Tutorial: Metabolic modelling of microbial communities

Metabolic network reconstruction and simulation with dynamic FBA

Steffen Waldherr, University of Vienna & Haris Zafeiropoulos, KU Leuven

Research topics

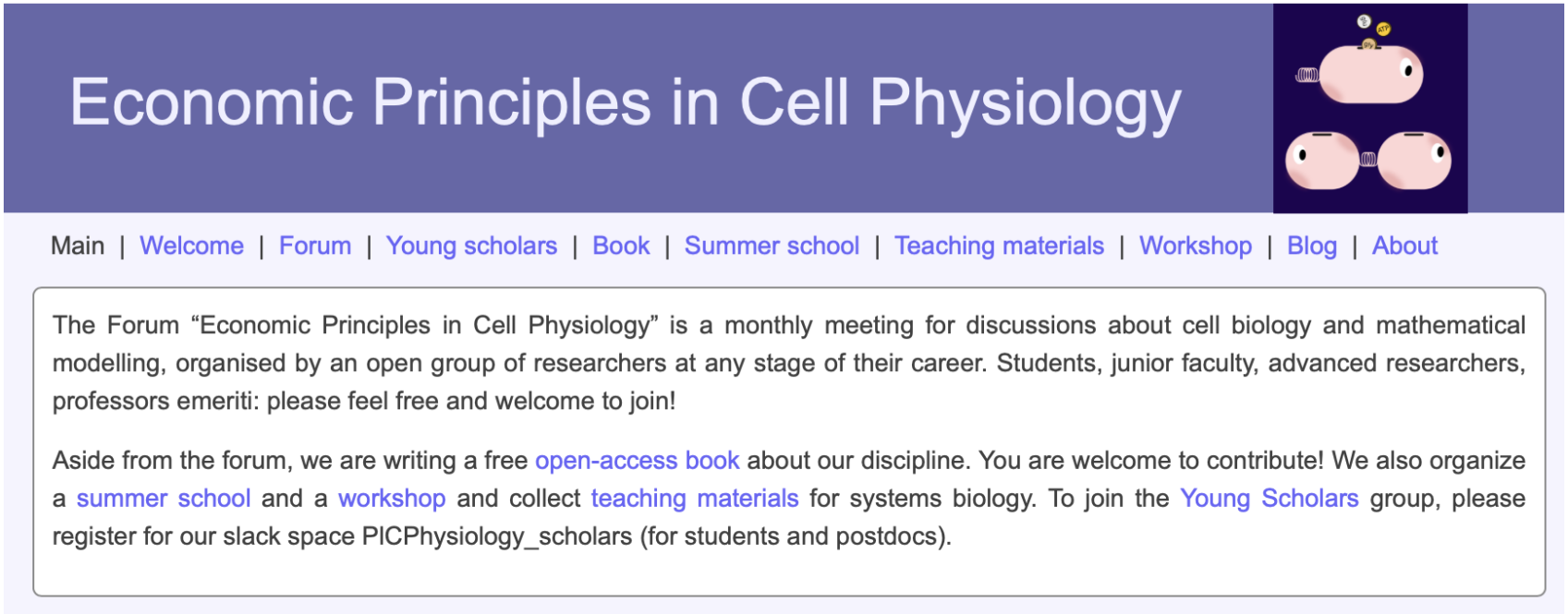
Research group “Computational methods for systems biology and biotechnology

Department of Functional and Evolutionary Ecology, Faculty of Life Sciences, University of Vienna

`https://ecology.univie.ac.at`

- Dynamic flux balance and resource allocation models
- Bioprocess optimization and control with constraint-based models
- Modelling and estimation of heterogeneous cell populations
- Machine learning for biochemical data and processes

