Metabolic Pathway Analysis

Stefan Schuster
Friedrich Schiller University Jena
Dept. of Bioinformatics
Introduction

- Analysis of metabolic systems requires theoretical methods due to high complexity
- Major challenge: clarifying relationship between structure and function in complex intracellular networks
- Study of robustness to enzyme deficiencies and knock-out mutations is of high medical and biotechnological relevance
Theoretical Methods

- Dynamic Simulation
- Stability and bifurcation analyses
- Metabolic Control Analysis (MCA)
- Metabolic Pathway Analysis
- Metabolic Flux Analysis (MFA)
- Optimization
- and others
Theoretical Methods

- Dynamic Simulation
- Stability and bifurcation analyses
- Metabolic Control Analysis (MCA)
- Metabolic Pathway Analysis
- Metabolic Flux Analysis (MFA)
- Optimization
- and others
Metabolic Pathway Analysis (or Metabolic Network Analysis)

- Decomposition of the network into the smallest functional entities (metabolic pathways)
- Does not require knowledge of kinetic parameters!!
- Uses stoichiometric coefficients and reversibility/irreversibility of reactions
History of pathway analysis

- „Direct mechanisms“ in chemistry (Milner 1964, Happel & Sellers 1982)
- Clarke 1980 „extreme currents“
- Seressiotis & Bailey 1986 „biochemical pathways“
- Leiser & Blum 1987 „fundamental modes“
- Mavrovouniotis et al. 1990 „biochemical pathways“
- Fell 1990 „linearly independent basis vectors“
- Schuster & Hilgetag 1994 „elementary flux modes“
- Liao et al. 1996 „basic reaction modes“
- Schilling, Letscher and Palsson 2000 „extreme pathways“
Mathematical background

Stoichiometry matrix

Example:

$N = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix}$
Steady-state condition

Balance equations for metabolites:

\[
\frac{dS_i}{dt} = \sum_j n_{ij} v_j
\]

\[
dS/dt = NV(S)
\]

At any stationary state, this simplifies to:

\[
NV(S) = 0
\]
Steady-state condition $\mathbf{N}V(S) = 0$

If the kinetic parameters were known, this could be solved for $S$.

If not, one can try to solve it for $V$. The equation system is linear in $V$. However, usually there is a manifold of solutions.

Mathematically: kernel (null-space) of $\mathbf{N}$. Spanned by basis vectors. These are not unique.
Use of null-space

The basis vectors can be gathered in a matrix, $K$. They can be interpreted as biochemical routes across the system.

If some row in $K$ is a null row, the corresponding reaction is at thermodynamic equilibrium in any steady state of the system.

Example:

$$K = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$
Use of null-space (2)

It allows one to determine „enzyme subsets“ = sets of enzymes that always operate together at steady, in fixed flux proportions.

The rows in $K$ corresponding to the reactions of an enzyme subset are proportional to each other.

Example:
Enzyme subsets: \{1,6\}, \{2,3\}, \{4,5\}

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

Extensions of the concept of „enzyme subsets“

Representation of rows of null-space matrix as vectors in space:

If \( \cos(\phi) = \pm 1 \), then the enzymes belong to the same subset

If \( \cos(\phi) = 0 \), then reactions uncoupled

Otherwise, enzymes partially coupled.

Extensions of the concept of „enzyme subsets“ (2)

Inclusion of information about irreversibility

(1) **Directional coupling** ($v_1 \Rightarrow v_2$), if a non-zero flux for $v_1$ implies a non-zero flux for $v_2$ but not necessarily the reverse.

(2) **Partial coupling** ($v_1 \leftrightarrow v_2$), if a non-zero flux for $v_1$ implies a non-zero, though variable, flux for $v_2$ and vice versa.

(3) **Full coupling** ($v_1 \Leftrightarrow v_2$), if a non-zero flux for $v_1$ implies not only a non-zero but also a fixed flux for $v_2$ and vice versa. – Enzyme subset.

**Flux coupling analysis**

**Drawbacks of null-space**

- The basis vectors are not given uniquely.
- They are not necessarily the simplest possible.
- They do not necessarily comply with the directionality of irreversible reactions.
- They do not always properly describe knock-outs.

\[ K = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \]
Drawbacks of null-space

They do not always properly describe knock-outs.

After knock-out of enzyme 1, the route \{-2, 3\} remains!


