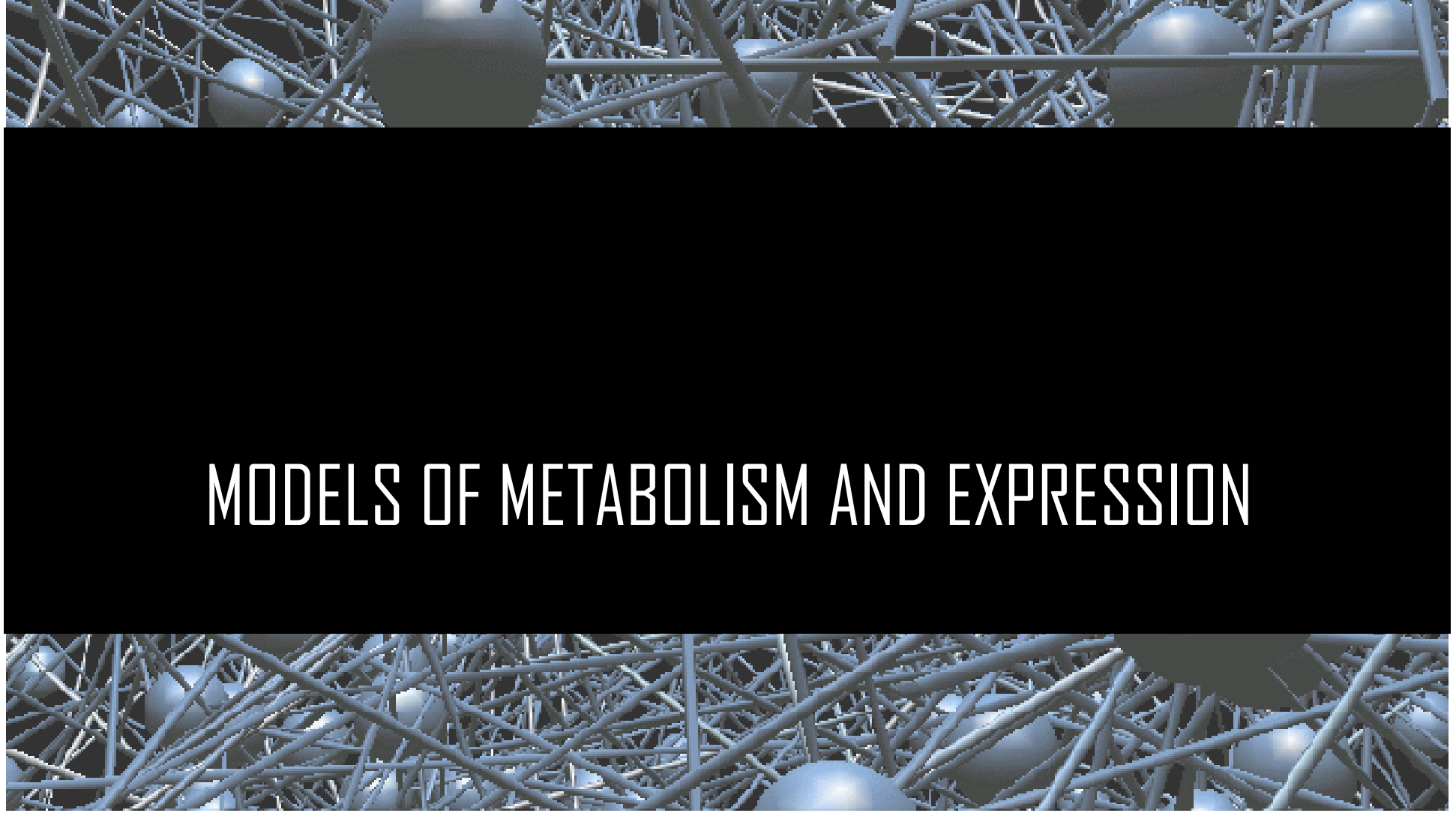


MODELS OF METABOLISM AND EXPRESSION

Principles and Applications of Systems Biology

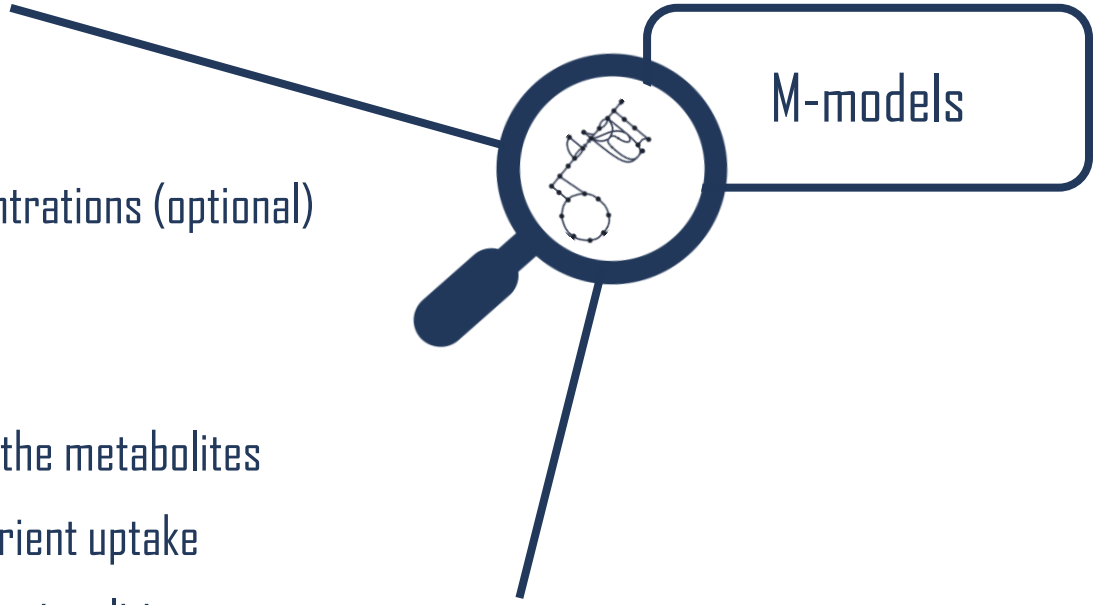
EPFL

Vassily Hatzimanikatis

November 2025

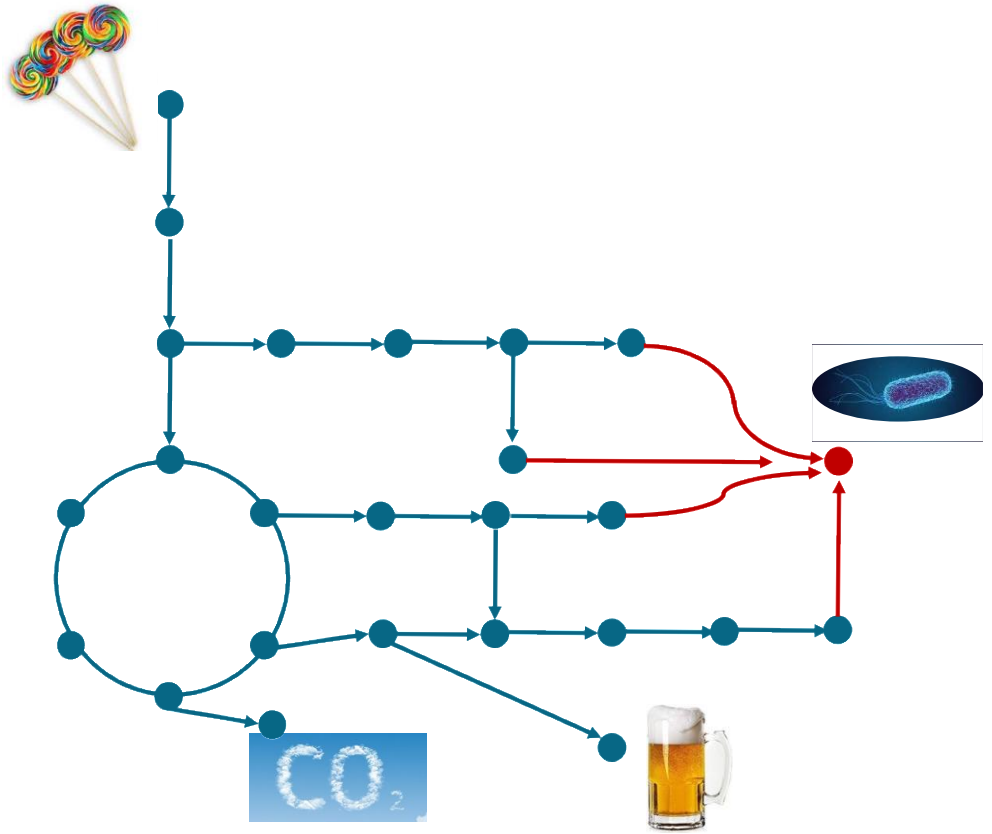
Variables

- Reaction fluxes
- Metabolite concentrations (optional)



Constraints

- Mass balance for the metabolites
- Limitations of nutrient uptake
- Pre-assigned directionalities



max 0.53 Ala + 0.18 Arg + 1e-6 Heme + … 55.3 ATP + 55.3 H2O -> biomass + 55.3 ADP + 55.3 H+ + 55.3 phosphate

Limitations in enzyme capacity

- The metabolic reactions are **mediated by enzymes or transporters**
- The catalytic capacity is limited:
  - **Abundance of the enzyme**
  - **Catalytic efficiency**
- In FBA, the relation between genes and reactions is **Boolean**
- The catalytic capacity is **unlimited**

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**
- Protein translation requires **energy**

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**
- Protein translation requires **energy**
- **mRNA transcripts** are needed to relay coding sequence from geneome to the translation machinery

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**
- Protein translation requires **energy**
- **mRNA transcripts** are needed to relay coding sequence from geneome to the translation machinery
- To synthesize RNA, a sequence of nucleotides must be assembled → **RNA transcription**

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**
- Protein translation requires **energy**

