Metabolic Networks

Ulf Leser and Michael Weidlich
This Lecture

- **Introduction**
  - Systems biology & modelling
  - Metabolism & metabolic networks

- **Network reconstruction**
  - Strategy & workflow

- **Mathematical representation**
  - The stoichiometric matrix
  - Convex analysis & solution space
  - Elementary Modes & Extreme Pathways

- **Constraint based network analysis**
  - Flux Balance Analysis & Optimization
  - Tools
Motivation

Complex systems in biology

Metabolic network

Neural network
„More than the sum of its parts“

- Many biological systems have emergent properties (they are called irreducible since they cannot be fully comprehended when broken down into smaller segments).

- Studying a neuron will not explain how information is stored in the brain? (fire patterns)
"Systems biology...is about putting together rather than taking apart, integration rather than reduction. It requires that we develop ways of thinking about integration that are as rigorous as our reductionist programmes, but different....It means changing our philosophy, in the full sense of the term" - Dennis Noble
Systems biology - Bridging the scales

From component to systems analysis

20th century

Cell

Tissue

Organ

Whole body

21st century

Ion channel gating

Diffusion

Motility

Mitosis

Protein turnover

Human lifetime
Excursion – Virtual Liver Network

Virtual Liver Network
The Virtual Liver will be a dynamic model that represents, rather than fully replicates, human liver physiology morphology and function, integrating quantitative data from all levels of organisation.

- Interdisciplinary **competence network** of experimental and theoretical research groups (since 2004)
- Liver - A most relevant research object for applications in **medicine, pharma research and nutrition**
- Example for a **german systems biology initiative**
What is a Model?

A Model is a simplified image of reality.

- Models are able to reduce reality to certain aspects.
- Definition of these aspects is guided by our comprehension of which aspects are crucial (for functionality)
- Simplification leads to understanding of complex systems
- Models should generate testable predictions (purpose driven)
Process of modelling

- The **direct problem** consists in generating data from models (i.e. doing simulations) → relativly easy
- The **inverse problem** consists in generating a model from data → generally difficult

- Modelling is an **iterative process** of both problems
Process of modelling

• The **direct problem** consists in generating data from models (i.e. doing simulations) \(\rightarrow\) **relatively easy**

• The **inverse problem** consists in generating a model from data \(\rightarrow\) **generally difficult**

• Modelling is an **iterative process** of both problems
Okay i got it!
What are we going to model ?

Metabolism
Metabolism

- is the *set of chemical reactions* that happen in living organisms to maintain life, simply spoken it is the uptake, transport and conversion of chemical entities in an organism
- These processes allow for growth and reproduction, maintenance of structures and responsiveness to environmental stress.
Metabolism and Disease

- Metabolism is in a **balanced state** which it tries to maintain
- any permanent perturbation of this balance can lead to diseases (i.e. diabetes, cancer, etc.)
Metabolism and Evolution

- Metabolism is subjected to evolutionary optimization which leads to robustness and modularity

- Redundancy - there is often more than one solution to a certain objective

- Pathways - TCA and glycolysis are shared among many different organisms
Key-players in metabolism

Chemical Compounds / Metabolites
• small molecules that are imported/exported and/or synthesized/degraded inside an organism.

Biochemical Reactions
• produce a set of metabolites (called products) from another set of compounds (called substrates)
• can be reversible or irreversible
• some are spontaneous, but most are enzymatic
• a set of reactions is often referred to as a pathway (still there is no consensus on the boundaries)
Key-players in metabolism

Enzyme
- a protein or a protein complex coded by one or several genes
- a single enzyme may accept distinct substrates and may catalyze several reactions
- a single reaction may be catalyzed by several enzymes
- the links between genes, proteins, and reactions (called GPR relationship) is nontrivial

Cofactors
- small molecules
- essential to allow the catalysis of a reaction by an enzyme
Example - tricarboxylic acid cycle (TCA)

Chemicals
Enzymes
Cofactors
Metabolic networks – Graphical representation

Metabolic networks can be represented as graphs

\[
\begin{array}{c}
A + B \rightarrow C \\
C \rightarrow D
\end{array}
\]