(C) Bruno Klomfar

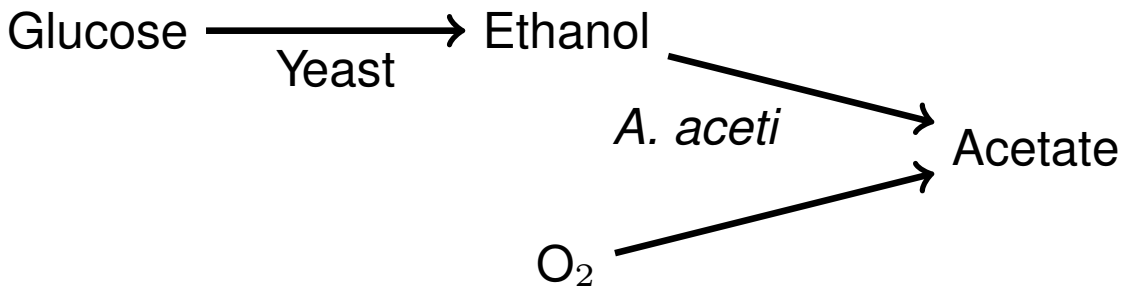


`https://principlescellphysiology.org`

- Subscribe to our email list (Website $\rightarrow$ Welcome $\rightarrow$ Where to find what)
- Join our monthly forum meeting: first Tuesday on a month
- Read our book “Economic principles in cell biology”, and especially the chapter “Optimal cell behavior in time”

Tutorial exercise: glucose → ethanol → acetate

- We try to model a community of yeast (*Saccharomyces cerevisiae*) and acetic acid bacteria (*Acetobacter aceti*) that should convert glucose into ethanol and ethanol into acetate, under conditions of limited oxygen transfer.



Metabolic network models

- For yeast, suitable metabolic network reconstructions for FBA and dynamic FBA are available.
- No appropriate FBA model for *A. aceti* ⇒ **metabolic network reconstruction**

Dynamic FBA simulations

- Instead of *A. aceti*, the ethanol metabolization will be done by an adapted *E. coli* model.

Overview

- Metabolic network reconstruction
  - Sequence homology search
  - Linking gene information to metabolic function
- Construction of dynamic FBA (DFBA) models
  - Constraints from enzyme kinetics / gas transfer in FBA
  - Coupling FBA part to dynamic part
  - Dynamic FBA for microbial communities

Network reconstruction: gene sequence to gene function

Aim

Analyse gene sequence to identify enzymes that are available for an organism

Starting point

(Annotated) genome sequence(s) of an organism that is to be modelled

- Full genome / chromosome sequence
- Sequences of individual gene products
  - Open reading frames (ORFs) in prokaryotes identified from start / stop codons
  - RNA sequencing (RNA-seq) for messenger RNA

Result

List of enzymes **possibly** expressed by an organism, with the genes coding for elements of each enzyme

Gene sequences

- Gene sequences are commonly available from databases:

By gene

- US National Center for Biotechnology Information (NCBI): `https://www.ncbi.nlm.nih.gov/gene`
- KEGG: `https://www.kegg.jp/kegg/genes.html`
- European Nucleotide Archive: `https://www.ebi.ac.uk/ena/`

By organism (genome)

- US National Center for Biotechnology Information (NCBI): `https://www.ncbi.nlm.nih.gov/genome`
- KEGG: `https://www.kegg.jp/kegg/genome/`
- European Bioinformatics Institute (EBI): `https://www.ensembl.org`, `https://ensemblgenomes.org`

- Can be downloaded in **FASTA** or **GenBank** format (both text-based)

  FASTA contains a 1-line description of each sequence (e.g. gene identifier, organism, proposed gene product / function, ...) followed by a single-letter-code for a nucleotide / amino acid sequence

  GenBank contains different types of information (references, descriptions, comments, gene locus annotations, **sequence**, . . .) in a basic text format

Sequence letter codes (IUPAC)

DNA

Code	Base
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

Protein

1-letter code	3-letter code	Amino acid
A	Ala	Alanine
B	Asx	Aspartic acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Any amino acid
Y	Tyr	Tyrosine
Z	Glx	Glutamine or Glutamic acid