**non-elementary flux mode**

**elementary flux modes**
An **elementary mode** is a minimal set of enzymes that can operate at steady state with all irreversible reactions used in the appropriate direction.

The enzymes are **weighted** by the relative flux they carry.

The elementary modes are **unique** up to scaling.

All flux distributions in the living cell are **non-negative linear combinations** of elementary modes.
Non-Decomposability property:

For any elementary mode, there is no other flux vector that uses only a proper subset of the enzymes used by the elementary mode.

For example, \{HK, PGI, PFK, FBPase\} is not elementary if \{HK, PGI, PFK\} is an admissible flux distribution.
**Simple example:**

Elementary modes:

\[
\begin{pmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & -1 \\
\end{pmatrix}
\]

They describe knock-outs properly.
Mathematical background (cont.)

Steady-state condition $\mathbf{N} \mathbf{V} = 0$
Sign restriction for irreversible fluxes: $\mathbf{V}^{\text{irr}} \geq 0$

This represents a linear equation/inequality system.

Solution is a convex region.

All edges correspond to elementary modes.

In addition, there may be elementary modes in the interior.
Geometrical interpretation

Elementary modes correspond to generating vectors (edges) of a convex polyhedral cone (= pyramid) in flux space (if all modes are irreversible)
flux1
flux2
flux3

generating vectors
If the system involves reversible reactions, there may be elementary modes in the interior of the cone.

*Example:*
Flux cone:

There are elementary modes in the interior of the cone.
Mathematical properties of elementary modes

Any vector representing an elementary mode involves at least \( \text{dim(null-space of } N) - 1 \) zero components.

**Example:**

\[
\begin{pmatrix}
1 & 1 \\
1 & 0 \\
0 & 1 \\
\end{pmatrix}
\]

\( \text{dim(null-space of } N) = 2 \)

Elementary modes:

\[
\begin{pmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & -1 \\
\end{pmatrix}
\]

(Schuster et al., *J. Math. Biol.* 2002, after results in theoretical chemistry by Milner et al.)
Mathematical properties of elementary modes (2)

A flux mode $V$ is elementary if and only if the null-space of the submatrix of $N$ that only involves the reactions of $V$ is of dimension one.


e.g. elementary mode:

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

$N = (1 \ -1) \Rightarrow \dim = 1$
Biochemical examples
Part of monosaccharide metabolism

Red: external metabolites
1\textsuperscript{st} elementary mode: glycolysis
2nd elementary mode: fructose-bisphosphate cycle
4 out of 7 elementary modes in glycolysis-pentose-phosphate system
Optimization: Maximizing molar yields

**Chemical Reactions**

- **CO₂** + **Ru5P** → **X5P** + **S7P** + **G6P**
- **Ru5P** → **R5P** → **G6P**
- **G6P** → **F6P**
- **F6P** → **GAP**
- **GAP** → **DHAP**
- **DHAP** → **NAD** → **NADH**
- **ATP** → **PEP**
- **ATP** → **3PG**
- **ATP** → **1.3BPG**
- **ATP** → **2PG**

**Molar Yields**

- **ATP:G6P yield = 3**
- **ATP:G6P yield = 2**
Synthesis of lysine in *E. coli*
Elementary mode with the highest lysine : phosphoglycerate yield

(thick arrows: twofold value of flux)
Maximization of tryptophan:glucose yield

Model of 65 reactions in the central metabolism of *E. coli*. 26 elementary modes. 2 modes with highest tryptophan:glucose yield: 0.451.

Can sugars be produced from lipids?

(Work with David Fell, Oxford, Luis Figueiredo and Christoph Kaleta, Jena)

- Known in biochemistry for a long time that many bacteria and plants can produce sugars from even-chain fatty acids (via C2 units) while animals cannot.
AcCoA is linked with glucose by a chain of reactions. However, no elementary mode realizes this conversion along that chain.
Elementary mode representing conversion of AcCoA into glucose. It requires the glyoxylate shunt.
The glyoxylate shunt is present in green plants, yeast, many bacteria (e.g. *E. coli*) and others and – as the only clade of animals – in nematodes.

This example shows that a description by usual graphs in the sense of graph theory is insufficient…


A successful theoretical prediction

Red elementary mode: Usual TCA cycle
Blue elementary mode: Catabolic pathway
predicted in Liao et al. (1996) and Schuster et al. (1999) for *E. coli*. 