Possibility A

Definition of a graph: \( G = (V,E) \)
- \( V \) is the set of nodes (metabolites)
- \( E \) is the set of edges (reactions)

directed graph:
- Directed edge \( e = (v_1,v_2) \) with \( v_1 \) (start node) and \( v_2 \) (end node) of \( e \).
Metabolic networks

Metabolic networks can be represented as graphs

\[ A + B \rightarrow C \]
\[ C \rightarrow D \]

Possibility B

Bipartite Graph \( G = (U,V,E) \)

- \( U \) is one type of nodes (metabolites)
- \( V \) is the other type of nodes (reactions)
- \( E \) is the set of edges with \( e = (u,v) \) or \( e = (v,u) \)
- directed bipartite graphs are very common (not shown here)

Which graph to choose depends on the type of questions asked.
Graph analysis

- Topological network properties like clustering, centrality, degree, density, distance and neighborhood etc. can be calculated (already covered in previous lecture)

- Metabolic Networks (Global properties):
  - scale free (hubs)
  - high clustering coefficient (modules)
  - degree distribution of nodes follows a power law
This Lecture

• Introduction
  – Systems biology & modelling
  – Metabolism & metabolic networks

• Network reconstruction
  – Strategy & workflow

• Mathematical representation
  – The stoichiometric matrix
  – Convex analysis & solution space
  – Elementary Modes & Extreme Pathways

• Constraint based network analysis
  – Flux Balance Analysis & Optimization
  – Tools
(Genome Scale) Network Reconstruction

**Top-Down Strategy**

- Identify proteins based on dna sequence
- Use gene/protein annotations for enzyme activity
- Link enzyme activity with reactions
- Compile the reaction list
Sequential order can differ (network reconstruction of *S. cerevisiae* started with list of already known enzymes)

Intensive literature research

Iterative approach (Simulation)
Gene Protein Reaction (GPR) association

- Not all genes have a one-to-one relationship with their corresponding enzymes or metabolic reactions

  - Promiscuous enzymes catalyzing more than one reaction

  - Enzyme complexes consisting of different protein subunits which all need to be present for the reaction to occur

  - Iso enzymes /Isozymes catalyzing the same reaction
Database Screenshots
This Lecture

- Introduction
  - Systems biology & modelling
  - Metabolism & metabolic networks

- Network reconstruction
  - Strategy & workflow

- Mathematical representation
  - The stoichiometric matrix
  - Convex analysis & solution space
  - Elementary Modes & Extreme Pathways

- Constraint based network analysis
  - Flux Balance Analysis & Optimization
  - Tools
Stoichiometry

- The set of chemical reactions that comprise a network can be represented as a set of chemical equations

- Embedded in these chemical equations is information about reaction stoichiometry (the quantitative relationships of the reaction's reactants and products)

- Stoichiometry is invariant between organisms for the same reactions and does not change with pressure, temperature, or other conditions.  
  \[ 2 \text{H}_2 + \text{O}_2 \rightarrow 2 \text{H}_2\text{O} \]

- All this stoichiometric information can be represented in a matrix form; the stoichiometric matrix, denoted by \( S \).
Stoichiometric matrix

- **m rows**      m metabolites
- **n columns**   n reactions
- $S_{ij}$       number of molecules of the $i^{th}$ metabolite produced in $j^{th}$ reaction.