Example FASTA file content

```
>ENA|MSL57368|MSL57368.1 Escherichia coli hexokinase
ATGGAGAAAAATATTTTTAAGTTAGATAATGAACAGCTCAAAGCAATAGCCCGTTCATTT
AAAGAAAAAGTAGAAAAAGGATTGAACACTGAAAACGCTGAAATCCAATGCATTCCTACC
TTTATTACTCCGAAGGCTGACAATATCAACGGTAAATCACTTGTGCTTGATCTCGGAGGA
ACCAATTATCGGGTAGCACTCGTTGATTTCAGCAAATCGGTACCGGACATTCATCCCAAC
AATGGTTGGAAGAAAGATATGTCGATCATGAAATCGCTGGGGTATACCCAAGAGGAATTA
TTTAAAGAGTTGGCAGATATGATCACCGGAATAAAACGGGAGGAGGAAATGCCTATCGGC
TATTGCTTTTCTTATCCGACCGAATCCGTGCCCGGCGGGGATGCAAAACTGCTGCGCTGG
ACAAAGGGAGTTGACATCAAAGAGATGATTGGAAAATATATCGGGAATCCCCTACTCAAC
TACCTGAATGAAAAGAATAAAATCAAGTTTACGGATATAAAAGTATTGAATGACACCGTA
GCCAGTTTATTTGCAGGACTTACAGACAACAGTTATGATGCATATATAGGACTGATTGTA
GGAACAGGCACTAATATGGCTACTTTCATCCCAGCCGACAAAATAAAAAAGCTGAATCCA
GCGGATAATATTCAAGGCATGATTCCCGTCAATTTGGAATCCGGGAATTTTCATCCGCCA
TTTCTTACCGGAGTGGATAATACAGTCGATGTAATTTCCGGTAACCCCAGAAAACAACGT
TTCGAGAAAGCAGTATCCGGTATGTATCTGGGAGATATTTTAAAAGCAACTTTTCCTTTA
GAAGAATTTGAAGAAAAATTTGATGCGCAAAAACTTACCGCTATCATGAACTATCCGGAT
ATATACAAGGATGTATATGTACAAGTGGCGCAATGGATATATACCAGATCGGCACAGTTG
GTGGCCGCTTCAATTACAGGGCTTGTCATGTTGTTGAAATCATACAATAAAGATATACGT
AGGATTTGCCTGGTGGCTGAAGGCAGTCTTTTCTGGAGTGAGAACAGAAAAGATAAGAAT
TATCACAGTATTGTTACAGAGGAATTAAAGGAGCTTTTCGACTTGTTCGGTTTGAAAGAT
GTTACAGTTGATATAAAAAGTATGAATAATGCGAATCTGATAGGTACAGCCATTGCGGCA
TTATCGTAA
```

Ways to determine gene function

Experimental

- direct assay / *in vitro*: isolate gene product and test in e.g. enzyme assay
- gene knockout studies $\rightarrow$ experimental analysis of metabolic pathways in mutant phenotype

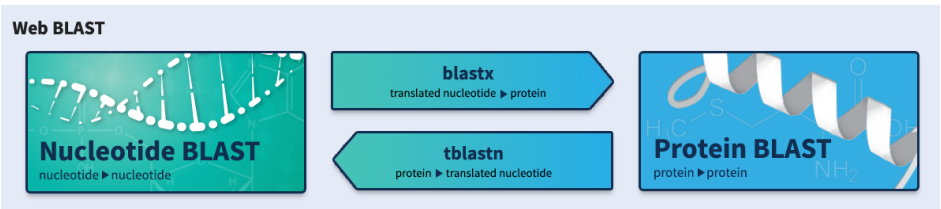
Computational

- **Sequence / structure similarity to a gene of known function, usually in another organism**
- Genomic context, e.g. presence in an operon with other genes of known function

Gene ontology evidence codes:
`geneontology.org/docs/guide-go-evidence-codes`

Homology search

- Sequence alignment of “query” gene to “database” of gene sequences with known function
- Computational tool: BLAST
  - Can work with nucleotide or amino acid sequences, or mixed (requires codon table)
  - Online operation: `blast.ncbi.nlm.nih.gov`



`www.ebi.ac.uk/Tools/sss/ncbiblast/nucleotide.html`

BLAST results and interpretation

- List of similar sequences in the database
- Each match is characterized by:
  - bit score: alignment of sequences based on a statistical model (higher is better)
  - E-value: expected number of sequence matches of similar score that can be found by chance in the database
    (good for metabolic reconstruction: $<\approx 10^{-50}$
  - sequence identity: Percentage of matching letters (less useful than other characteristics)