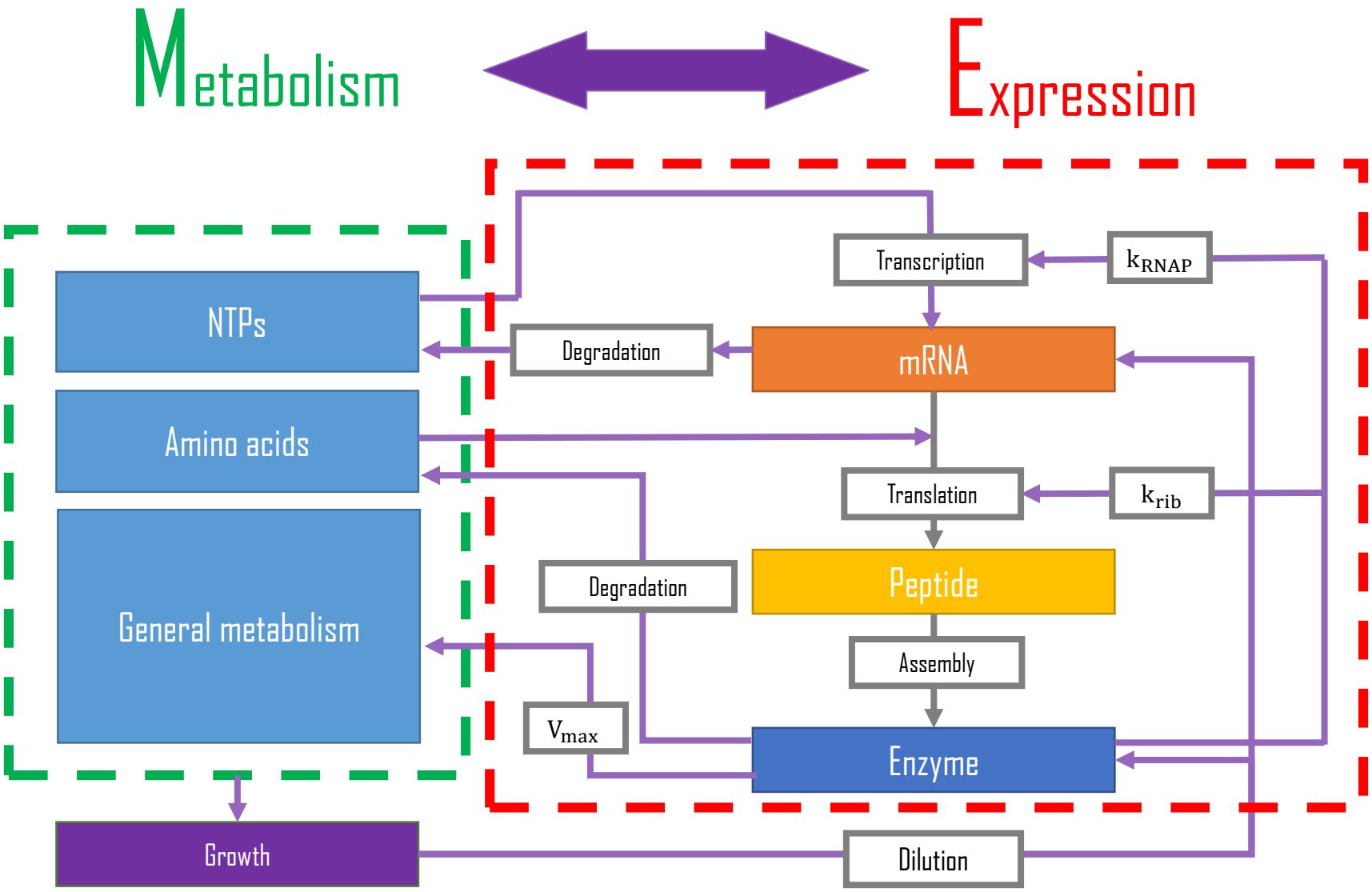
- **mRNA transcripts** are needed to relay coding sequence from genome to the translation machinery
- To synthesize RNA, a sequence of nucleotides must be assembled → **RNA transcription**
- Translation (**Ribosome**) and transcription (**RNA polymerase**) machinery are proteins

Metabolic requirements of proteins

- To synthesize proteins, a sequence of amino acids must be assembled → **Protein translation**
- Protein translation requires **energy**

- **mRNA transcripts** are needed to relay coding sequence from geneome to the translation machinery
- To synthesize RNA, a sequence of nucleotides must be assembled → **RNA transcription**
- Translation (**Ribosome**) and transcription (**RNA polymerase**) machinery are proteins
- In FBA, all these costs are modelled by biomass reaction:

0.53 Ala + 0.18 Arg + 0.03 CTP + … 55.3 ATP + 55.3 H2O -> **biomass** + 55.3 ADP + 55.3 H+ + 55.3 phosphate

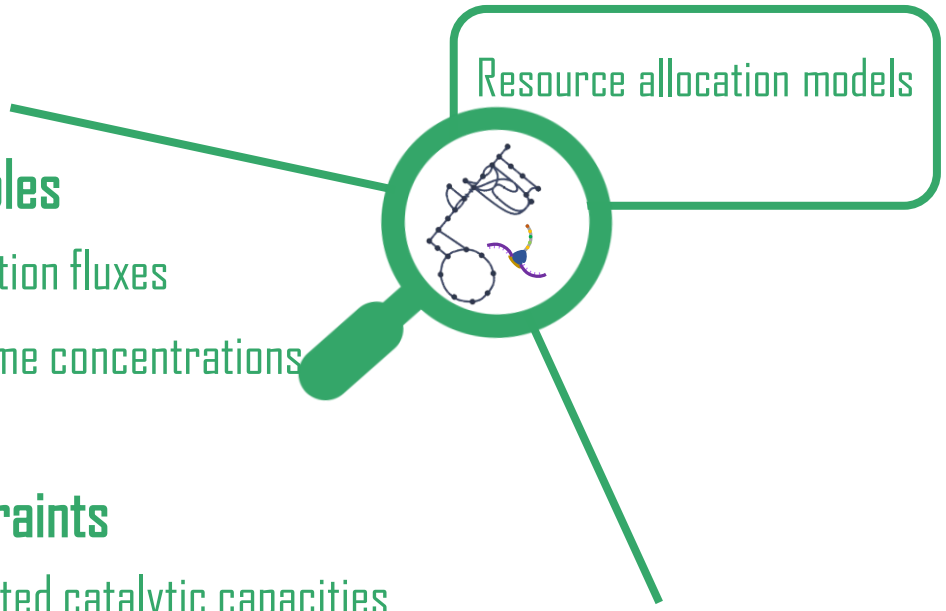

Metabolism
Expression
NTPs
Amino acids
General metabolism
Growth
Transcription
k_{RNAP}
Degradation
mRNA
Translation
k_{rib}
Degradation
Peptide
Assembly
V_{max}
Enzyme
Dilution

Variables

- Reaction fluxes
- Metabolite concentrations (optional)

Constraints

- Mass balance for the metabolites
- Limitations of nutrient uptake
- Pre-assigned directionalities



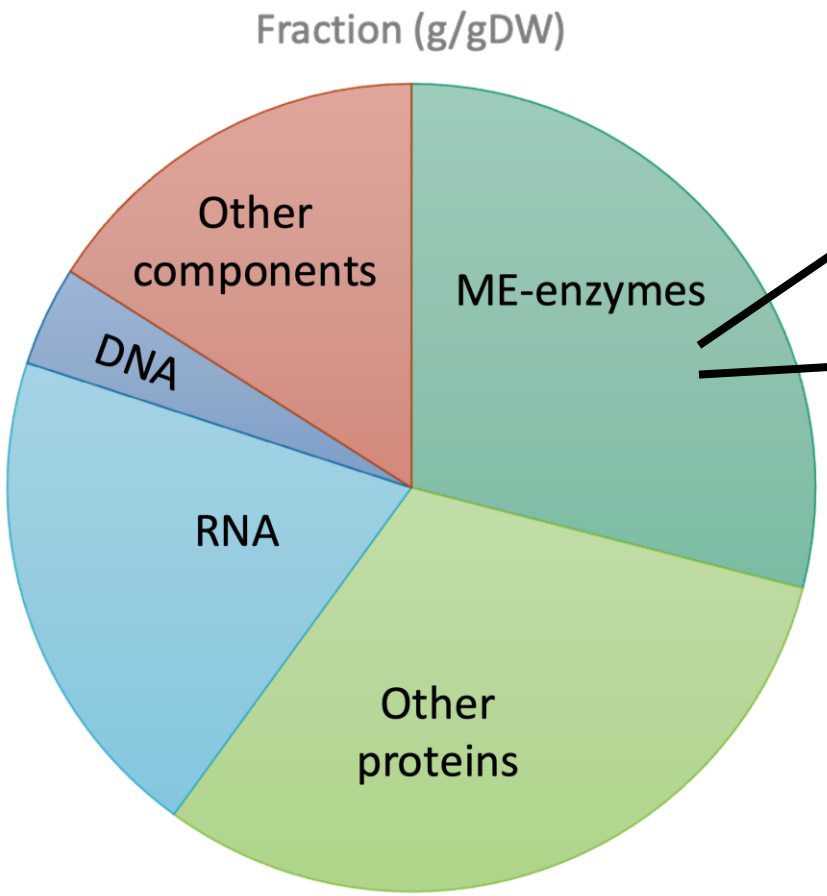
Variables

- Reaction fluxes
- Enzyme concentrations

Constraints

- Limited catalytic capacities
- Cost of protein synthesis
- Mass balance for the proteins (optional)
- Limited cellular capacities (optional)

Sánchez, Benjamín J., et al. "Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints." Molecular systems biology 13.8 (2017): 935.
Mori, Matteo, et al. "Constrained allocation flux balance analysis." PLoS computational biology 12.6 (2016): e1004913.
Goelzer, Anne, et al. "Quantitative prediction of genome-wide resource allocation in bacteria." Metabolic engineering 32 (2015): 232-243.