- Consumption is understood as negative production.
Example 1 - Reversible Conversion

\[ \text{CP} \iff \text{PC} \]

Simple chemical rearrangement of the molecule without any change in its elemental composition. Isomerase catalyze such reaction.

\[
\mathbf{s} = \begin{pmatrix}
\text{CP} \\
\text{PC}
\end{pmatrix}
\begin{pmatrix}
-1 & 1 \\
1 & -1
\end{pmatrix}
\]

Forward     Backward
Example 2 - Bimolecular association

\[ C + P \rightleftharpoons CP \]

Combination of two moieties, C and P, to form a new compound, e.g., the dimerization of two protein molecules, or the initial binding of a substrate to an active site on an enzyme molecules.

\[
\mathbf{S} = \begin{pmatrix}
C & -1 & 1 \\
P & -1 & 1 \\
CP & 1 & -1 \\
\end{pmatrix}
\]

Forward
Backward
Example 3 - Cofactor-Coupled Reaction

\[ C + AP \rightleftharpoons CP + A \]

One compound AP donates a moiety P to another compound C. Such reaction can be decomposed into two bimolecular association reactions.

\[ s = \begin{pmatrix}
  C & -1 & 1 \\
  AP & -1 & 1 \\
  CP & 1 & -1 \\
  A & 1 & -1
\end{pmatrix} \]
Mathematical Interpretation

Mathematically, the stoichiometric matrix $S$ is a linear transformation of the flux vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ to a vector of time derivatives of the concentration $\frac{dx}{dt} = \left( \frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_m}{dt} \right)$ or

$$S \cdot \mathbf{v} = \frac{dx}{dt}$$
Example 4 – system of 5 metabolites

\[
S \cdot \mathbf{v} = \frac{\mathrm{d} \mathbf{x}}{\mathrm{d} t}
\]

\[
\begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 1 & -1 & \\
0 & 0 & 0 & 1 & -1 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\dot{v}_1 \\
\dot{v}_2 \\
\dot{v}_3 \\
\dot{v}_4 \\
\dot{v}_5 \\
\dot{v}_6 \\
\end{pmatrix}
= 
\begin{pmatrix}
\dot{A} \\
\dot{B} \\
\dot{C} \\
\dot{D} \\
\dot{E} \\
\end{pmatrix}
\]
Example 4 – viewed from a row

\[ \frac{dC}{dt} = 0v_1 + 1v_2 - 1v_3 - 1v_4 + 1v_5 - 1v_6 \]

Fluxes that form C

Fluxes that degrade C
First constraint

• If we think about metabolism, we remember that it tends to maintain a balanced state.

• So a good assumption might be that under normal environmental conditions a constant flow through the system would be optimal.

• This implies that the relative concentration of metabolites would remain constant over time, which in fact is the central constraint for flux balance analysis.

\[ S v = 0 \]
Steady state

**Balance**

- A state in which no macroscopic events take place
- For every reaction the forward and reverse direction cancel each other out
- Analogy: „calm lake“

**Steady state**

- A state in which every variable is temporally constant
- For every substance generating and consuming reactions cancel each other out $\rightarrow$ such a system can „live“
- Analogy: „calm river“

\[ S_\nu = 0 \]
Convex analysis

- Equalities (in this case $Sv = 0$) and inequalities (in this case, $0 \leq v_i \leq v_{i,max}$) lead to convex analysis.
- It leads to the definition of a set of nonnegative generating vectors that span the solution space, the so called Extreme Pathways.
- All possible flux distributions of a metabolic network lie within the cone circumscribed by these pathways.
Elementary Flux Modes

Three Conditions for $S_{\text{EFM}}$

1. **Steady state** $S \cdot v = 0$
2. **Thermodynamic feasibility** $v_i \geq 0$ if reaction $i$ irreversible
3. **Non-decomposability** no nontrivial vector $e$ fulfilling the two conditions above and such that $P(e) \subset P(v)$

Where $P(e)$ denotes the set of reaction with non-zero flux in the mode $e$.

- A EFM is a set of enzymes that operate together at steady state and a mode is elementary when the removal of one enzyme causes to fail.
Extreme Pathways

Two additional Conditions for $S_{EP}$

4. **Network reconfiguration**: reactions must be classified either as exchange or internal. All reversible internal reactions are splitted into two separate, irreversible reactions with opposite direction. Internal flux can only admit non-negative value. Exchange reactions can be reversible.