Finding gene function information

Useful functional information

- Name of enzyme encoded by gene product(s)
- EC number

Procedure

- Find suitable matching genes in databases (NCBI, ENA, KEGG)
- Look for annotation such as:
  - Gene / Protein name
  - EC number
  - References to other databases that are more protein/function-oriented:
    - UniProt: `www.uniprot.org`
    - InterPro: `www.ebi.ac.uk/interpro`
    - NCBI Protein: `www.ncbi.nlm.nih.gov/protein`
    - BRENDA: `www.brenda-enzymes.org`
- Follow-up with the found information to determine catalyzed metabolic reaction(s)

Hands-on part I

- Repository for tutorial materials (Jupyter notebooks):

  `https://github.com/swaldherr/dfba-microbedesign-leuven2023`

- Can be executed on Google Colaboratory:

  `https://colab.research.google.com`

Overview

- Metabolic network reconstruction
  - Sequence homology search
  - Linking gene information to metabolic function
- Construction of dynamic FBA (DFBA) models
  - Constraints from enzyme kinetics / gas transfer in FBA
  - Coupling FBA part to dynamic part
  - Dynamic FBA for microbial communities

General idea of DFBA

- Starting from a mass balancing model like the Monod model:

$$\frac{dX}{dt} = \mu(c)X$$
$$\frac{dc}{dt} = -\frac{\mu(c)}{Y_{X/c}}X$$

- replace the growth rate $\mu(c)$ by an “optimal” growth rate from FBA model
- replace the substrate / product rates by exchange fluxes from FBA model

Key steps / questions

- How do we set the reaction constraints (mostly transport capacity) based on the changing nutrient availability?
- Connect the FBA-based part (optimization problem) to the dynamic part (differential equation model)

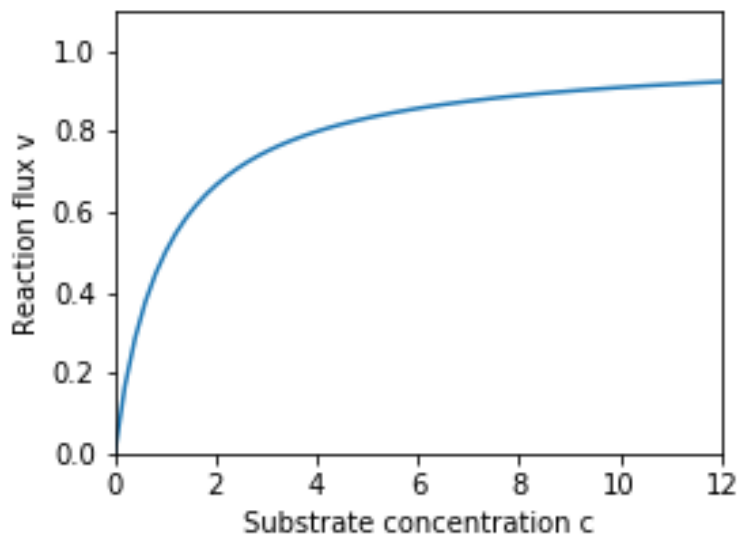
Enzyme kinetics

How do nutrient concentrations affect uptake rates?

- Constraints on enzymatic uptake fluxes (active transporters) are best represented by **enzyme kinetics**
- Michaelis-Menten kinetics:

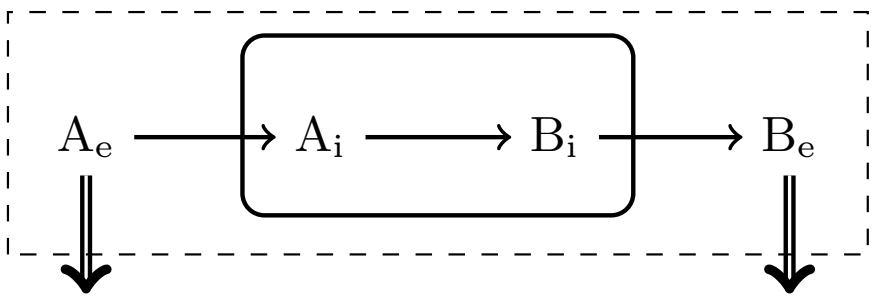
$$v = \frac{v_{max} c}{K_m + c}$$

  - Maximal reaction rate $v_{max} = k_{cat} E$, with k_{cat} the enzyme catalytic constant and E the enzyme concentration
  - Enzyme concentration E (relative to biomass) is assumed to be constant for DFBA.
  - Michaelis-Menten constant K_m: enzyme specific parameter