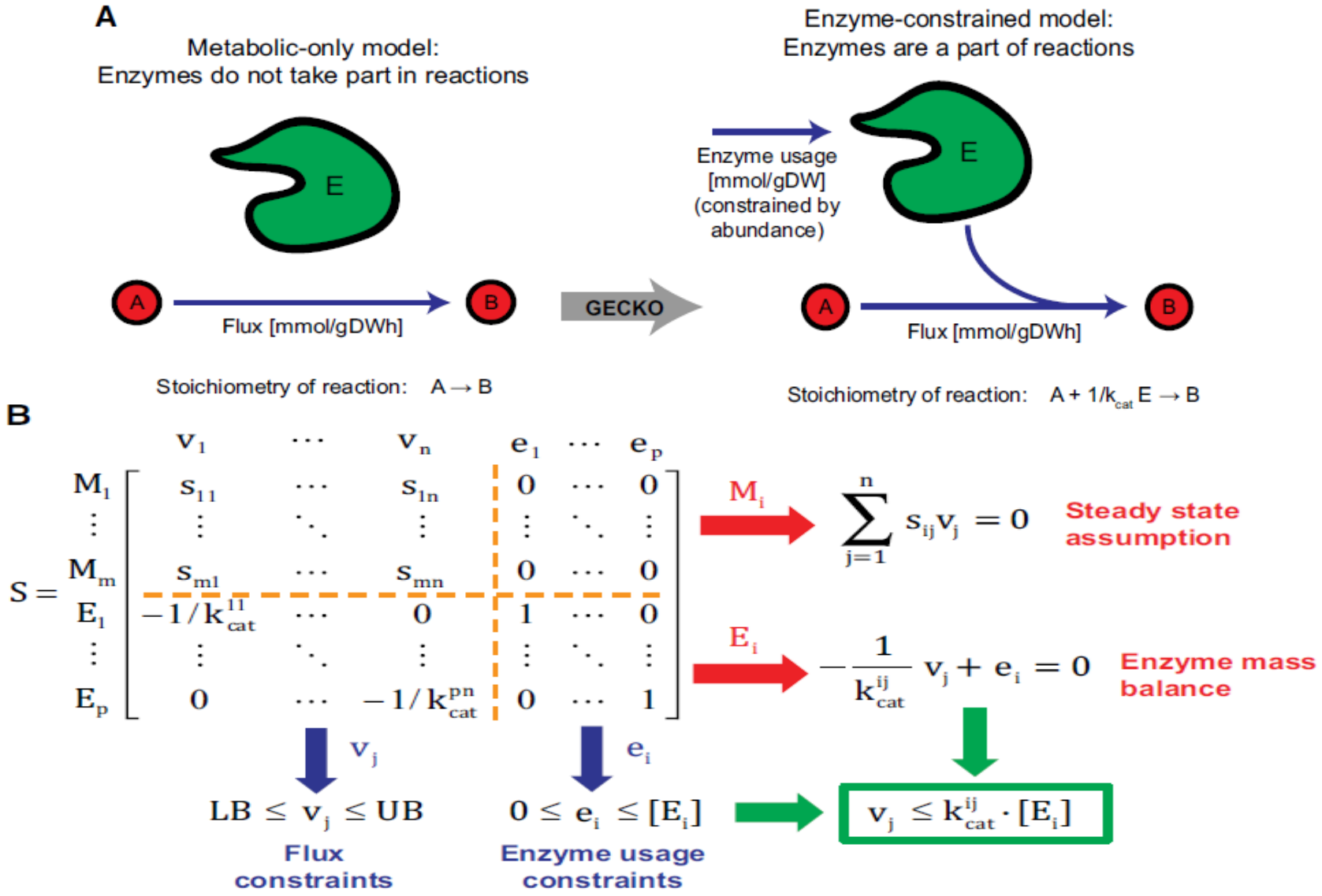
Enzyme 1 Enzyme 2 Enzyme 3

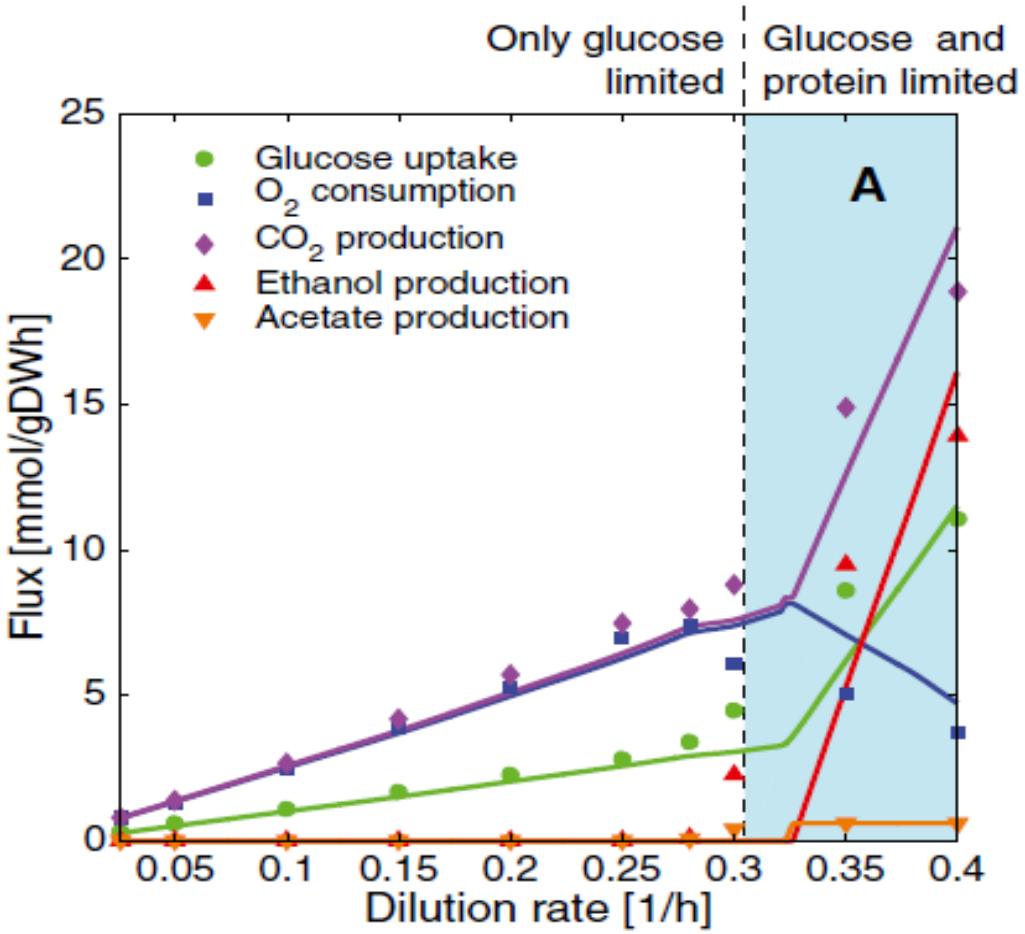
A resource allocation constraint defines the allocation of cellular resources to various enzymes

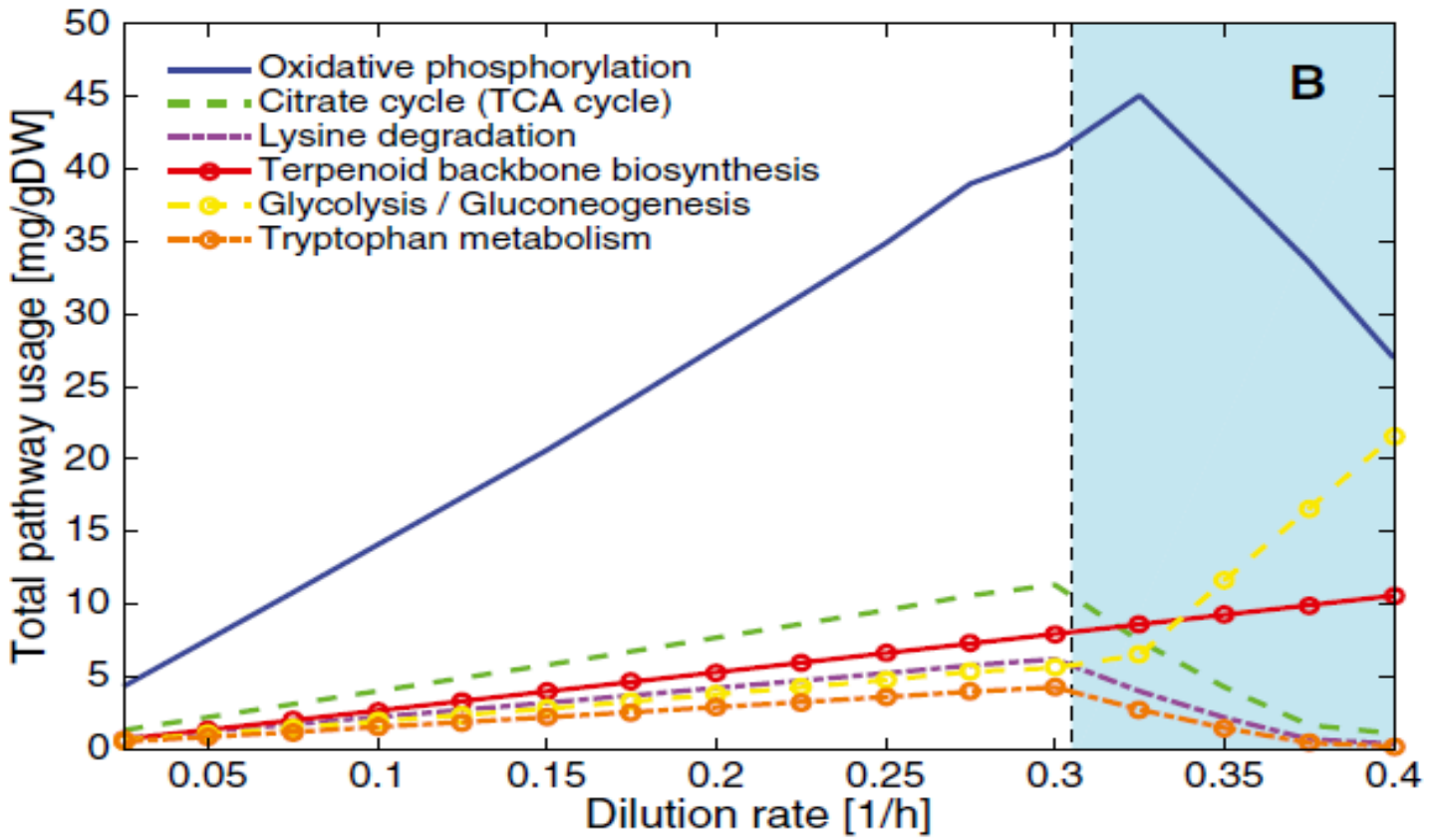
The enzymes are weighted based on their molecular weight

These methods are **phenomenological and coarse-grained**

GECKO
(Genome-scale model with Enzymatic Constraints using Kinetic and Omics data)
Concept

