Advanced Computational Methods in Biotechnology

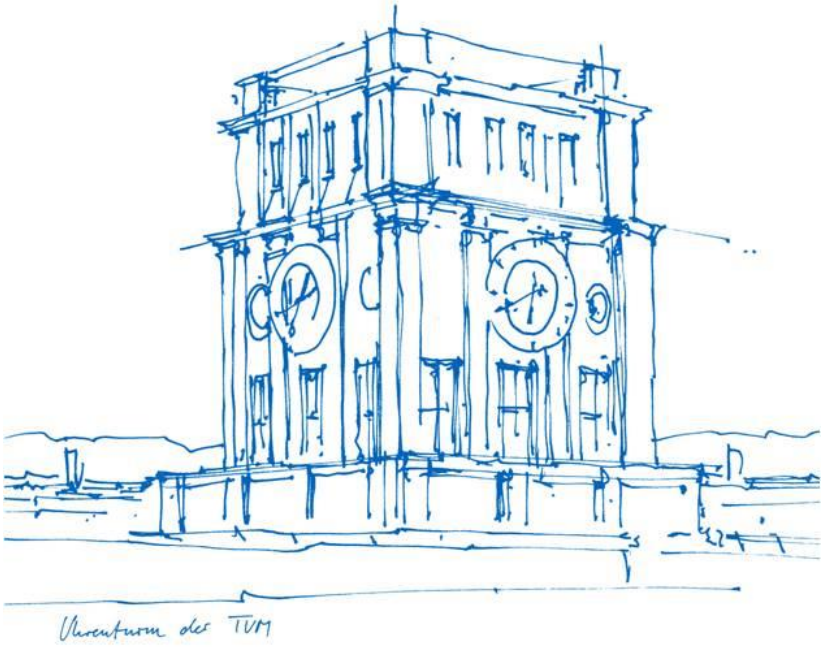
M. Sc. Marleen Beentjes

Professor Andreas Kremling

Technische Universität München

Professorship of Systems Biotechnology

Garching, 23th of June 2021



Course Evaluation

Online:
https://evasys.zv.tum.de/evasys/public/online/index

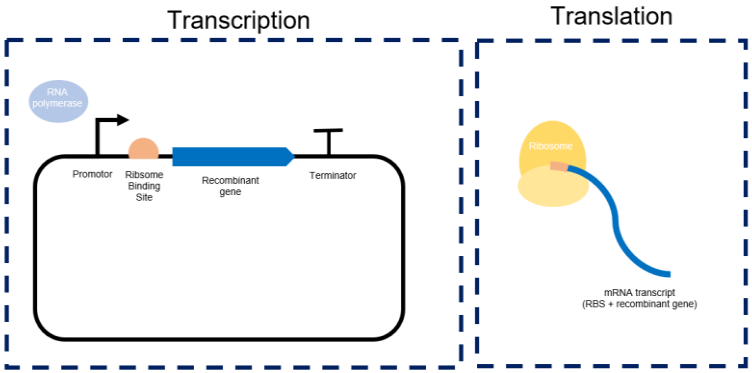
Code: K3MSL

Thank you in advance ☺

Scales of Cellular Processes

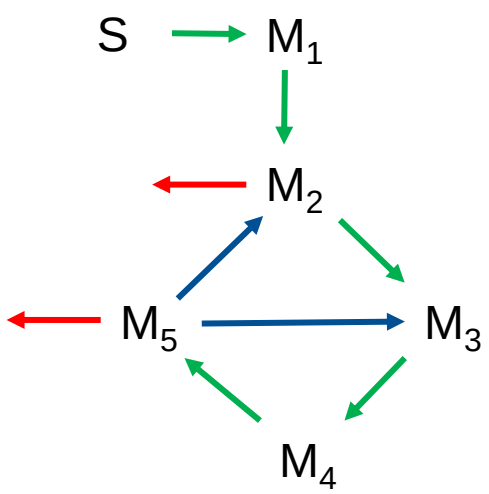