Exchange reactions

- Exchange reactions are added for all metabolites that are either **consumed** or **produced** in a metabolic steady state.
- They normally involve only **extracellular** metabolites.
- By convention, the reaction **direction** is towards the outside of the system



$\longrightarrow$ normal reaction

$\Longrightarrow$ exchange reaction

Positive vs. negative flux on exchange reaction

- Negative flux = actually goes into the system = **supply** (consumption) of a metabolite
- Positive flux = goes outside of system = **removal** (production) of a metabolite

Using enzyme kinetics to constrain FBA fluxes

- **Extracellular** concentration values are incorporated in DFBA model
- This concentration values can be used in enzyme kinetics to put **constraints** on the corresponding reactions in an FBA model.

Example:

FBA constraint

$$-10 \leq v \leq 0$$

DFBA constraint

$$-\frac{v_{max}c}{K_m + c} \leq v \leq 0$$

- More complicated kinetics (e.g. inhibition) can be modelled by corresponding changes in the constraint.

Gas transfer rates

- Gas transfer is usually limited by the transport from the gaseous phase to the liquid phase.
- Each gas chemical species has a **saturation concentration** c^* in the liquid which depends on the fraction of this gas in the gaseous phase.
  - Oxygen saturation concentration water / air: $c^* \approx 10 \frac{\mathrm{mg}}{\mathrm{L}}$ at 10 – 20 °C
- Once dissolved, common gases (oxygen) diffuse through the phospholipid membrane

Henry's law

Gas transfer is proportional to the difference between saturation concentration and dissolved concentration:

$$\frac{dc}{dt} = k_L a (c^* - c)$$

- Gas transfer constant $k_L a$ depends on geometry / setup of the growth environment (e.g. water-air surface ratio)

Integrating the DFBA model parts I

FBA part

$$\begin{aligned} &\max \mu \\ &\text{s.t. } Sv = 0 \\ &\quad v_{i,min}(c) \leq v_i \leq v_{i,max}(c) \end{aligned}$$

- Enzyme kinetics for bounds $v_{i,min}(c)$, $v_{i,max}(c)$: usually only a couple of (uptake) reactions

$$\Downarrow$$

Optimal solution

- Optimal growth rate μ^* in units of 1 / time
- Optimal exchange fluxes $v_{e,c}$ in units of molar amount / biomass / time

$$\Downarrow$$

Dynamic model equations

Integrating the DFBA model parts II

Optimal solution

- Optimal growth rate μ^* in units of 1 / time
- Optimal exchange fluxes $v_{e,c}$ in units of molar amount / biomass / time