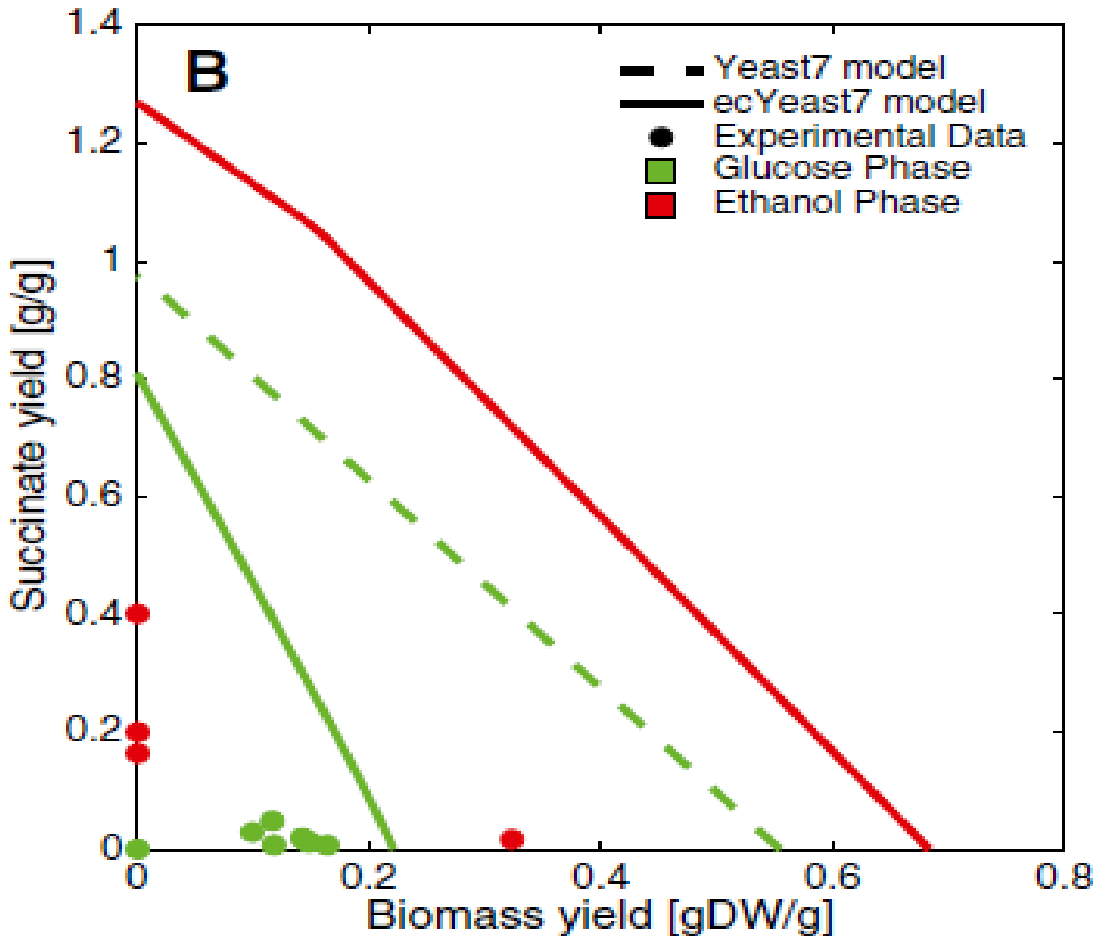




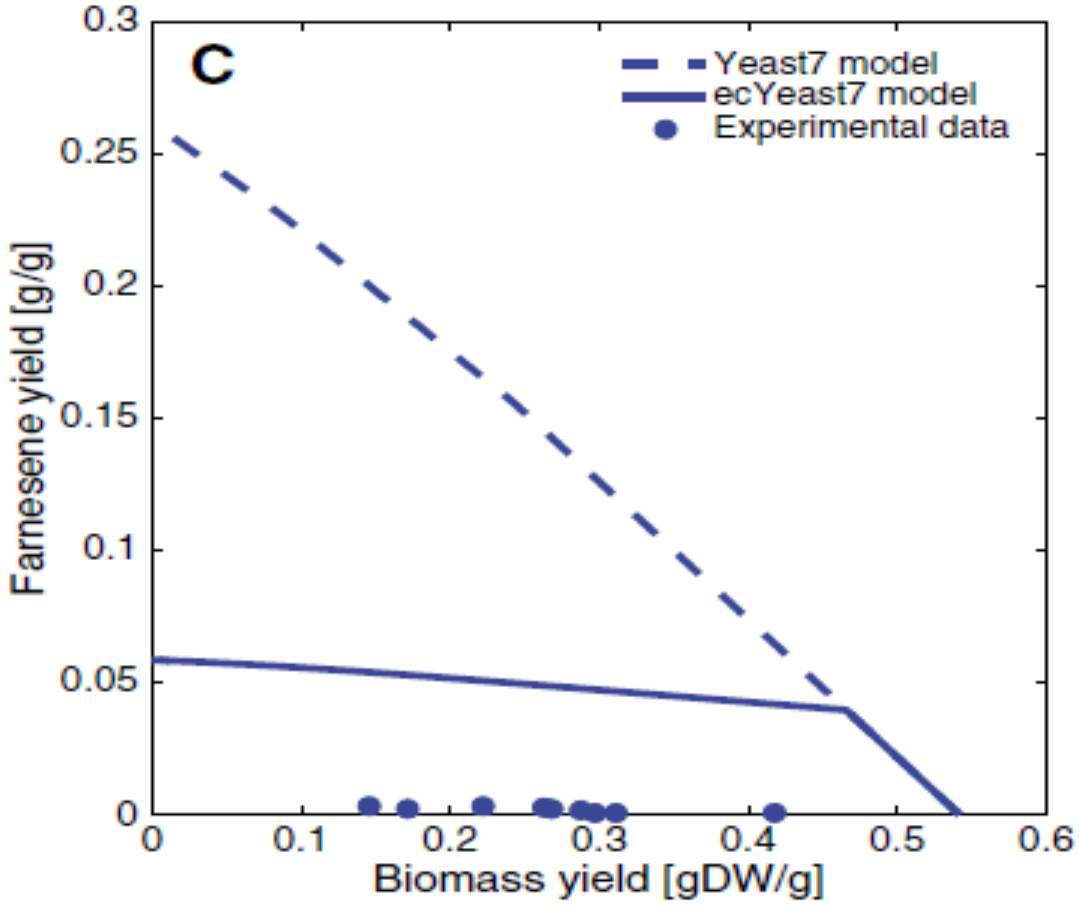


Crabtree effect: above 0.3 h^{-1} from respiration to respiration and fermentation

Results: Metabolic Engineering Application



Efficient Pathway
Classical Engineering Approaches



Inefficient Pathway
Protein Engineering Approaches

Variables

- Reaction fluxes
- Metabolite concentrations (optional)

Constraints

- Mass balance for the metabolites
- Limitations of nutrient uptake
- Pre-assigned directionalities

Aminoacids and nucleutides are not part of the lumped biomass reaction as the sythesis of individual enzymes and RNA transcripts is explicitly modeled

Constraints

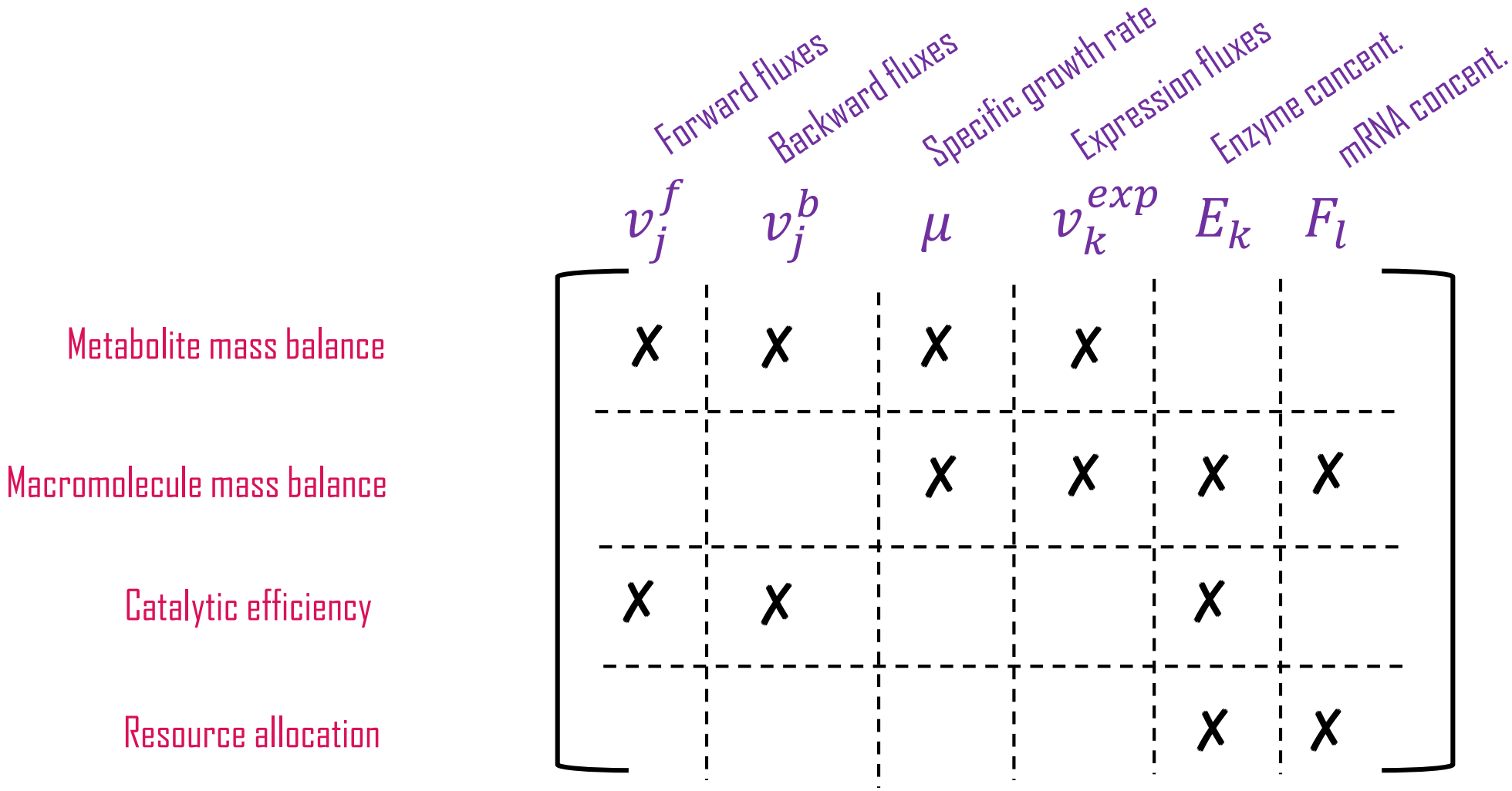
- Limited catalytic capacities
- Cost of protein and mRNA synthesis
- Mass balance for the macro-molecules
- Limited cellular capacities

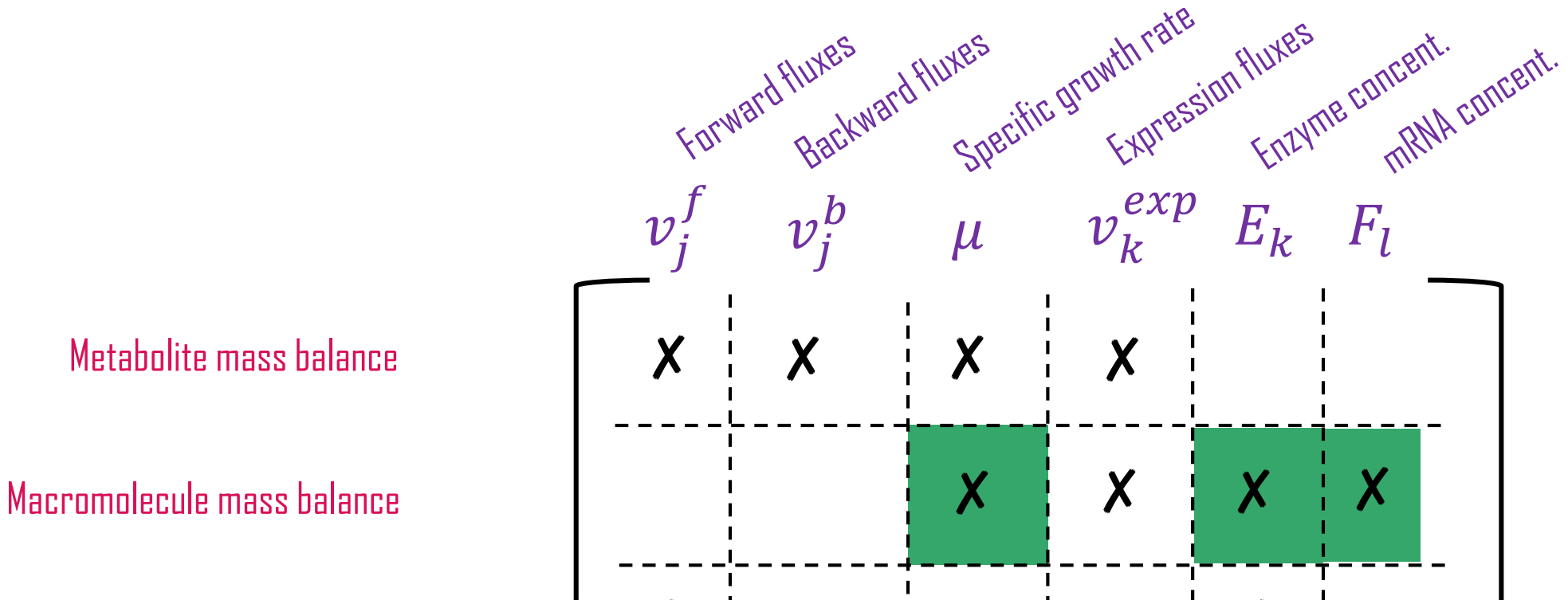
Variables

- Reaction fluxes
- Enzyme concentrations
- mRNA concentrations

Lerman, Joshua A., et al. "In silico method for modelling metabolism and gene product expression at genome scale." Nature communications 3.1 (2012)

Salvy, Pierre, and Vassily Hatzimanikatis. "The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models." Nature Communications 11.1 (2020)


Forward fluxes
Backward fluxes
Specific growth rate
Expression fluxes
Enzyme concent.
mRNA concent.
v_j^f
v_j^b
μ
v_k^{exp}
E_k
F_l
Metabolite mass balance
Macromolecule mass balance
Catalytic efficiency
Resource allocation



- Macromolecule mass balance:

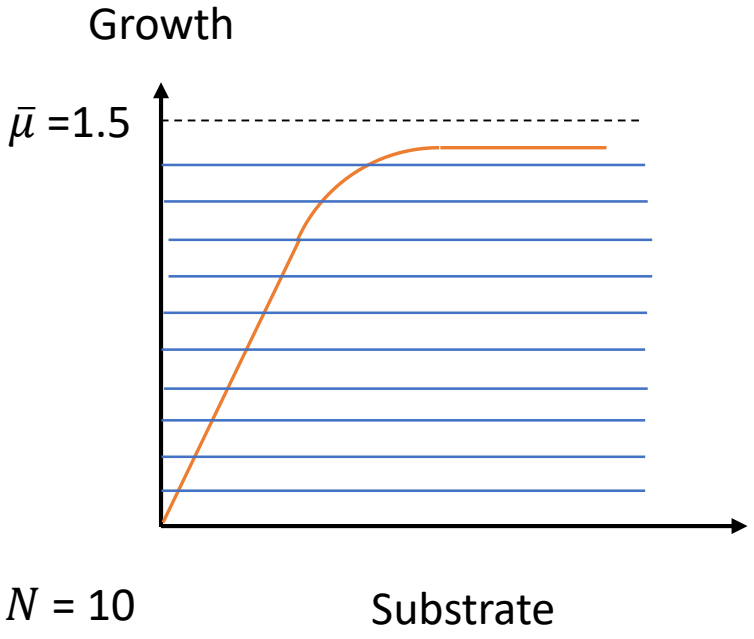
$$v_l^{syn} - v_l^{deg} - v_l^{dil} = 0 \qquad \forall l \in Macromolecules$$

$$v_l^{syn} - k_l^{deg} G_l - \mu G_l = 0$$

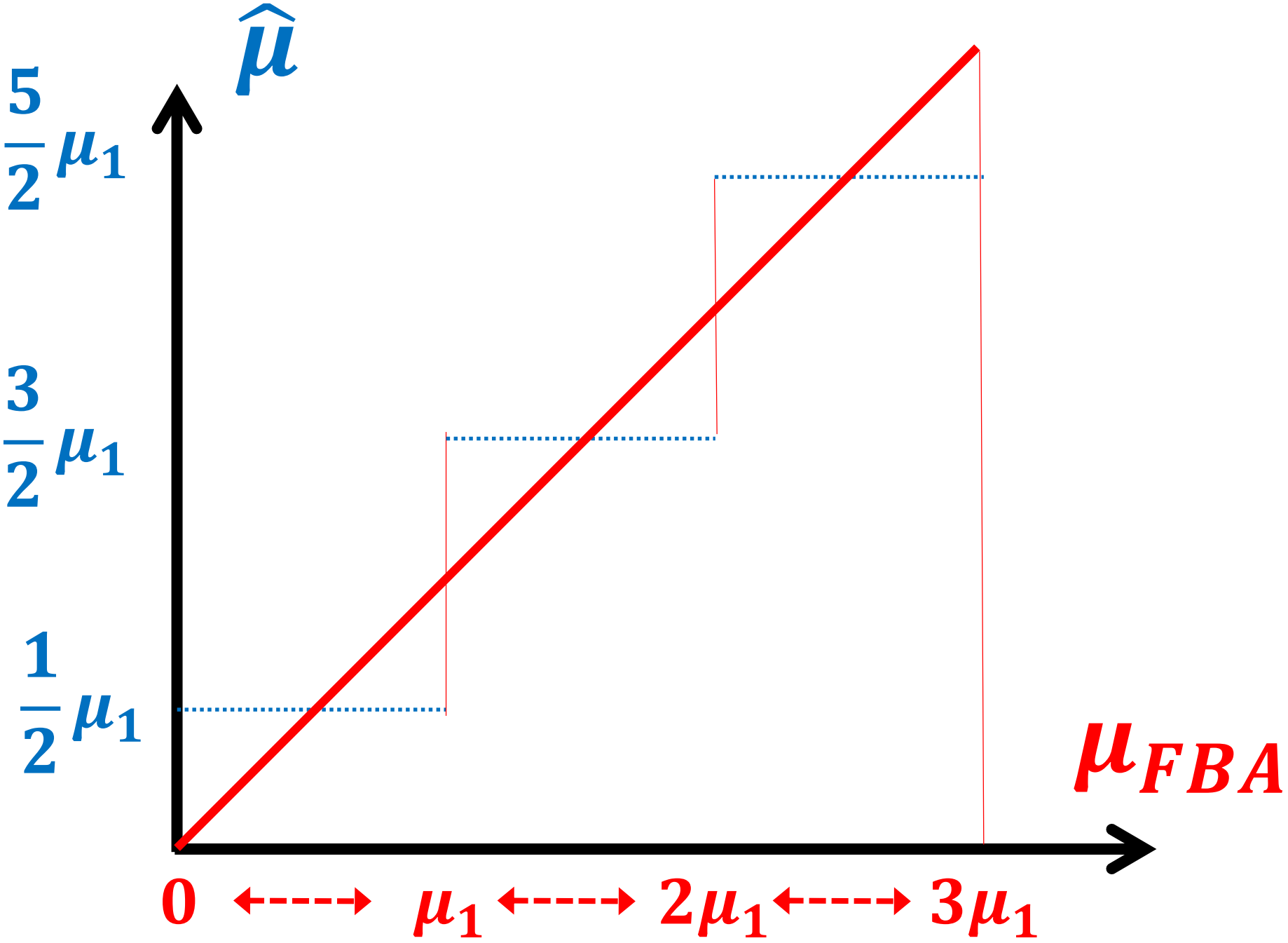
Nonlinearity

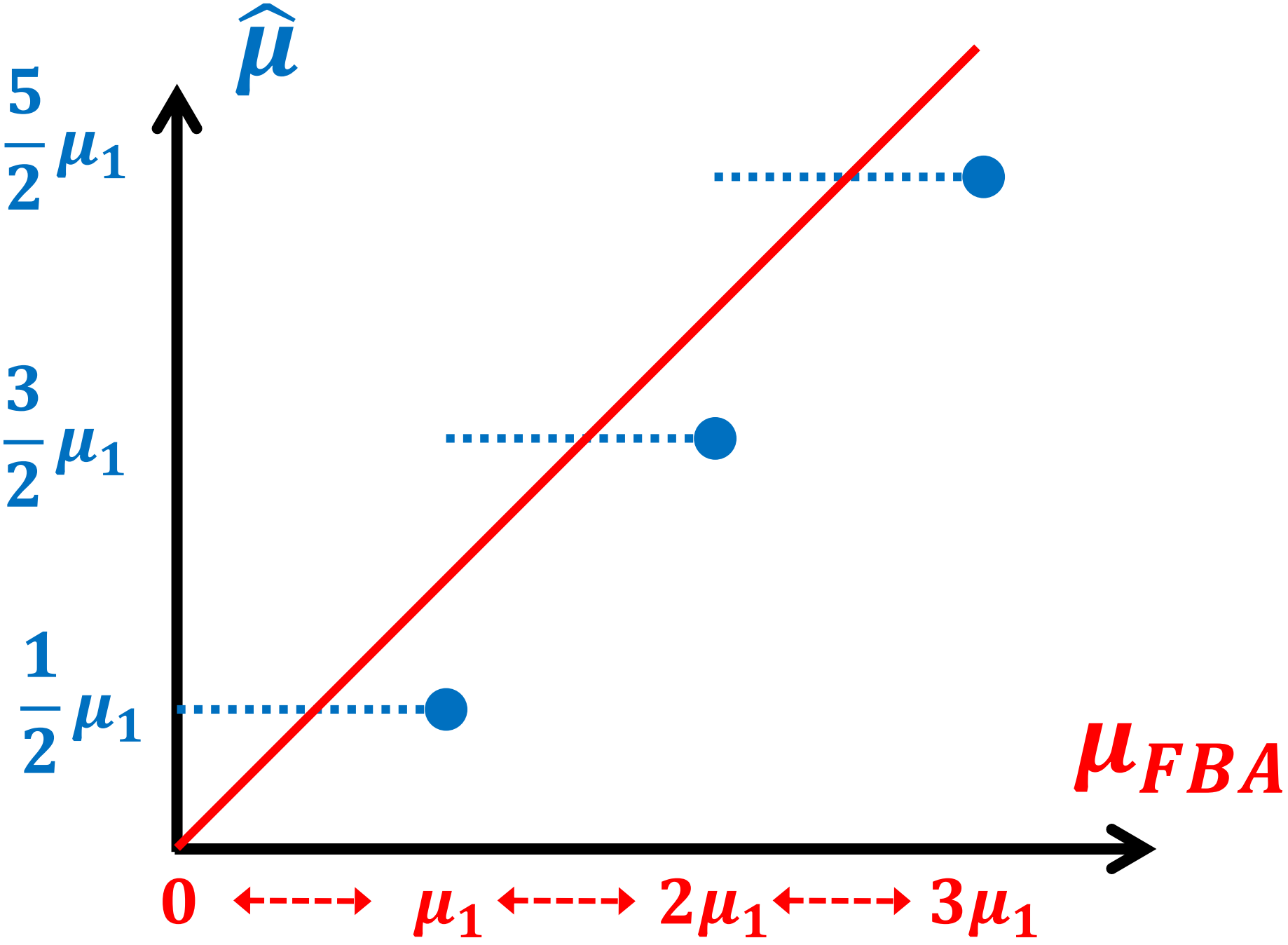
- Growth rate (μ) is discretized

$$v_l^{syn} - k_l^{deg} G_l - \mu G_l = 0$$



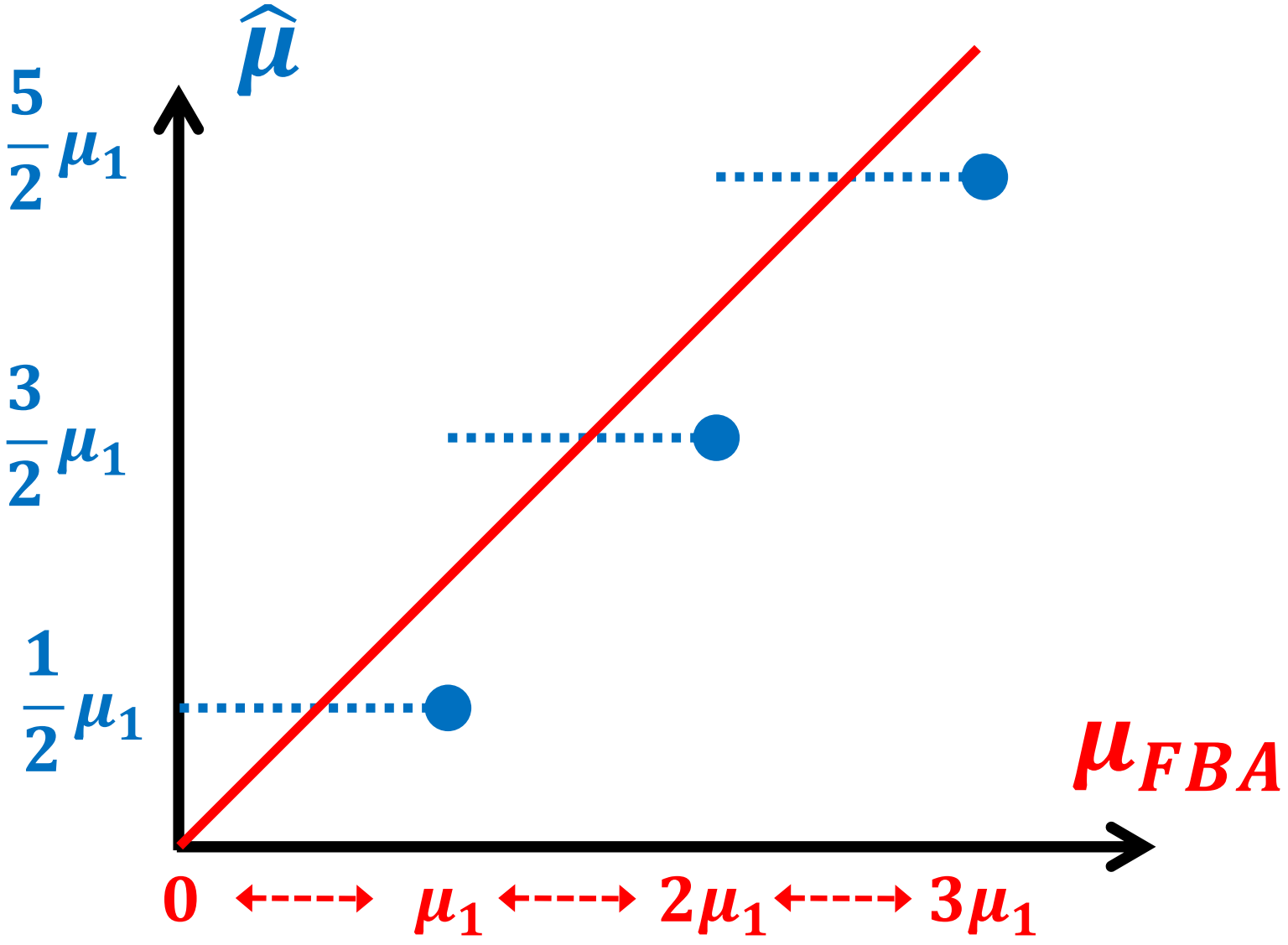
ETFL


$\hat{\mu}$
$\frac{5}{2}\mu_1$
$\frac{3}{2}\mu_1$
$\frac{1}{2}\mu_1$
μ_{FBA}
0
μ_1
$2\mu_1$
$3\mu_1$