Topic 3

Signalling networks &
Gene Expression

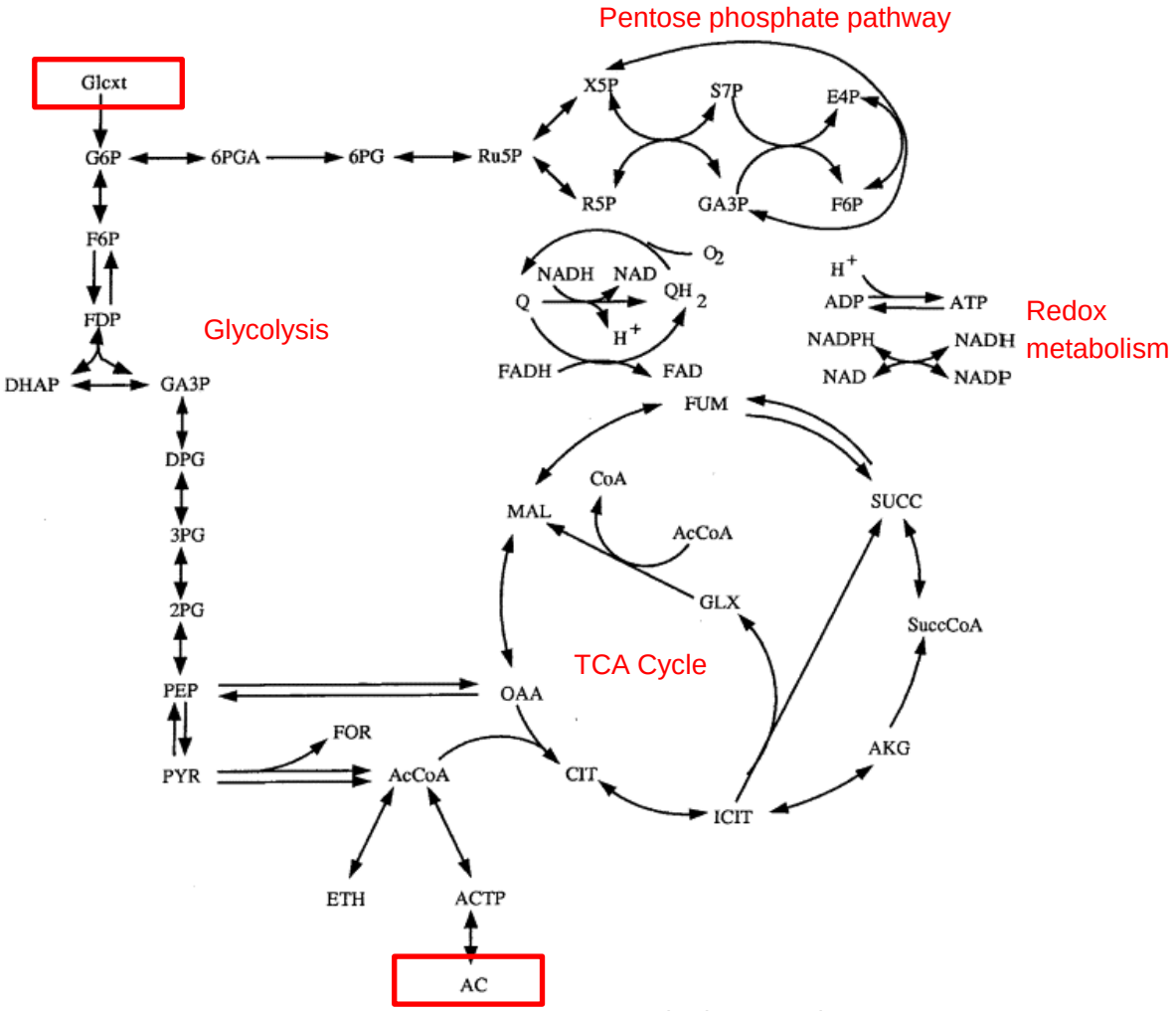


Proteins (including enzymes)

Metabolic Networks



Metabolic Networks



Mahadevan et al. 2002

Annotation:

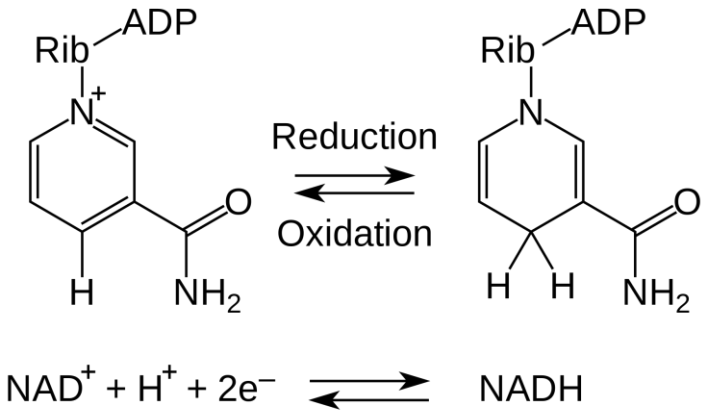
- Reactants
- Products
- Direction of reactions
- Co-factors
- Reducing equivalents
- Enzymes

Reducing equivalents

$RH_2 + NAD^+ \rightarrow NADH + H^+ + R$

From the hydride electron pair:

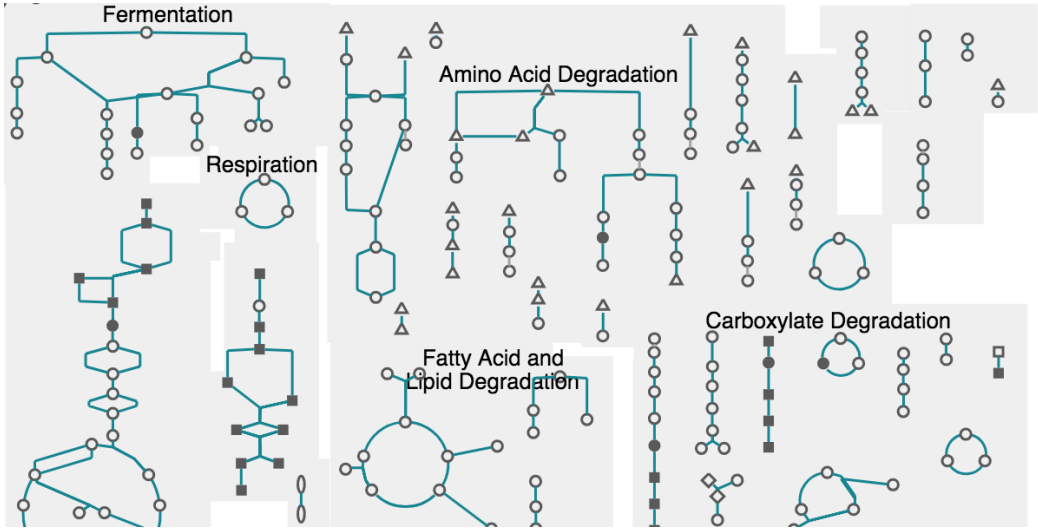
- one electron is transferred to the positively charged nitrogen of the nicotinamide ring of NAD^+
- the second hydrogen atom is transferred to the C4 carbon atom opposite this nitrogen
- In case study:
  - Not interested in H^+ as compound (not included in stoichiometric matrix)



How to access Metabolic Networks

Different options:

- KEGG
- MetaCyc
- BioCyc
- ENZYME
- BRENDA
- Reactome
- KaPPA-View4



- For our assignment: KEGG

KEGG

- Kyoto Encyclopedia of Genes and Genomes

- Different databases in four main categories
  - Systems information
  - Genomic information
  - Chemical information
  - Health information

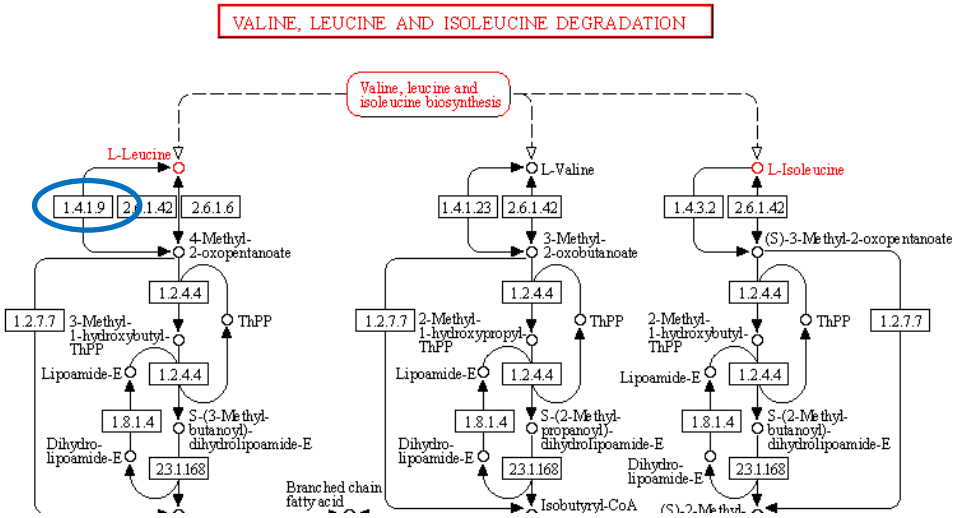
- KEGG PATHWAY - https://www.genome.jp/kegg/pathway.html

- Allows you to analyse specific (metabolic) pathways

- Allows you to find the occurrence of specific biochemical substances in different pathways