5. **Systemic independance**: the set of EPs in a network is the minimal set of EFMs such that all feasible steady-state fluxes are non-negative linear combinations of these extremal pathways.
Extreme Pathways (Definition)

Comparison of network-based pathway analysis methods

Jason A. Papin¹, Joerg Stelling², Nathan D. Price¹, Steffen Klamt², Stefan Schuster³ and Bernhard O. Palsson¹

Extreme pathways
Extreme pathways (μ) are a set of convex basis vectors derived from the stoichiometric matrix [5]. They have the following properties:
(I) There is a unique set of extreme pathways for a given network.
(II) Each extreme pathway consists of the minimum number of reactions that it needs to exist as a functional unit.
(III) The extreme pathways are the systemically independent subset of elementary modes; that is, no extreme pathway can be represented as a nonnegative linear combination of any other extreme pathways.

Papin 2004 Comparison of network-based pathway analysis methods - PMID: 15283984
• EP is the **minimal set** of EFM in the sense that no element in EP is the convex combination of other elements.

• **EP = EFM** in the case where all **exchange reactions** (reactions connecting a metabolite with the outside of a metabolic system) are **irreversible** (e.g. glucose can only go into and ammonia can only go out of the system).
Examples (EP & EFM)
How to compute Extreme Pathways (1)

Recall the stoichiometric matrix

\[
N = \begin{bmatrix}
1 & -1 & 0 & 0 & -1 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & -1 \\
\end{bmatrix}
\]
How to compute Extreme Pathways (2)

Rewrite \( N^T \) by grouping internal reactions and external reactions.

\[
\begin{align*}
N^T &= \\
\begin{pmatrix}
A & B & C & D \\
v_2 & \begin{pmatrix}
-1 & 1 & 0 & 0 \\
0 & -1 & 1 & 0 \\
-1 & 0 & 0 & 1 \\
0 & 0 & 1 & -1 \\
1 & 0 & 0 & 0 \\
0 & 0 & -1 & 0
\end{pmatrix}
\end{align*}
\]
How to compute Extreme Pathways (3)

Use extended identity matrix, left panel record the history of reaction usage, right panel record the current net change of metabolic species.

\[
\begin{pmatrix}
  v_2 & v_3 & v_5 & v_6 & v_1 & v_4 & A & B & C & D \\
  v_2 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\
  v_3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
  v_5 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
  v_6 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\
  v_1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
  v_4 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 
\end{pmatrix}
\]
How to compute Extreme Pathways (4)

Identify metabolites that are not involved in external reactions. In this example, $B$ and $D$. Consider $B$ first, copy rows that correspond to internal reactions which do not involve $B$ (in this case $v_5$ and $v_6$). Combine reaction pairs which run opposite on $B$ (in this case $v_2$ and $v_3$). External reactions remain the same.

\[
\begin{pmatrix}
v_2 + v_3 & v_2 & v_3 & v_5 & v_6 & v_1 & v_4 & A & B & C & D \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0
\end{pmatrix}
\]
How to compute Extreme Pathways (5)

Repeat the same for $D$

$$
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0
\end{pmatrix}
$$
How to compute Extreme Pathways (6)

Use external reactions to balance internal net changes of all species remained. add $v_1$ onto $v_2 + v_3$ and $v_5 + v_6$ and remove $v_1$ (after balance $A$)

\[
\begin{align*}
&v_2 & v_3 & v_5 & v_6 & v_1 & v_4 & A & B & C & D \\
v_1 + v_2 + v_3 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
v_1 + v_5 + v_6 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
v_4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0
\end{align*}
\]

add $v_4$ onto $v_1 + v_2 + v_3$ and $v_1 + v_5 + v_6$ and remove $v_4$ (after balance $C$)

\[
\begin{align*}
&v_2 & v_3 & v_5 & v_6 & v_1 & v_4 & A & B & C & D \\
v_1 + v_2 + v_3 + v_4 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
v_1 + v_4 + v_5 + v_6 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0
\end{align*}
\]
How to compute Extreme Pathways (7)

- Now we balanced all the species (net change is zero) and the flux combination recorded on the left panel corresponds to the extremal pathways.

\[
EP_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}
\]
\[
EP_2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}
\]
Extreme Pathway - Research