$\hat{\mu}$
$\frac{5}{2}\mu_1$
$\frac{3}{2}\mu_1$
$\frac{1}{2}\mu_1$
μ_{FBA}
0
μ_1
$2\mu_1$
$3\mu_1$

Overcoming the *Nonlinear*:

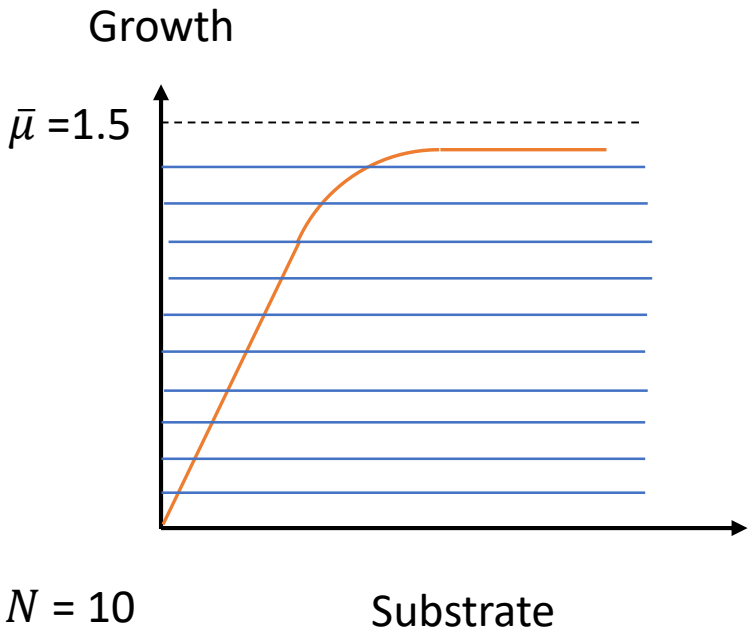
A piece-wise linear μ for Expression



Nonlinearity

- Growth rate (μ) is discretized

$$\mu \approx \sum_{s=0}^{\lceil \log_2 N \rceil} 2^s \delta_s \frac{\bar{\mu}}{N}$$



$$z_l^s = \delta_s G_l$$

$$\begin{cases} z_l^s \leq G_l \\ G_l + M\delta_s - M \leq z_l^s \\ z_l^s \leq M\delta_s \\ z_l^s \geq 0 \end{cases}$$

$$v_l^{syn} - k_l^{deg} G_l - \mu G_l = 0$$

$$v_l^{syn} - k_l^{deg} G_l - \sum_{s=0}^{\lceil \log_2 N \rceil} 2^s \frac{\bar{\mu}}{N} \delta_s G_l = 0$$

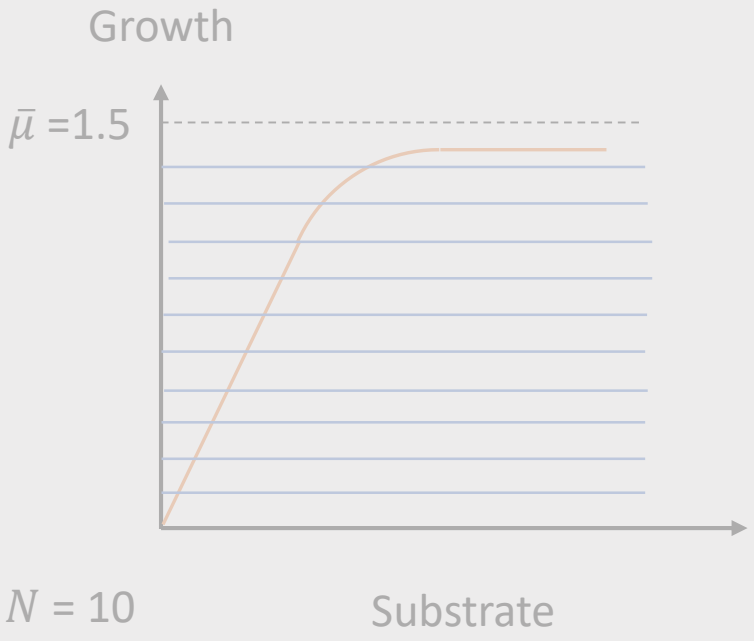
$$v_l^{syn} - k_l^{deg} G_l - \sum_{s=0}^{\lceil \log_2 N \rceil} 2^s \frac{\bar{\mu}}{N} z_l^s = 0$$

ETFL

Nonlinearity

- Growth rate (μ) is discretized

$$\mu \approx \sum_{s=0}^{\lceil \log_2 N \rceil} 2^s \delta_s \frac{\bar{\mu}}{N}$$



$$z_l^s = \delta_s G_l$$

$$\begin{cases} z_l^s \leq G_l \\ G_l + M\delta_s - M \leq z_l^s \\ z_l^s \leq M\delta_s \\ z_l^s \geq 0 \end{cases}$$

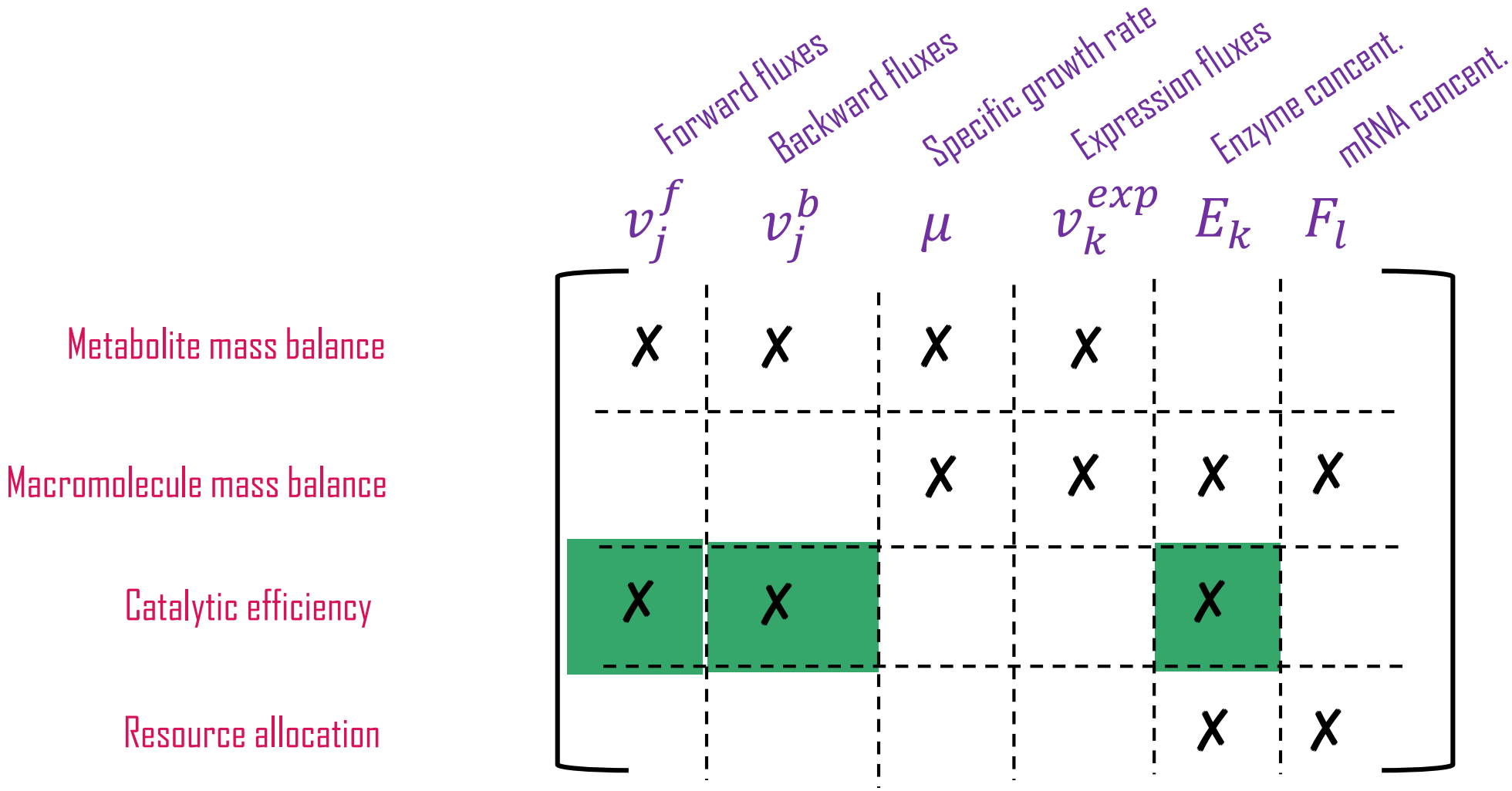
$$v_l^{syn} - k_l^{deg} G_l - \mu G_l = 0$$

$$v_l^{syn} - k_l^{deg} G_l - \sum_{s=0}^{\lceil \log_2 N \rceil} 2^s \frac{\bar{\mu}}{N} \delta_s G_l = 0$$

The resulting problem is:

- MILP (~ min)
- Solvable with double precision solvers
- Flexible

ETFL



They are related by the enzyme efficiency (k_{cat}s)

Efficiency

The rate of the recations: $v_j \leq v_{j,max} = k_{cat} E_j$

k_{cat} values are scarcely available

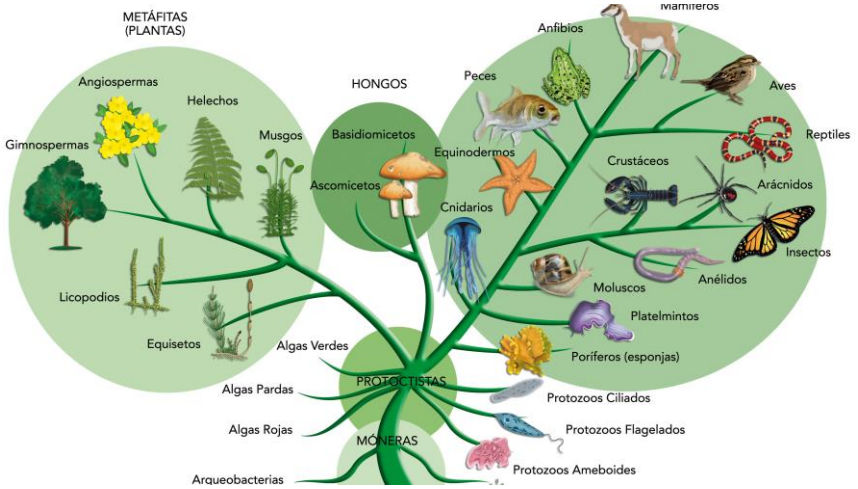
Efficiency

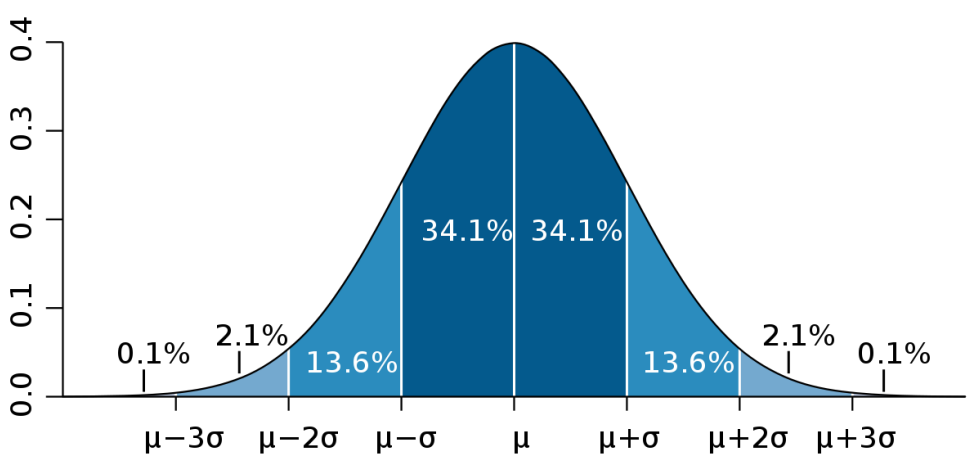
The rate of the recations: $v_j \leq v_{j,max} = k_{cat} E_j$

k_{cat} values are scarcely available

Two methods to infer k_{cat}s:

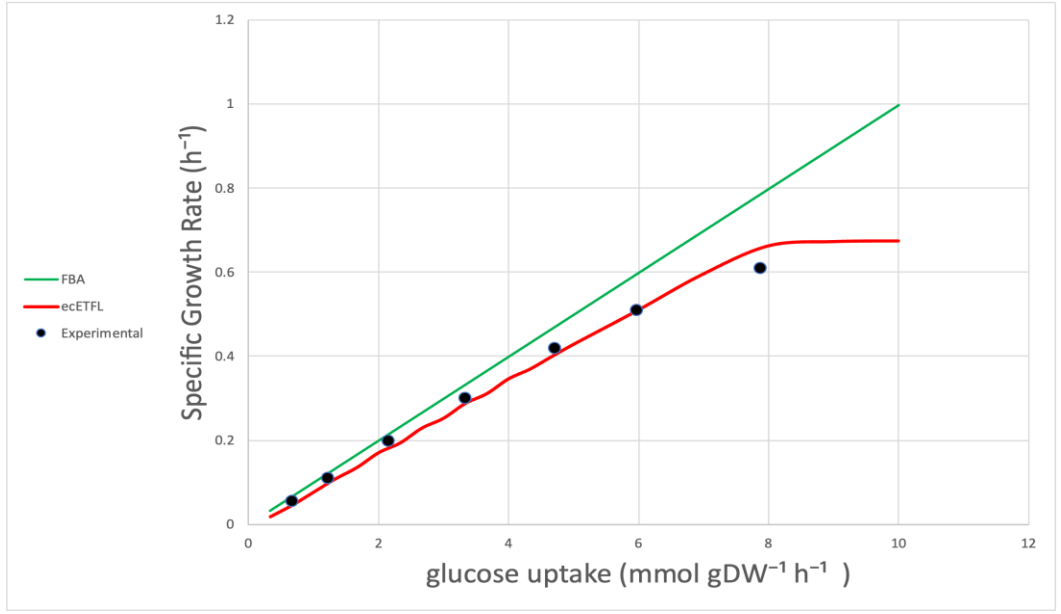
- Biological → based on bio-chemical similarity
- Statiscal → based on statistical distribution (also ML)



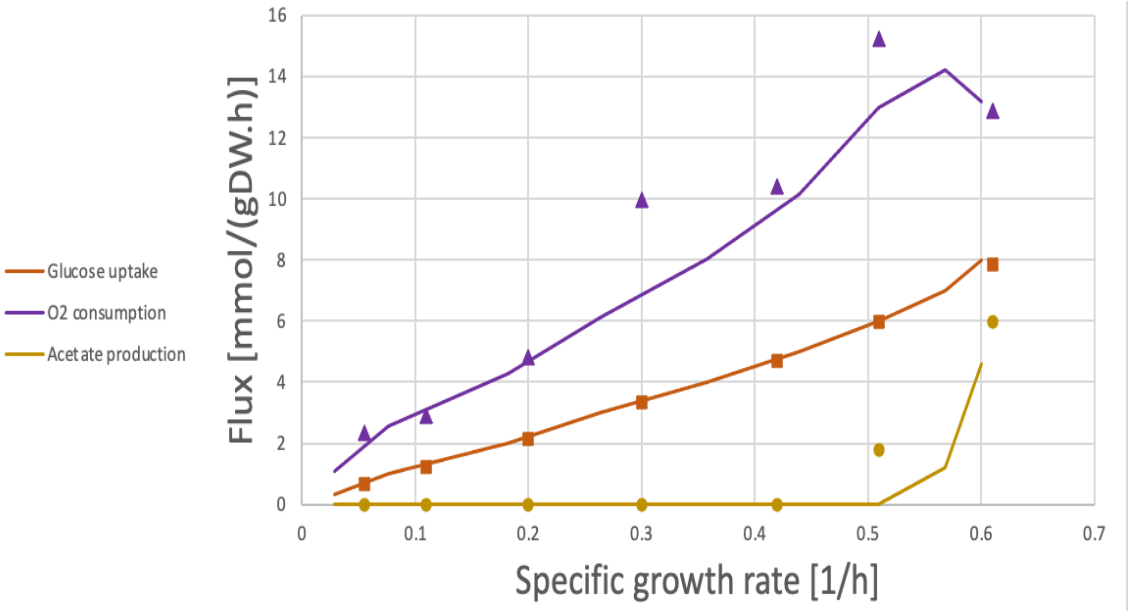


ME-model for *E. coli* (ETFL)

- Improved parameterization of the model → energetic requirements and resource allocation



Growth curve

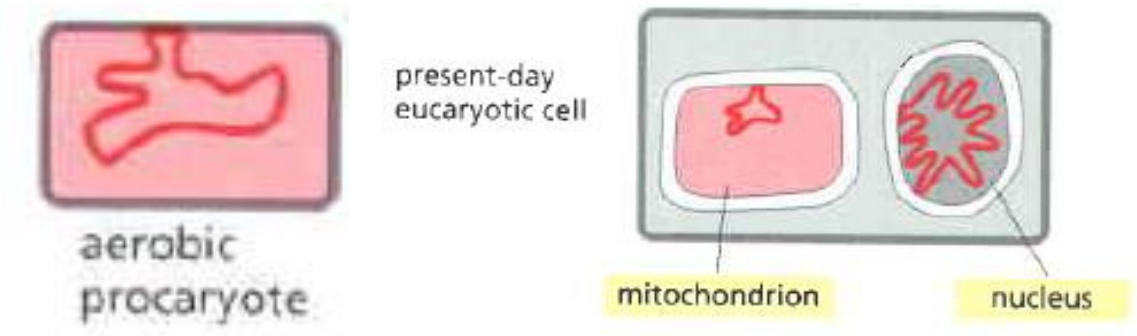


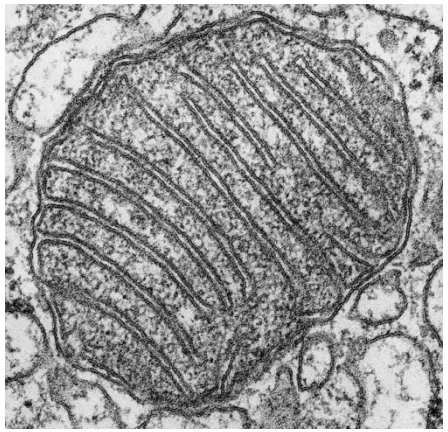
Overflow metabolism

ME-models for eukaryotes

Mitochondrial expression system:

- Defining an additional ribosome and RNA polymerase



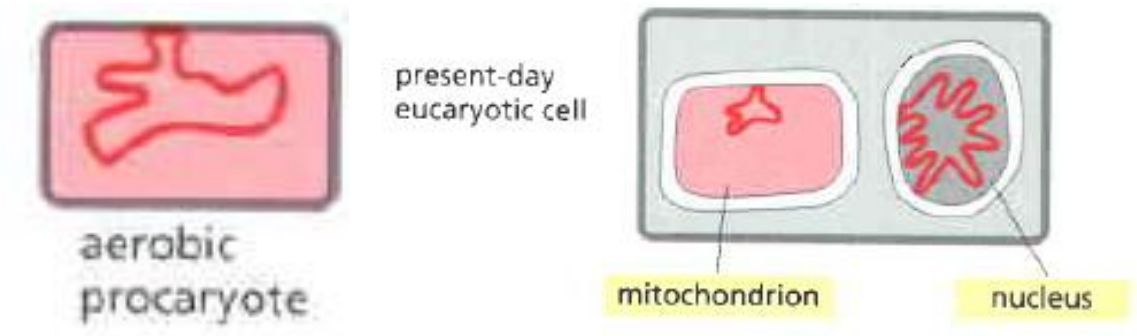


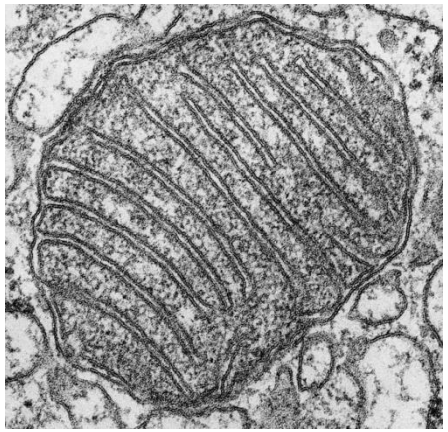
Alberts, Bruce. *Molecular Biology of the Cell: Hauptbd*. Garland, 2002

ME-models for eukaryotes

Mitochondrial expression system:

- Defining an additional ribosome and RNA polymerase
  - Finding data about composition and catalytic activity
  - Modifying the formulation to associate each gene with its corresponding RNA polymerase and ribosome





Alberts, Bruce. *Molecular Biology of the Cell: Hauptbd*. Garland, 2002

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-25158-6 OPEN

Check for updates

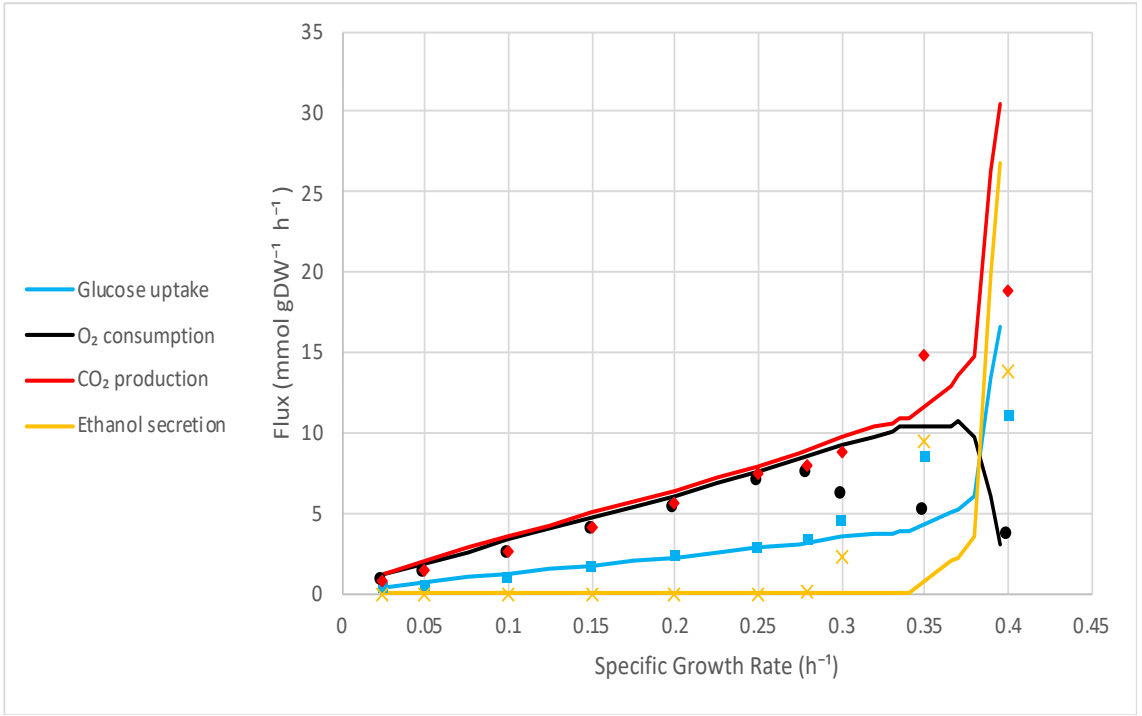
ME-model for *S. cerevisiae* (ETFL)