Example: L-Leucine

- Occurs in 34 different pathways in KEGG Database



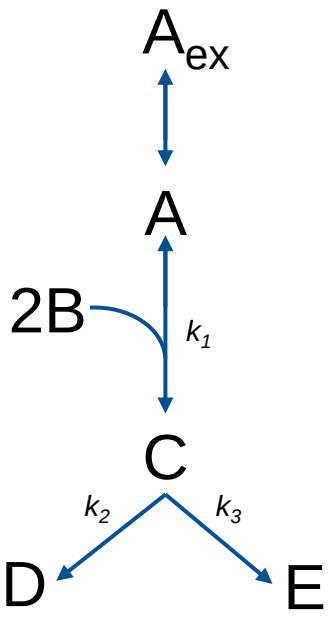
KEGG **ENZYME: 1.4.1.9** Help

Entry	EC 1.4.1.9 Enzyme
Name	leucine dehydrogenase; L-leucine dehydrogenase; L-leucine:NAD+ oxidoreductase, deaminating; LeuDH
Class	Oxidoreductases; Acting on the CH-NH2 group of donors; With NAD+ or NADP+ as acceptor BRITE hierarchy
Sysname	L-leucine:NAD+ oxidoreductase (deaminating)
Reaction(IUBMB)	L-leucine + H2O + NAD+ = 4-methyl-2-oxopentanoate + NH3 + NADH + H+ [RN:R01088]
Reaction(KEGG)	R01088; (other) R00145 R00146 R01434 R02196 Reaction
Substrate	L-leucine [CPD:C00123]; H2O [CPD:C00001]; NAD+ [CPD:C00003]
Product	4-methyl-2-oxopentanoate [CPD:C00233]; NH3 [CPD:C00014]; NADH [CPD:C00004]; H+ [CPD:C00080]

Computational analysis of Metabolic Networks

Mathematical description of metabolite dynamics:

$$\frac{dc_i}{dt} = production\ rate - consumption\ rate$$



Example Component C:

$$\frac{dc_C}{dt} = k_1 \cdot c_A \cdot {c_B}^2 - (k_2 + k_3) \cdot c_C$$

- Concentrations of intracellular components hard to determine
- Kinetic parameters often not known
- Great computational effort for complex networks

What is a (metabolic) flux?

Flux (various meanings)

- Physics:

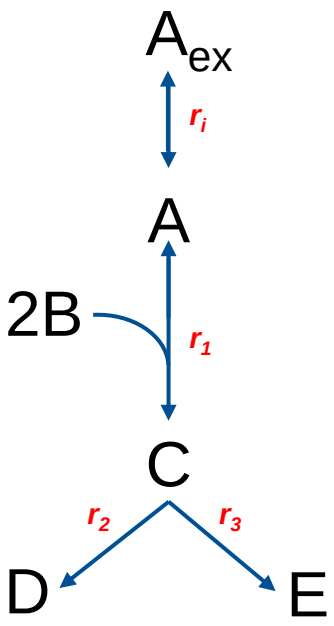
Flux describes any effect that appears to pass or travel (whether it actually moves or not) through a surface or substance

- Biology:

Flux, or metabolic flux is the rate of turnover of molecules through a metabolic pathway. Flux is regulated by the enzymes involved in a pathway

Flux Balance Analysis

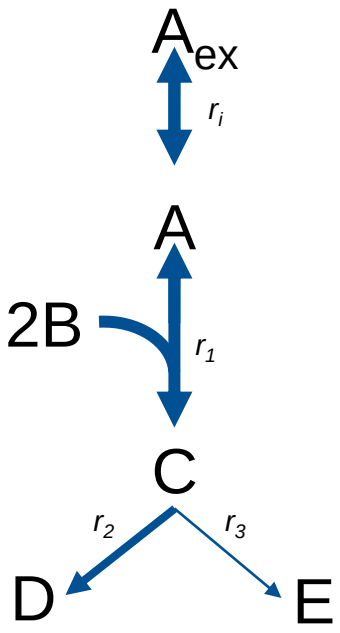
- Computational capacity is limited → simpler representation is preffered



Flux Balance Analysis

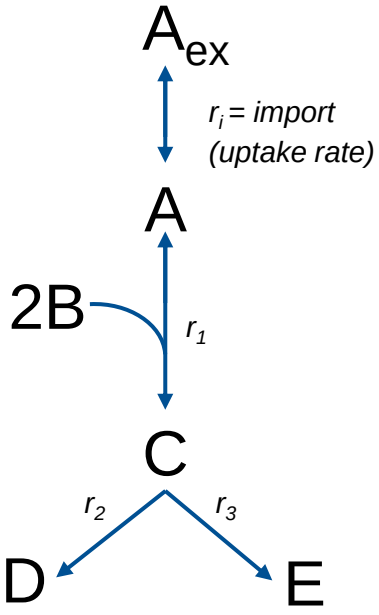
- Computational capacity is limited → simpler representation is preffered

- Flux Balance Analysis
  - First conceptualized in the 1980s
  - Simulation of flux distrubition in metabolic networks
  - Assumes all reactions in steady state (Constant over time)
  - Simulated using linear programming