Glucose

- PEP
- Pyr
- AcCoA
- Oxac
- Cit
- IsoCit
- OG
- CO₂
- Mal
- Gly
- Fum
- Succ
- SucCoA
- CO₂
A successful theoretical prediction

Glucose

Red elementary mode: Usual TCA cycle
Blue elementary mode: Catabolic pathway
predicted in Liao et al. (1996) and Schuster et al. (1999) Experimental hints in Wick et al. (2001). Experimental proof in:

E. Fischer and U. Sauer: A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli,
Crassulacean Acid Metabolism (CAM)

(Work with David Fell, Oxford)

- Variant of photosynthesis employed by a range of plants (e.g. cacti) as an adaptation to arid conditions
- To reduce water loss, stomata are closed during daytime
- At nighttime, $\text{PEP} + \text{CO}_2 \rightarrow \text{oxaloacetate} \rightarrow \text{malate}$
- At daytime, malate $\rightarrow$ pyruvate (or PEP) + CO$_2$ $\rightarrow$ carbohydrates
CAM metabolism during daytime

[Diagram showing the metabolic pathways involving hexose, PEP, TP, CO₂, Pi, mal, oxac, cytosol, chloroplast, and starch.]
Elementary modes

A) Hexose synthesis via malic enzyme as occurring in Agavaceae and Dracaenaceae

B) Starch synthesis via malic enzyme as occurring in Cactaceae and Crassulacea

Dracaena

Ferocactus
Simultaneous starch and hexose synthesis via malic enzyme as occurring in:

**Clusia minor**

Hexose synthesis via PEPCK as occurring in *Clusia rosea* and in:

*Ananus comosus* = pineapple
Starch synthesis via PEPCK as occurring in Asclepiadaceae

Simultaneous starch and hexose synthesis via PEPCK as occurring in:

Caralluma hexagona

Aloe vera
„Pure“ pathways

- In a review by Christopher and Holtum (1996), only cases A), B), D), and E) were given as “pure” functionalities. F) was considered as a superposition, and C) was not mentioned.
- However, F) is an elementary mode as well, although it produces two products. It does not use the triose phosphate transporter.
- The systematic overview provided by elementary modes enables one to look for missing examples. Case C) is indeed realized in *Clusia minor* (Borland et al, 1994).
- Interestingly, (almost) pure elementary modes are realized here. No redundancy?

Algorithms for computing elementary modes

1. Modified Gauss-Jordan method starting with tableau ($N^T \ I$). Pairwise combination of rows so that one column of $N^T$ after the other becomes null vector.

2. Column operations on the null-space matrix.
Empirically faster than 1. on biochemical networks.
Example:

\[ T^{(0)} = \begin{pmatrix} 1 & 0 & : & 1 & 0 & 0 & 0 \\ -1 & 0 & : & 0 & 1 & 0 & 0 \\ -1 & 1 & : & 0 & 0 & 1 & 0 \\ 1 & -1 & : & 0 & 0 & 0 & 1 \end{pmatrix} \]
$$T^{(0)} = \begin{pmatrix}
1 & 0 & \vdots & 1 & 0 & 0 & 0 \\
-1 & 0 & \vdots & 0 & 1 & 0 & 0 \\
-1 & 1 & \vdots & 0 & 0 & 1 & 0 \\
1 & -1 & \vdots & 0 & 0 & 0 & 1
\end{pmatrix}$$

$$T^{(1)} = \begin{pmatrix}
0 & 0 & \vdots & 1 & 1 & 0 & 0 \\
0 & 1 & \vdots & 1 & 0 & 1 & 0 \\
0 & -1 & \vdots & 0 & 1 & 0 & 1 \\
0 & 0 & \vdots & 0 & 0 & 1 & 1
\end{pmatrix}$$

These two rows should not be combined.
Final tableau:

\[
T^{(2)} = \begin{pmatrix}
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\]
Algorithm is faster, if this column is processed first.

\[
T^{(0)} = \begin{pmatrix}
1 & 0 & \vdots & 1 & 0 & 0 & 0 \\
-1 & 0 & \vdots & 0 & 1 & 0 & 0 \\
-1 & 1 & \vdots & 0 & 0 & 1 & 0 \\
1 & -1 & \vdots & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]
Runtime complexity

