001]

Atom Transition Network

Estimating fluxes from isotopomere patterns is an inverse problem.

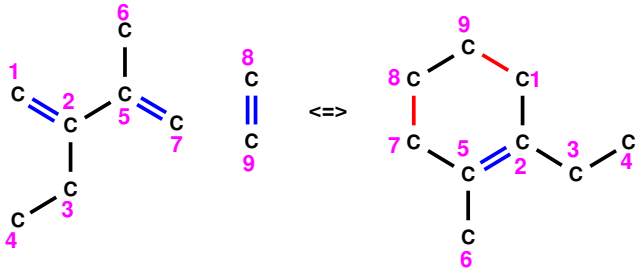


The *atom-atom mapping* between reaction educts and products must be known.

Getting this information is an *NP*-hard problem.

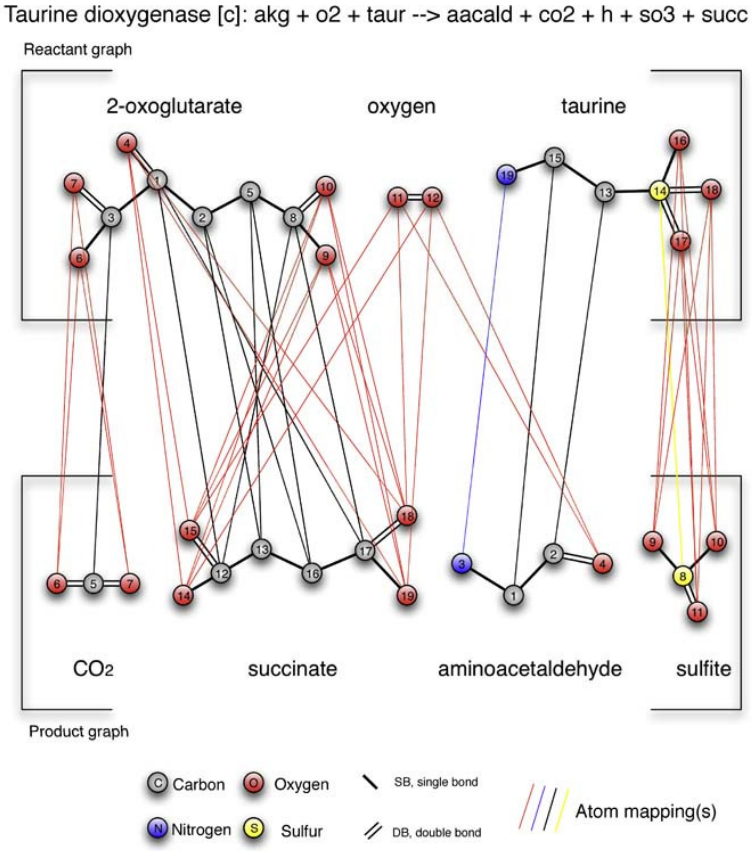
Therefore most approaches use a heuristic optimization principle e.g.

"minimal chemical distance".



Chen WL, Chen DZ, Taylor KT (2013) Automatic reaction mapping and reaction center detection. *WIREs Comput Mol Sci* | doi:10.1002/wcms.1140

Degeneracy of Atom-Atom Mappings



Increasing degeneracy:

- Rotational symmetric molecules.
- Equivalence of oxygen atoms in carboxyl or phosphate groups.

Restricting degeneracy:

- Prochiral carbone centers.
- Molecules with a center of inversion but lacking of a rotational axis.

Ravikirthi P et al (2011), Construction of an *E. Coli* Genome-Scale Atom Mapping Model for MFA Calculations, *Biotechnol Bioeng* **108**:1372-1382 | doi:10.1002/bit.23070

Further Reading

Gagneur J, Klamt S.
Computation of elementary modes: a unifying framework and the new binary approach.
BMC Bioinf **5**:175 2004 | doi:10.1186/1471-2105-5-175

Wagner C, Urbanczik R.
The Geometry of the Flux Cone of a Metabolic Network.
Biophys J **89**:38373845, 2005 | doi:10.1529/biophysj.104.055129

Price ND, Reed JL, Palsson BO.
Genome-scale models of microbial cells: Evaluating the consequences of constraints.
Nature Rev Nicrobiol **2**:886-897, 2004 | doi:10.1038/nrmicro1023

Orth JD, Ines Thiele I, Palsson BO
What is flux balance analysis?
Nature Biotech **28**:245-248, 2010 | doi:10.1038/nbt.1614

Antoniewicz MR, Kelleher JK, Stephanopoulos G.
Elementary metabolic units (EMU): A novel framework for modeling isotopic distributions.
Metab Eng **9**:68-86, 2007 | doi:10.1016/j.ymben.2006.09.001

Sauer U.
Metabolic networks in motion: ^{13}C flux analysis.
Mol Sys Biol **2**:62, 2006 | doi:10.1038/msb4100109

Wiechert W.
^{13}C Metabolic Flux Analysis.
Metab Eng **3**(3):195–206, 2001 | doi:10.1006/mben.2001.0187

Chen WL, Chen DZ, Taylor KT.
Automated reaction mapping and reaction center detection.
WIREs Comput Mol Sci 2013 | doi:10.1002/wcms.1140