General overview

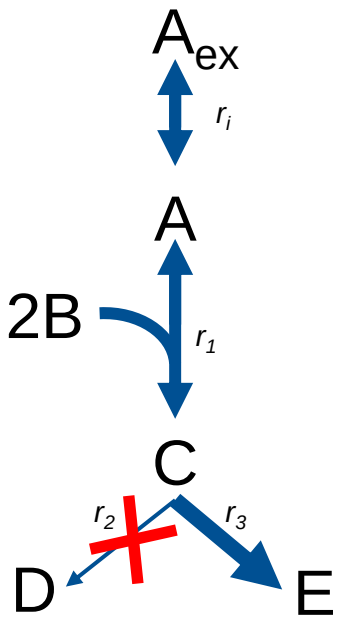
- Stoichiometric representation of network
- More unknown rates as known rates → Many solutions possible
- Finding a meaningfull solution → objective function
- Define an **objective function**
  - In the case of FBA, maximize or minimize a certain flux
- Perform optimization using linear programming



- We will come back to the mathematical and computational description of the problem later on

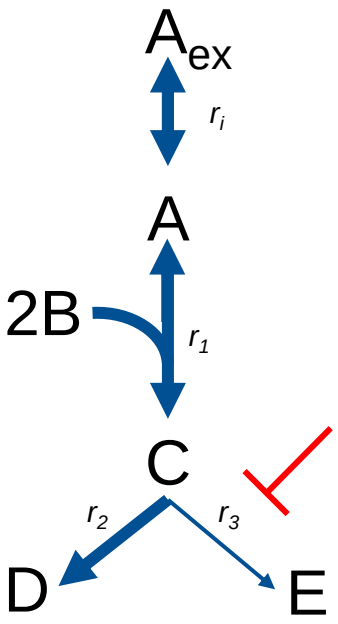
Why do we use it?

- Indicate targets for genetic engineering



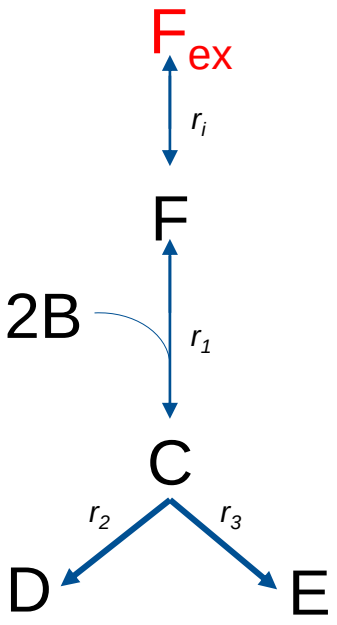
Why do we use it?

- Indicate targets for genetic engineering
- Investigate effect of reaction inhibitions



Why do we use it?

- Indicate targets for genetic engineering
- Investigate effect of reaction inhibitions
- Estimate the impact of different carbon sources



How to improve quality of analysis?

Add experimental data to the model

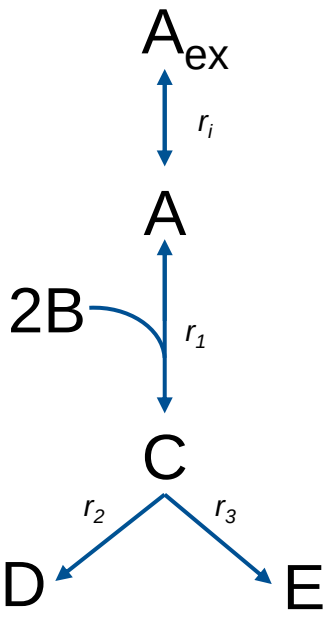
Constraints:

- Thermodynamic (reversiblity of reactions)

- Substrate uptake
  - Extracellular concentration
  - Diffusion / Active transport

- Measured fluxes

  → NMR or isotope labelling



How to improve quality of analysis?

Add experimental data to the model

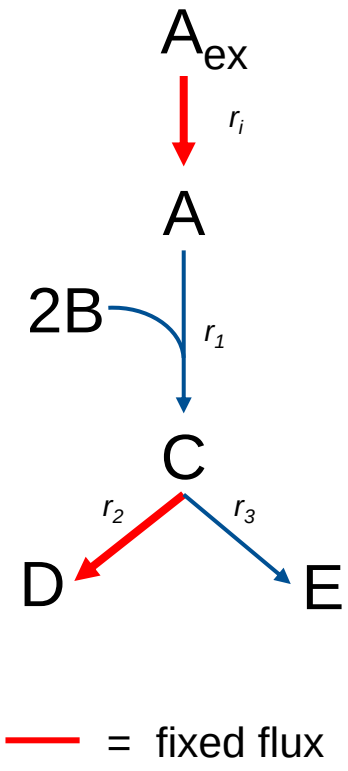
Constraints:

- Thermodynamic (reversiblity of reactions)

- Substrate uptake
  - Extracellular concentration
  - Diffusion / Active transport

- Measured fluxes

  → NMR or isotope labelling



Mathematical representation FBA

Stoichiometric representation of reactions:
(compare lectures on Reactor Problem)

Reaction equation:

$$|\gamma_A|A + |\gamma_B|B \rightarrow |\gamma_C|C + |\gamma_D|D$$

Mass balance equation:

$$0 = \gamma_C \cdot C + \gamma_D \cdot D + \gamma_A \cdot A + \gamma_B \cdot B$$

Vector/Matrix representation:

$$0 = \underline{n^T} \cdot \underline{K} \qquad \underline{n} = \begin{pmatrix} \gamma_A \\ \gamma_B \\ \gamma_C \\ \gamma_D \end{pmatrix} \qquad \underline{K} = \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix}$$

Mathematical representation FBA

FBA specific:

- $\underline{r}$: Vector containing all fluxes (reactions) in network
- N : Matrix containing stoichiometric coefficient of all reactions in network
  - Nr. columns = Nr. reactions
  - Nr. rows = Nr. metabolites

$$N = \begin{array}{c} \\ comp_1 \\ comp_2 \\ \dots \\ comp_N \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & \dots & r_M \end{array} \\ \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,N} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{N,1} & \gamma_{N,2} & \cdots & \gamma_{N,M} \end{pmatrix} \end{array}$$

- *Mass balance equation:*

$$\dot{c} = N \cdot \underline{r}$$

- **Assumptions:**
  - $0 = N \cdot \underline{r}$ **(steady state)**
  - Metabolite dilution negligible
  - Cell growth negligible (Proxy in form of ATP)

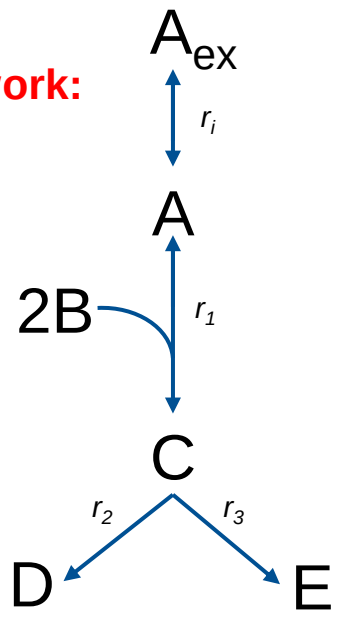
Mathematical representation FBA

Example:

- $\underline{r}$: Vector containing all fluxes (reactions) in network
- N : Matrix containing stoichiometric coefficient of all reactions in network
  - Nr. columns = Nr. reactions
  - Nr. rows = Nr. metabolites

$$N = \begin{array}{c} \\ comp_1 \\ comp_2 \\ \dots \\ comp_N \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & \dots & r_M \end{array} \\ \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,N} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{N,1} & \gamma_{N,2} & \cdots & \gamma_{N,M} \end{pmatrix} \end{array}$$

Construct the vector $\underline{r}$ and matrix N for the following example network:

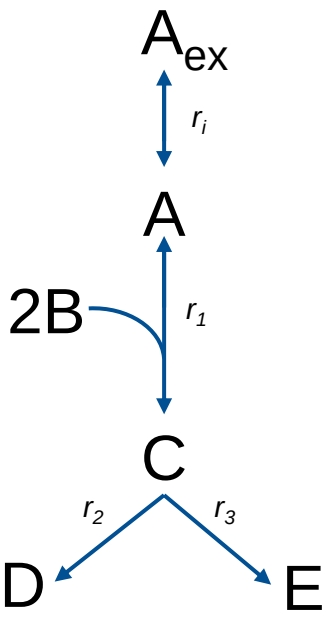


Mathematical representation FBA

Example solution:

$$r = \begin{pmatrix} r_i \\ r_1 \\ r_2 \\ r_3 \end{pmatrix}$$

$$N = \begin{array}{c} \begin{matrix} r_i & r_1 & r_2 & r_3 \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} A_{ex} \\ A \\ B \\ C \\ D \\ E \end{matrix}$$



Mathematical representation FBA

How can be FBA be validated experimentally?

Continuously stirred tank reactor (CSTR):

- $q_{in} = q_{out}$
- $\mu = (q_{in}/V)$ is constant
- System at steady-state

Optimization problem

Problem formulation

FBA aims to present the optimal distribution of all fluxes (*solution*) in the given metabolic network that are necessary to either maximize or minimize one or more specific fluxes (*objective function*)

$$\max\ c^T \cdot \underline{r}$$
$$\text{s.t. } N \cdot \underline{r} = 0$$
$$\underline{lb} < \underline{r} < \underline{ub}$$

The goal is to maximize the rates as indicated in the c^T vector subject to the steady state condition ($N \cdot \underline{r} = 0$) and the fixed boundaries ($\underline{lb} < \underline{r} < \underline{ub}$)

Objective function

What is most applied objective function in FBA?