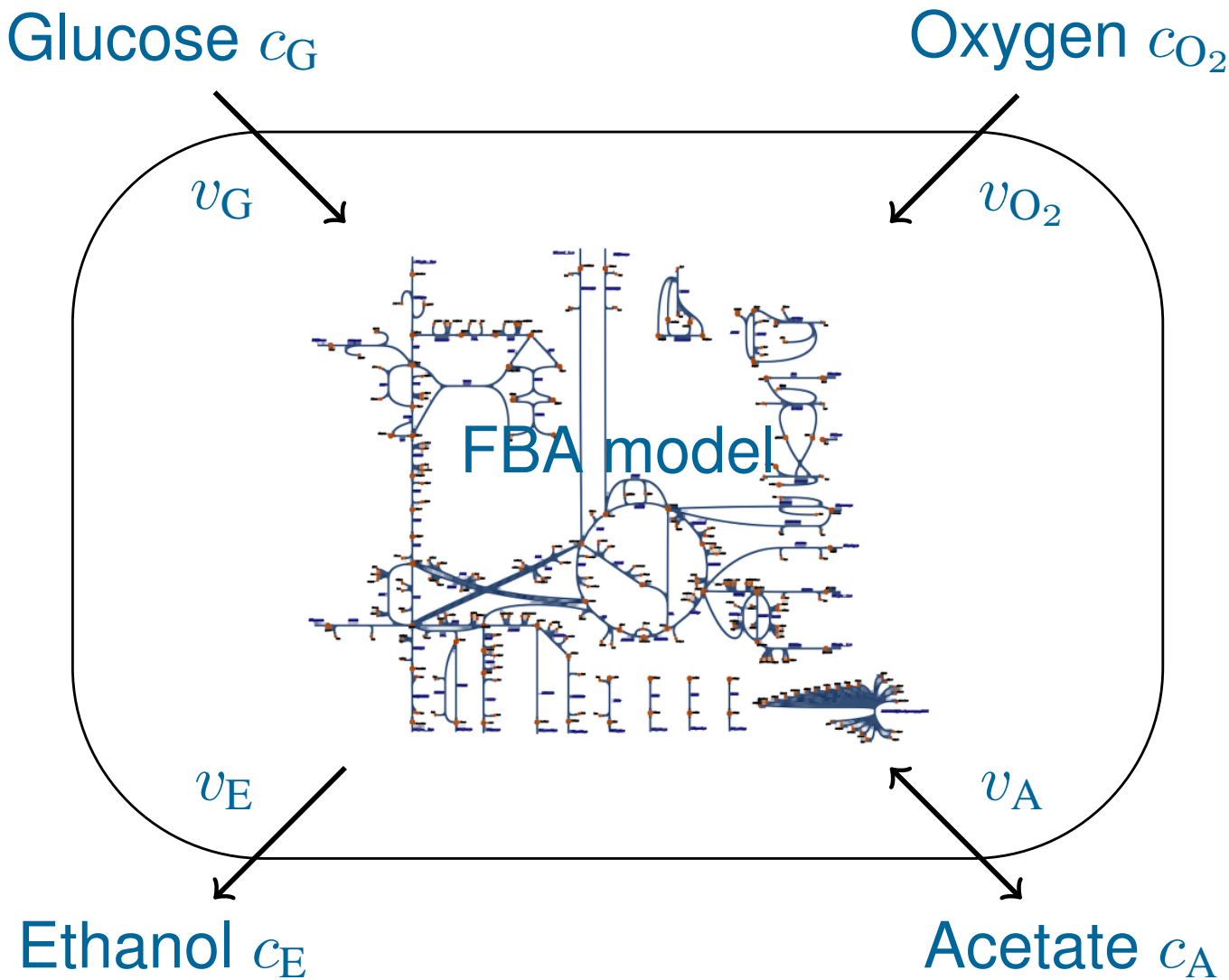
$$\Downarrow$$

Dynamic model equations

$$\frac{dX}{dt} = \mu^* X$$
$$\frac{dc}{dt} = v^*_{e,c} X$$

- Biomass X in units of mass or mass / growth environment volume
- Metabolite (concentration) c in molar amount or molar amount / growth environment volume
- Metabolites can also be considered in units of mass, requires multiplication with molar weight

Example based on *E. coli* core model



Book “Economic Principles in Cell Biology”, chapter “Optimal cell behavior in time”

Example based on *E. coli* core model

Exchange constraints

$$-10.5\,\frac{\text{mmol}}{\text{gDW h}}\,\frac{c_\text{G}}{2.7\,\frac{\text{mg}}{\text{L}}+c_\text{G}} \le v_\text{G} \le 0$$

$$-30\,\frac{\text{mmol}}{\text{gDW h}}\,\frac{c_{\text{O}_2}}{20\,\frac{\text{mg}}{\text{L}}+c_{\text{O}_2}} \le v_{\text{O}_2} \le 0$$

$$0 \le v_E$$

$$-30\,\frac{\text{mmol}}{\text{gDW h}}\,\frac{c_\text{A}}{100\,\frac{\text{mg}}{\text{L}}+c_\text{A}} \le v_\text{A}$$

