Predicting pathways in genome-scale metabolic networks

Stefan Schuster
Dept. of Bioinformatics
Friedrich Schiller University Jena
Famous people at Jena University

Friedrich Schiller (1759-1805)

Matthias Schleiden (1804-1881)
Discoverer of the living plant cell

Ernst Haeckel (1834-1919, Biogenetic rule)
Introduction

• Metabolism is bridge between genotype and phenotype
• Technological relevance of metabolism: Synthesis of specific products (antibiotics, amino acids, ethanol, citric acid, dyes, odorants), degradation of xenobiotics
• Medical relevance, e.g. diseases based on enzyme deficiencies
• Metabolic networks are complex due to their size and the presence of bimolecular reactions
Introduction (2)

• Structure and (nonlinear) dynamics of metabolic networks cannot be understood intuitively
• Theoretical methods needed
• Traditional graph theory is here insufficient
• These methods should be systemic rather than too reductionist (Systems Biology) and they should be able to cope with genome-scale models
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 (these data can be integrated into large-scale models relatively easily)

non-elementary flux mode

S1 → S2 → S3 → S4 → P2
P1 → S1 → S2 → S4 → P2

elementary flux modes
ON ELEMENTARY FLUX MODES IN BIOCHEMICAL REACTION SYSTEMS AT STEADY STATE

STEFAN SCHUSTER and CLAUS HILGETAG
Fachbereich Biologie, Institut für Biophysik, Humboldt-Universität zu Berlin, Invalidenstr. 42, D-10115 Berlin, Germany

Received 11 August 1993
Revised 3 January 1994

ABSTRACT
A mathematical definition of the concept of elementary mode is given so as to apply to biochemical reaction systems subsisting at steady state. This definition relates to existing concepts of null-space vectors and includes a condition of simplicity. It is shown that for systems in which all fluxes have fixed signs, all elementary modes are given by the generating vectors of a convex cone and can, thus, be computed by an existing algorithm. The present analysis allows for the more general case that some reactions can proceed in either direction. Basic ideas on how to compute the complete set of elementary modes in this situation are outlined and verified by way of several examples, with one of them representing glycolysis and gluconeogenesis. These examples show that the elementary modes can be interpreted in terms of the particular biochemical functions of the network. The relationships to (futile) substrate cycles are elucidated.

Keywords: Biochemical pathways, mathematical modelling, elementary mode, substrate cycle.

1. Introduction
Investigation of steady states plays an important role in the modelling of biochemical reaction systems, because virtually stationary regimes are frequently encountered in experimental settings and under in-vivo conditions [3,9,16]. As for systems with oscillatory behaviour, stationary concentrations can be calculated as average values over a longer time span. In both situations, reaction rates have to fulfill balance
Thomas Pfeiffer

Now in Martin Nowak’s group at Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

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.
Simple example:

Elementary modes:

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

They describe knock-outs properly.
**Simple example:**

Elementary modes:

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

They describe knock-outs properly.
Mathematical background

Steady-state condition \( NV = 0 \)

Sign restriction for irreversible fluxes: \( V_{irr} \geq 0 \)

This represents a linear equation/inequality system.

Solution is a convex region.

All edges correspond to elementary modes.
Geometrical interpretation

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

flux1

flux2

flux3
**Mathematical properties of elementary modes**

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

*Example:*

![Diagram](image)

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

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

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:

$$
\begin{pmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & -1
\end{pmatrix}
\Rightarrow N = (1 \ -1) \Rightarrow \text{dim} = 1
$$
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.


![Tryptophan structure]
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
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) 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*,

Can even-chain fatty acids converted into glucose?

• Excess sugar in human diet is converted into storage lipids, mainly triglycerides
• Is reverse transformation feasible? Triglyceride → sugar?
If AcCoA, glucose, CO$_2$ and all cofactors are considered external, there is NO elementary mode consuming AcCoA, nor any one producing glucose.