- Biomass formation
- Atomic composition of biomass can be experimentally determined
- How is it represented?
  - Often lumped reactions
  - Sink for constituents of biomass (Fatty-acids, aminoacids, cell wall components)
  - In our exercise → Approximation (only ATP considered)

Table 7-4: Typical composition of bacteria cells

Constituent or element	Percent of dry weight
Major cellular material	
Protein	55.0
Polysaccharide	5.0
Lipid	9.1
DNA	3.1
RNA	20.5
Other (sugars, amino acids)	6.3
Inorganic ions	1.0
Sum:	100%
As cell elements	
Carbon	50.0
Oxygen	22.0
Nitrogen	12.0
Hydrogen	9.0
Phosphorus	2.0
Sulfur	1.0
Potassium	1.0
Sodium	1.0
Calcium	0.5
Magnesium	0.5
Chlorine	0.5
Iron	0.2
Other trace elements	0.3
Sum:	100%

Source: Metcalf & Eddy. Wastewater Engineering: Treatment and Reuse. 4th ed. Boston: McGraw-Hill, 2003.

Constraints

$$\underline{lb} < \underline{r} < \underline{ub}$$

$\underline{r}$: Set of fluxes that, multiplied by stoichiometric matrix *N*, should guarantee a steady state of the system

$\underline{lb}$: Lower bounds, indicate the **minimum** value for each flux. Size is equal to vector $\underline{r}$

$\underline{ub}$: Upper bounds, indicate the **maximum** value for each flux. Size is equal to vector $\underline{r}$

General: Fluxes with unknown value ±∞

More realistic:

- Reversible reactions [lb,ub]
- Irreversible reaction [0, ub] or [lb,0]

Computational implementation FBA

MATLAB function for solving linear programming problems:

x = linprog(f,A,b,Aeq,beq,lb,ub)

Description

Linear programming solver

Finds the minimum of a problem specified by

$$\min_x f^T x \text{ such that } \begin{cases} A \cdot x \leq b, \\ Aeq \cdot x = beq, \\ lb \leq x \leq ub. \end{cases}$$

f, x, b, beq, lb, and ub are vectors, and A and Aeq are matrices.

Output and inputs accepted by the linprog function:

f : Coefficient vector, specified as a real vector or real array. The coefficient vector represents the objective function f'*x.

A : Linear inequality constraints, specified as a real matrix. A is an M-by-N matrix, where M is the number of inequalities, and N is the number of variables (length of f).

b : Linear inequality constraints, specified as a real vector. b is an M-element vector related to the A matrix.

LB : Lower bounds, specified as a real vector or real array

UB : Upper bounds, specified as a real vector or real array

x: Solution, returned as a real vector or real array. The size of x is the same as the size of f.

Computational implementation FBA

FBA

$$\max c^T \cdot \underline{r}$$
$$\text{s.t. } N \cdot \underline{r} = 0$$
$$\underline{lb} < \underline{r} < \underline{ub}$$

Linprog

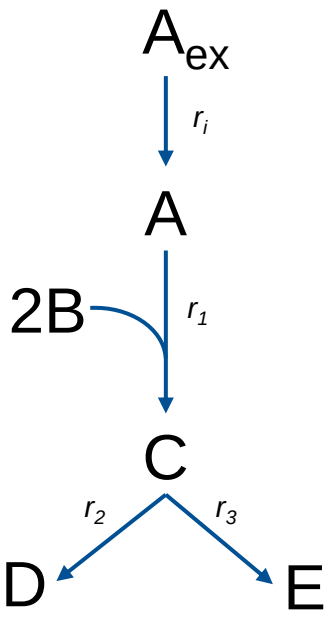
$$\min_x f^T x \text{ such that } \begin{cases} A \cdot x \leq b \text{ (inequality)} \\ Aeq \cdot x = beq, \\ \underline{lb} < \underline{x} < \underline{ub} \end{cases}$$

How to achieve maximization instead of minimization using linprog???