Dynamic equations

$$\dot{X} = \mu X$$

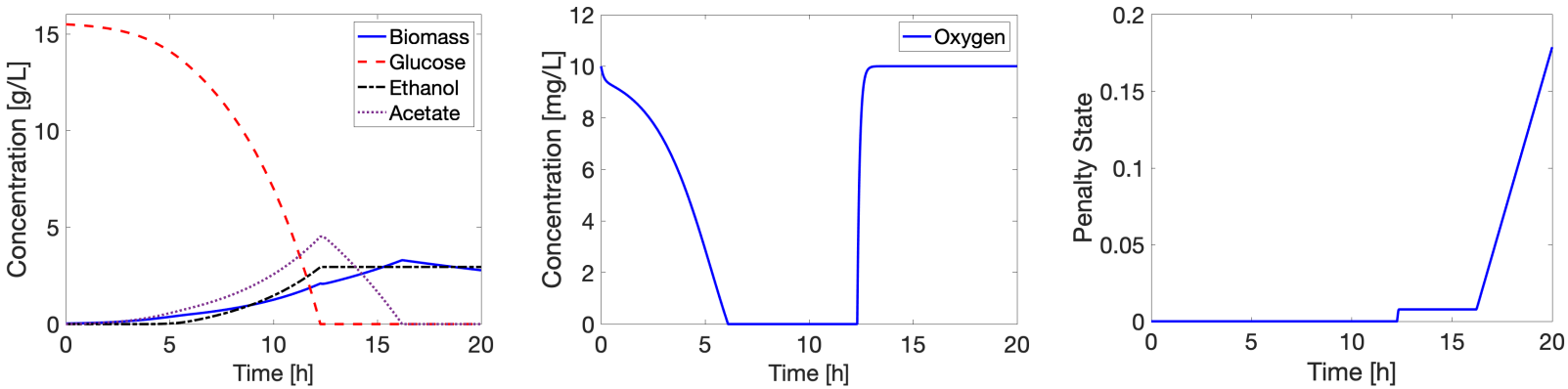
$$\dot{c}_{\text{O}_2} = v_{\text{O}_2} m_{\text{O}_2} X + k_L a(10\,\frac{\text{mg}}{\text{L}} - c_{\text{O}_2})$$

$$\dot{c}_\text{G} = v_\text{G} m_\text{G} X$$

$$\dot{c}_\text{E} = v_\text{E} m_\text{E} X$$

$$\dot{c}_\text{A} = v_\text{A} m_\text{A} X$$

Numerical solution for *E. coli* core example



Simulation with DFBAlab (`https://yoric.mit.edu/dfbalab`)

- Due to oxygen transfer limitations, an anaerobic period occurs during growth.
- Ethanol is only produced during that anaerobic period.
- Acetate is produced while glucose is available. After glucose depletion, oxygen levels recover and acetate that was previously secreted is re-metabolized.
- Penalty increases briefly during the glucose-acetate switch, and in stationary phase.

Exensions

- Microbial communities can be simulated by combining multiple FBA models with a single dynamic model
- Spatially distributed models (e.g. biofilms): replacing the (ordinary) differential equation with a partial differential equation (reaction-diffusion model)

Microbial communities with dynamic FBA

General approach

- Each community member is tracked by its biomass X_i.
- Exchange metabolites are modelled dynamically with their concentration in the extracellular medium.
- FBA models for community members are optimized independently based on current nutrient availability.
- Rate of change for extracellular metabolites is the sum of exchange fluxes from all community members.

Resulting dynamic model equations

$$\frac{dX_i}{dt} = \mu_i^* X_i$$
$$\frac{dc}{dt} = \sum_i v_{e_i,c}^* X_i$$

Caveats

- In this approach, each community member optimizes its own objective function, taking only its own metabolic network into account. If this is too unrealistic for a particular community (e.g. symbiotic), optimization of a combined network model will be necessary.
- The “naive” dynamic FBA simulation approach we use in the tutorial is quite inefficient. In actual modelling projects, better use dedicated DFBA solvers:
  - DFBAlab `https://yoric.mit.edu/dfbalab` (Matlab)
  - Dynamic-FBA `https://dynamic-fba.readthedocs.io` (Python)

Hands-on part II

- Repository for tutorial materials (Jupyter notebooks):

  `https://github.com/swaldherr/dfba-microbedesign-leuven2023`

- Can be executed on Google Colaboratory:

  `https://colab.research.google.com`

Conclusions

- Dynamic FBA: time courses of biomass and extracellular metabolite concentrations
- Constraints in FBA model are replaced by enzyme kinetics which depend on metabolite concentrations
- Advantages vs. Monod-type growth models:
  - Instead of empirical parameters on the whole-cell level, we have parameters on the molecular level
  - Complex nutrient transitions and multiple metabolic states can be described in a single model
- Can directly be used in “egoistic” type of community
  - Constant enzyme concentration / biomass composition is still assumed
  - Computationally challenging due to combination of optimization with differential equation solving