Intuitive explanation by regarding oxaloacetate or CO$_2$. 
Green plants, fungi, many bacteria (e.g. *E. coli*) and – as the only clade of animals – nematodes harbour the glyoxylate shunt. Then, there is an elementary mode representing conversion of AcCoA (and of fatty acids) into glucose.
This example shows that a description by usual graphs in the sense of graph theory is insufficient…


Interesting question: Is the conversion feasible in genome-scale networks?
Considering circadian rhythms

- Diploma student Sascha Schäuble computed EFMs for amino acid metabolism in *Chlamydomonas rheinhardtii* (cooperation with Maria Mittag, Jena)
- Circadian rhythm was taken into account by three distinct phases (i.e., sets of conditions)
  - In day-phase, ATP and NADH are sufficiently available, thus set to external status
  - In first phase of night, ATP and NADH need to be balanced (are internal), while triose phosphates are sufficiently available
  - In second phase of night, ATP, NADH and triose phosphates need to be balanced
Clustering of elementary modes

• In large networks, (very) large number of elementary modes
• As long as computation is feasible, clustering of EFMs is sensible for better handling and interpretation
Combinatorial explosion of elementary modes

A)

B)

2*3*2 modes
Decomposition procedure

A)

B)

C)

S external: 2+3 modes

2*3 modes

[S]
Mycoplasma pneumoniae

Schuster et al., *Bioinformatics*, 2002
Elementary flux patterns

• Delimit a smaller subsystem within a large (e.g. genome-scale) network
• Check which binary flux patterns in the subsystem are consistent with a flux distribution in the entire system

Computing elementary flux patterns

• Check binary flux patterns in the subsystem as to whether consistent with a flux distribution in the entire system
• Done by mixed-integer linear programming (similar to FBA)
• If consistent, reconstruct elementary mode from binary pattern.
• Not all $2^k$ binary flux patterns need be tested by using constraint that each new flux pattern cannot be written as a combination of previously found elementary flux patterns.
• Is fixed-parameter tractability problem
Figure 1. Examples for problems in elementary mode analysis. (A) Condensed network of glycolysis (dark reactions) and Entner–Doudoroff pathway (gray reactions). The pentose phosphate pathway has been omitted for clarity. In most analyses of glycolysis, the Entner–Doudoroff pathway is not considered. Hence, not all pathways from glucose into the TCA cycle are found, and the wrong conclusion might be drawn that only the pentose–phosphate pathway (not shown) can be used to bypass the knockout of one of the enzymes converting G6P to FDP. In vivo the knockout of the corresponding reactions is partially bypassed by a flux through the Entner–Doudoroff pathway (Fischer and Sauer 2003). (B) Modeling the Entner–Doudoroff pathway by adding an outflow of G6P and an inflow of C3P and Pyr only partially resolves the problem (thick arrows). Such an approach is often used in elementary mode analysis to avoid the consideration of some pathways in detail (e.g., the outflow of succinyl-CoA from the TCA cycle in Schuster et al. 1999, analyzed in Results). However, this can lead to fluxes like in C,
The TCA-cycle/glyoxylate shunt/amino acid synthesis system revisited

- Elementary-flux pattern analysis shows that only 10 of them are feasible.
- All modes producing succinyl-CoA are not because they require additional inputs from the system.
while aspartate is produced by three elementary flux patterns (Fig. 5B–D) and glutamate by four elementary flux patterns (Fig. 5E–H).

The elementary flux pattern producing alanine contains only two reactions, the production of alanine from pyruvate and glutamate as well as the outflow of extracellular alanine. Hence, the

into the subsystem would be a union of two elementary flux patterns. The first is the elementary flux pattern containing the reactions of the subsystem corresponding to the alternative glyoxylate-producing pathway; and the second, the aspartate-producing elementary flux pattern in Figure 6.