- Not yet completely clear
  - **Theorem 9.** Given a matrix $N$, *counting* the number of elementary modes is $\#P$-complete.
  - **Theorem 10.** In case all reactions in a metabolic network are reversible, the elementary modes can be *enumerated* in polynomial time.
- **Open question:** Can elementary modes be *enumerated* in polynomial time if some reactions are irreversible?
**#P (sharp P) Complexity class**

- An **NP problem** is often of the form, "Are there any solutions that satisfy certain constraints?" For example:
  - Are there any subsets of a list of integers that add up to zero? (subset sum problem)
  - Are there any Hamiltonian cycles in a given graph with cost less than 100?
- The corresponding **#P problems** ask "how many" rather than "are there any". For example:
  - How many subsets of a list of integers add up to zero?
  - How many Hamiltonian cycles in a given graph have cost less than 100?
Software involving routines for computing elementary modes

EMPATH - J. Woods
METATOOL - Th. Pfeiffer, F. Moldenhauer, A. von Kamp (In versions 5.x, Wagner algorithm)
GEPASI - P. Mendes
JARNAC - H. Sauro
In-Silico-Discovery™ - K. Mauch
FluxAnalyzer (in MATLAB) - S. Klamt
ScrumPy - M. Poolman
Alternative algorithm in MATLAB – C. Wagner, R. Urbanczik
PySCeS – B. Olivier et al.

On-line computation:
pHpMetatool - H. Höpfner, M. Lange
Related concept: Extreme pathways

C.H. Schilling, D. Letscher and B.O. Palsson, 

Distinction between internal and exchange reactions, all internal reversible reactions are split up into forward and reverse steps

Then, the convex basis is calculated. Spurious cyclic modes are discarded.
Advantages of extreme pathways:
• Smaller number
• Correspond to edges of flux cone

Drawbacks of extreme pathways:
• Flux cone is higher-dimensional
• Often not all relevant biochemical pathways represented
• Knock-outs not always properly described
• Often route with maximal yield not covered

However, this depends on network configuration. Originally, Schilling et al. (2000) proposed adding exchange reaction for each external metabolite.
Network reconfiguration

1. Decomposition of internal reversible reactions into forward and reverse steps

2. Optionally: inclusion of (non-decomposed) exchange reactions for each external metabolite.

Now, there is a 1:1 correspondence between extreme pathways and elementary modes!
Combinatorial explosion of elementary modes

A) S

B) S

C) [S]

2*3 modes

S external:

2+3 modes
Proposed decomposition procedure

- In addition to the pre-defined external metabolites, set all metabolites participating in more than 4 reactions to external status
- Thus, the network disintegrates into subnetworks
- Determine the elementary flux modes of the subnetworks separately

Mycoplasma pneumoniae

Subsystem 1
Sugar import

Subsystem 2
PPP, glycolysis, fragm. lipid metab.

Subsystem 3
Nucleotide metab.

Subsystem 4
Lower glycolysis

Subsystem 5
C1 pool

Subsystem 6
Arginine degrad.

Yellow boxes: additional external metabolites

Met
f-Met
Serine
Glycine
Formate
NH3
NH4
Ornithine
Carbamate
CO2
Summary

- Elementary modes are an appropriate concept to describe biochemical pathways in wild-type and mutants.

- Information about network structure can be used to derive far-reaching conclusions about performance of metabolism, e.g. about viability of mutants.

- Elementary modes reflect specific characteristics of metabolic networks such as steady-state mass flow, thermodynamic constraints and molar yields.
Pathway analysis is well-suited for computing maximal and submaximal molar yields.

Many metabolic systems in various organisms have been analysed in this way. In some cases new pathways discovered.

Relevant applications: knockout studies (biotechnology) and enzyme deficiencies (medicine).

Work still to be done on decomposition methods (combinatorial explosion).
Acknowledgements

• David Fell (Brookes U Oxford)
• Thomas Dandekar (U Würzburg)
• Steffen Klamt (MPI Magdeburg)
• Jörg Stelling (ETH Zürich)
• Thomas Pfeiffer (Harvard)
• late Reinhart Heinrich (HU Berlin)
• Ina Koch (TFH and MPI Berlin)
• and many others

• Acknowledgement to DFG, BMBF and FCT (Portugal) for financial support