Reaction Network

Stoichiometric Matrix

Pathway Matrix
Extreme Pathway - Research

- Extreme Pathways are the starting point for follow up calculations like Pathway Length or Reaction Participation

![Pathway Length Diagram]

Pathway Length

\[
P = \begin{pmatrix}
2 & 2 & 2 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}
\]

\[
\tilde{P} = \begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}
\]

\[
\tilde{P} \cdot P = \begin{pmatrix}
6 & 4 & 3 \\
5 & 6 & 3
\end{pmatrix}
\]

Comments:
1) The lengths of EP₁, EP₂, and EP₃ are 6, 6, and 7, respectively, the highlighted diagonal elements of the final matrix.
2) EP₁ and EP₂ have a shared length of 5 (indicated by the circle). As seen in the schematics above, they share reactions v₁, v₅, b₁, b₂, and b₅.

![Reaction Participation Diagram]

Reaction Participation

\[
P = \begin{pmatrix}
2 & 2 & 2 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}
\]

\[
\tilde{P} = \begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}
\]

\[
\tilde{P} \cdot P = \begin{pmatrix}
1 & 2 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\]

Comments:
1) The number of extreme pathways in which each reaction participates is indicated in the diagonal elements, as highlighted in the final matrix. These can then be expressed as a percentage of the total number of extreme pathways. For example, reaction v₂ has a participation value of 3. Since there are 3 extreme pathways, this can be expressed as 100% reaction participation.
2) The off diagonal terms can indicate correlated groups of reactions. Reactions v₂, b₂, b₃, and b₄ participate in 3 pathways. They also have a shared participation of 3, meaning they act as a correlated group (indicated by circles).
This Lecture

• Introduction
  – Systems biology & modelling
  – Metabolism & metabolic networks

• Network reconstruction
  – Strategy & workflow

• Mathematical representation
  – The stoichiometric matrix
  – Convex analysis & solution space
  – Elementary Modes & Extreme Pathways

• Constraint based network analysis
  – Flux Balance Analysis & Optimization
  – Tools
Optimization based methods (FBA)

- Identify an **objective function** $Z$ to be minimized or maximized depending on the task at hand
- Set **upper and lower bounds** on the flux variables $v_{i,\text{min}} \leq v_i \leq v_{i,\text{max}}$ for $i=\{1,\ldots,n\}$
- Solve **Linear Programming Problem** (i.e. Simplex Algorithm)
Possible Solutions for LP (visual)

\[ v_{i,\text{min}} \leq v_i \leq v_{i,\text{max}} \]

- empty
- unique
- degenerated
Minimization - Objective Functions

- Minimize ATP production: most energy-efficient state
- Minimize nutrient uptake: the fittest state under nutrient-shortage
- Minimize the Manhattan distance or Euclidian distance of the flux vector: minimize the overall flux
Maximization - Objective Functions

- Maximize metabolite production: the biochemical production capabilities of certain desirable metabolite such as lysine, phenylalanine etc.

- Maximize biomass formation: maximal growth rate

- Maximize biomass as well as metabolite production: the trade-off between cell-growth and forced metabolite production
Summary Flux Balance Analysis

Advantages

- For an objective function we get a 'fast' answer what the optimal flux distribution is.
- Successfully employed for several microorganisms.

Drawbacks

- Sensitivity to the definition of objective function.
- Optimal flux distribution might not be unique.
- Microorganisms could use different optimization criteria depending on their environment.
- The exploration of all suitable objective function is a difficult task.
FBA-Mode for synthesis of a VLDL-particle
Further Reading

- Llaneras 2008 Stoichiometric Modelling of Cell Metabolism PMID: 18295713

- Papin 2004 Comparison of network-based pathway analysis methods PMID: 15283984

- Harald Marx 2007 „Ausarbeitung Extreme Pathways“

- More kinetic analysis in cellular systems visit the lectures (Systemsbiology, Modelle zellulärer Prozesse) of Prof. Dr. Edda Klipp (Head of Theoretical Biophysics)