Output and inputs accepted by the linprog function (in FBA context):
f : Selects the fluxes for maximization/minimization (c^T ,same dimension as $\underline{r}$)
A : stoichiometric matrix (*N)*
b : vector to impose steady-state solution (contains only zeros, dimension is the height of *N*)
LB : lower boundaries for reactions
UB : upper boundaries for reactions
x: optimal flux distribution to achieve the max/min fluxes of f ($\underline{r}$)

Computational implementation FBA

Example



Stoichiometric matrix

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

$$f = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Objective: maximize r_3

$$b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

SS-condition

$$lb = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

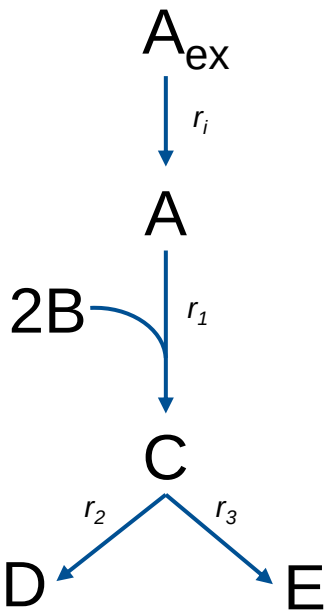
Lower boundries
(irreversible reactions)

$$ub = \begin{pmatrix} 1 \\ 100 \\ 100 \\ 100 \end{pmatrix}$$

Upper boundries
(fixed uptake rate)

Computational implementation FBA

Example



Stoichiometric matrix

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

$$f = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

Objective: minimize r_2

Objective: maximize r_3

$$b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

SS-condition

$$lb = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Lower boundries
(irreversible reactions)

$$ub = \begin{pmatrix} 1 \\ 100 \\ 100 \\ 100 \end{pmatrix}$$

Upper boundries
(fixed uptake rate)

Input for the linprog function

x = linprog(f,A,b,Aeq,beq,lb,ub)

We are interested in the flux distribution for steady state condition (Aeq, beq)

r_opt = linprog (f, [ ], [ ], N, zeros(height(N),1), lb, ub)

[ ] means no entry

Different input possible for linprog:

Known fluxes fixed by boundaries vs known fluxes in vector *b*

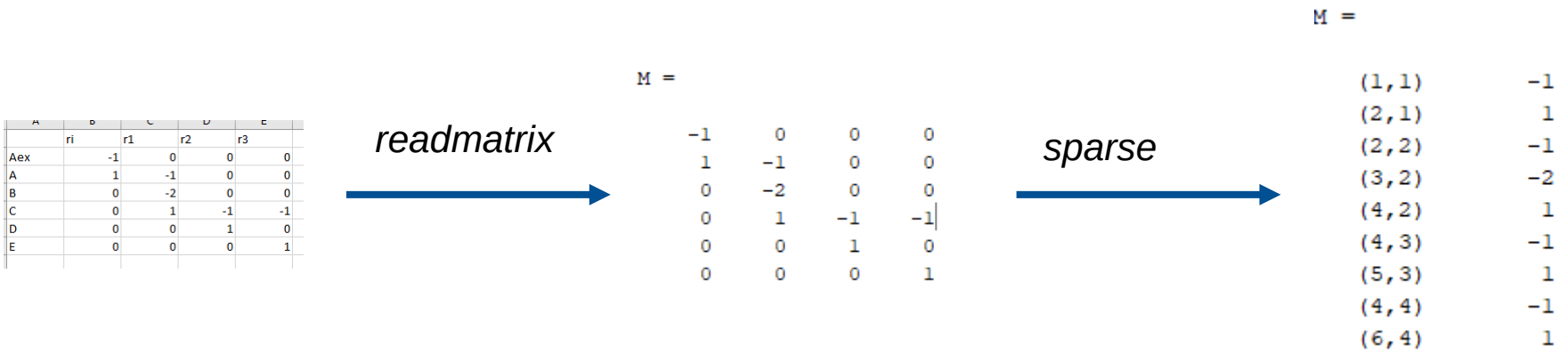
$N \cdot \underline{r} = 0$ *vs* $N' \cdot \underline{r'} = -\underline{r_{known}}$

N' should not include r_{known} in column

Saving space and memory in MATLAB

Working with large matrices:

- Construct large matrices in Excel
- Load them in a separate MATLAB file (e.g. *readmatrix*) and define matrix
  - NOTE: Excel file should be in same folder as MATLAB file
- Save Matrix as a .mat file (*save*)
- Load Matrix file (.mat) into working file
- Save memory by applying *sparse* function to matrix
- Some functions (like rank and null) will not work with sparsed matrices (Taks 1!)



Matrices on Moodle

- N_3 for task 3 (PHB synthesis)
- N_4 for task 4 (Glycerin Metabolism)
- N_6 for task 6 (Calvin-Cycle)

- In given matrices, r_{PHB} (r_{24}) represents a flux of 1 PHB and not ½ PHB!

- N_5 in task 5 (Alternative PHP synthesis) → Please include stoichiometric matrix file with indication of components and reactions!!!