Figure 5. Amino acid producing elementary flux patterns. (A) Alanine; (B–D) aspartate; (E–H) glutamate. Black reactions belong to the elementary flux pattern; gray reactions are the remaining reactions of the subsystem. A list of abbreviations can be found in Supplemental materials S2 and S3.
Predicting pathways

By elementary flux patterns, several interesting pathways in *E. coli* could be predicted. For example, a bypass of part of the TCA cycle.
GABA shunt in *E. coli*

**Figure 4.** Alternative pathways in the central metabolism of *E. coli*. Dashed arrows represent condensed reactions. (A) Alternative glyoxylate producing pathway; (B) Entner–Doudoroff pathway; (C) glycerate pathway; (D) GABA-shunt. The lower pathway represents the standard route from oxoglutarate to succinate in the TCA cycle, and the upper pathway depicts the GABA shunt. A list of abbreviations can be found in Supplemental materials S2 and S3.
Known: GABA shunt in plants

Probably hitherto unknown pathway: Synthesis of glyoxylate from purines

Figure 4. Alternative pathways in the central metabolism of E. coli. Dashed arrows represent condensed reactions. (A) Alternative glyoxylate producing pathway; (B) Entner–Doudoroff pathway; (C) glycerate pathway; (D) GABA-shunt. The lower pathway represents the standard route from oxoglutarate to succinate in the TCA cycle, and the upper pathway depicts the GABA shunt. A list of abbreviations can be found in Supplemental materials S2 and S3.
Computing the shortest elementary modes in genome-scale metabolic networks

• First compute the shortest elementary mode, then the second-shortest and so on.
• Done by mixed-integer linear programming.
• Assigning a binary variable $z_i$ to each reaction $i$ such that $z_i = 1$ if reaction is operative and 0 otherwise
• Minimize $\sum z_i$ under certain side constraints such as $\sum z_i >= 1$ and steady-state condition

Computing the shortest elementary modes

• Is NP-hard in length of elementary mode. However, if this is small, then feasible
• Elementary modes can be computed consecutively although computation time for each one is longer than in previous algorithms
• We applied this to lysine production in a genome-scale network of Corynebacterium glutamicum
Signalling in enzyme cascades

Obviously, elementary signalling routes
How define elementary signalling routes mathematically?

• Signalling systems are not always at steady state. Propagation of signals is time-dependent process.
• However: Averaged over longer time spans, also signalling systems must fulfill steady-state condition because system must regenerate.
Calculating elementary modes gives trivial result that each cycle corresponds to one mode. Flow of information is not reflected.
Signal amplification

• Mass flow not linked with information flow.
• However: Signal amplification requires that each activated enzyme must catalyse at least one further activation.
• Minimum condition: Each activated enzyme catalyses exactly one further activation.
• Thus, operational stoichiometric coupling of cascade levels.
• $E_1^* + E_2 \rightarrow E_1 + E_2^*$
The elementary routes thus calculated exactly give the signalling routes.

Conclusions

- 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.
- Two tendencies in modelling: large-scale vs. medium-scale.
- Analysis of both types of models allows interesting conclusions.
- Some questions can only be answered in whole-cell models, for example: Can some product principally be synthesized from a given substrate?
Conclusions (2)

- Many metabolic systems in various organisms have been analysed in this way. Several new pathways discovered
- Elementary modes compatible with flux distributions in whole cells can be computed
- Elementary-mode analysis is applicable to signalling in enzyme cascades
Dept. of Bioinformatics group at the School of Biology and Pharmaceutics, Jena University

Pathway analysis group:
Luis de Figueiredo, Christoph Kaleta, Jörn Behre, Sascha Schäuble, Martin Kötzing, partly: Dr. Ines Heiland, Kathrin Bohl
Cooperations

- David Fell (Brookes U Oxford)
- Thomas Pfeiffer (Harvard)
- Thomas Dandekar (Würzburg U)
- Eytan Ruppin (Tel Aviv U)
- Francisco Planes (U de Navarra)
- John Beasley (Brunel U)
- Steffen Klamt (MPI Magdeburg)
- Thomas Wilhelm (IFR Norwich)
- Maria Mittag, Jena
- and many others …

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