- The first ME-model for the yeast

A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics

Omid Oftadeh[1], Pierre Salvy[1,2], Maria Masid[1], Maxime Curvat[1,3], Ljubisa Miskovic[1] & Vassily Hatzimanikatis[1✉]

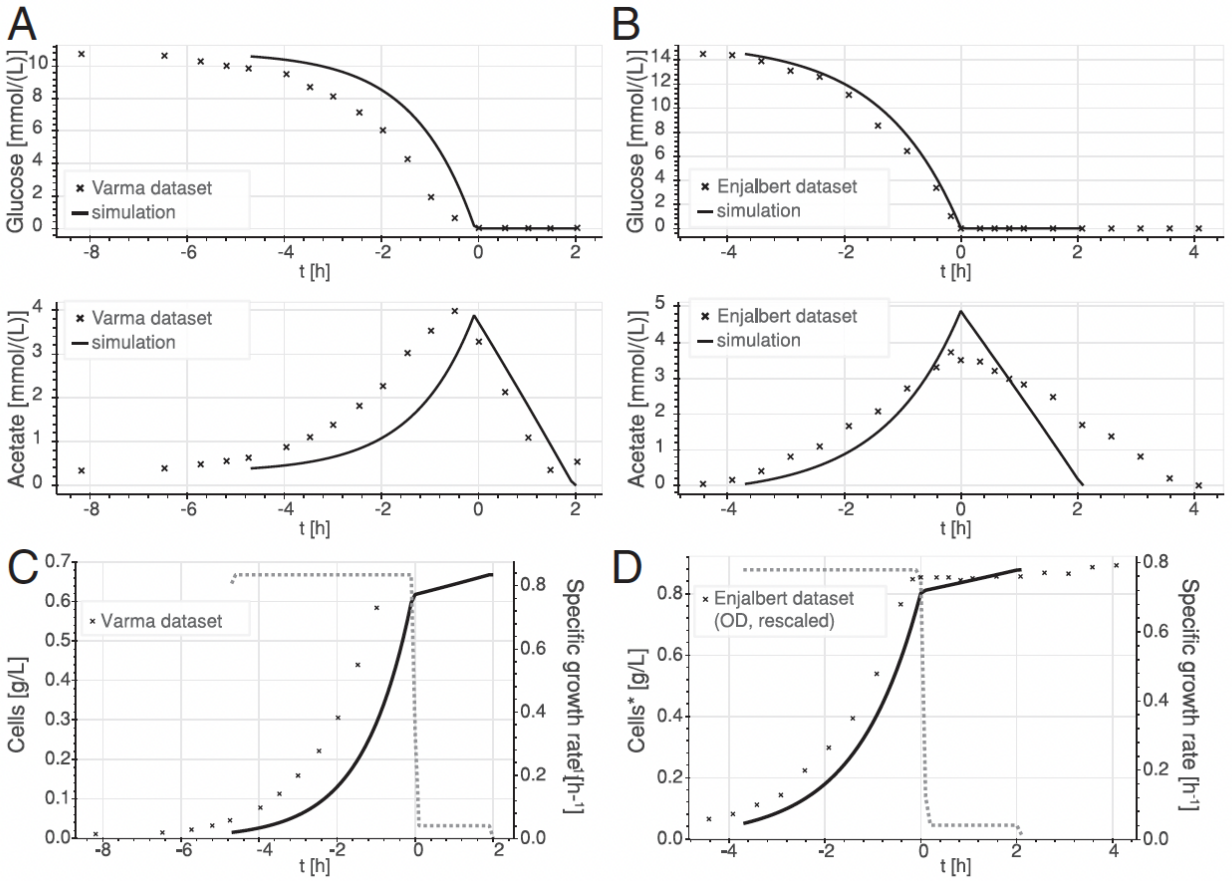
Growth curve



Overflow metabolism

Dynamic ME-models

- Temporal variation of fluxes and concentrations



PNAS

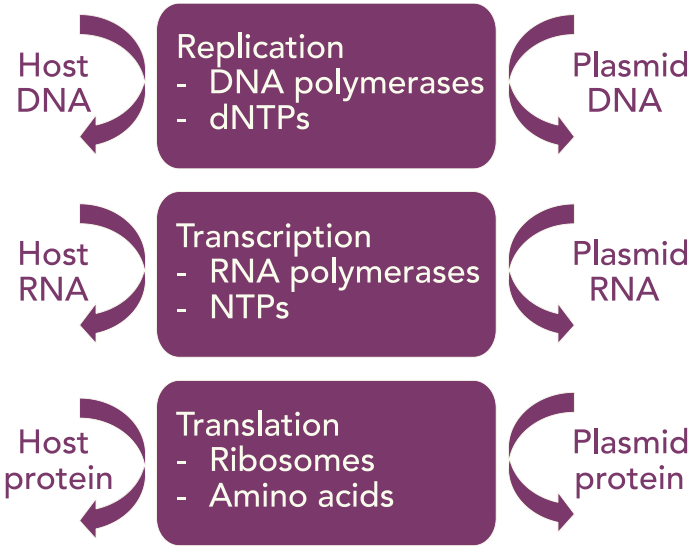
Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism

Pierre Salvy[a] and Vassily Hatzimanikatis[a,1]

Extending ME-models to recombinant organisms

The allocation of expression resources to the plasmid:

- Building blocks
- Machinery



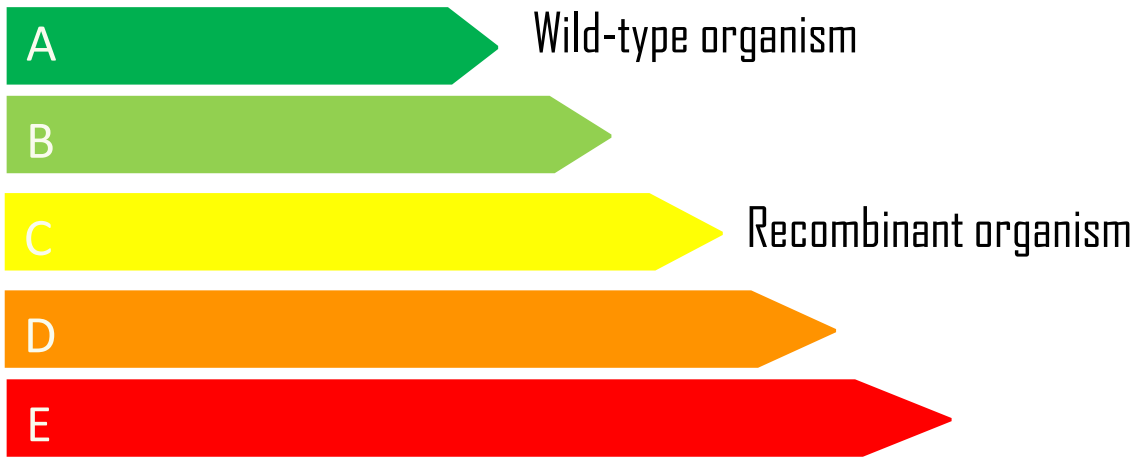
Extending ME-models to recombinant organisms

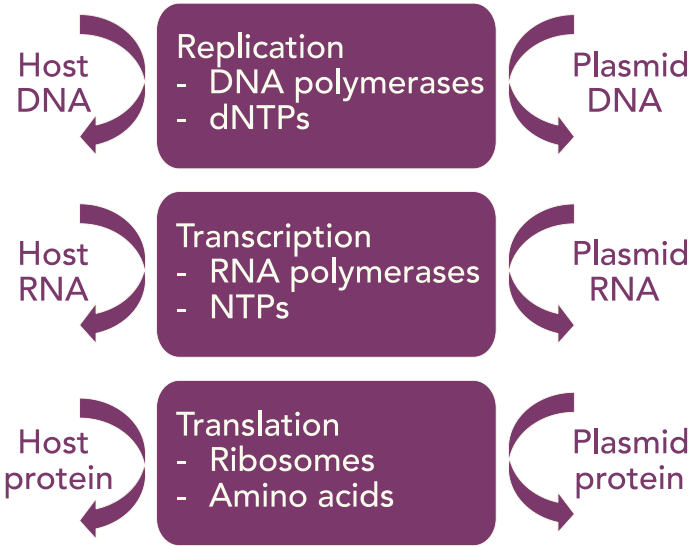
The allocation of expression resources to the plasmid:

- Building blocks
- Machinery

The energetic inefficiency caused by plasmid

- Refitting Non-growth associated maintenance





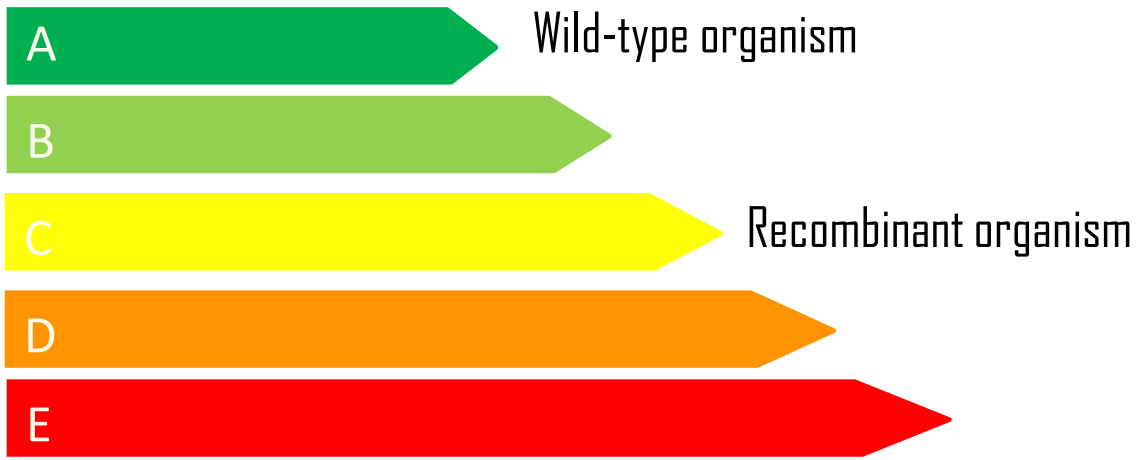
Extending ME-models to recombinant organisms

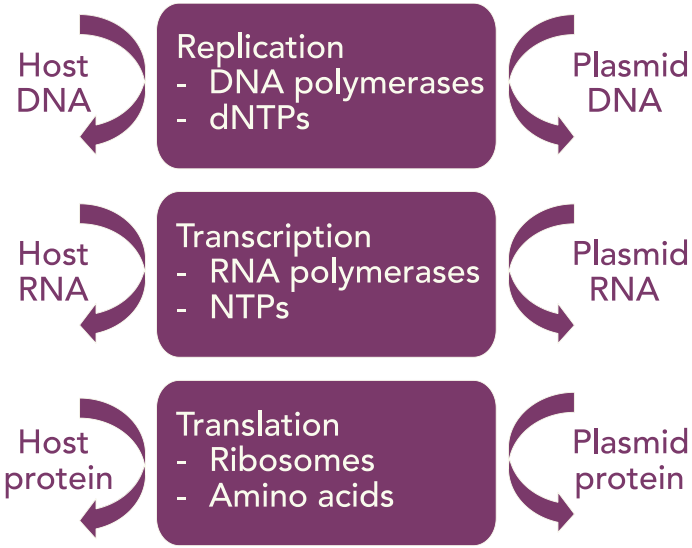
The allocation of expression resources to the plasmid:

- Building blocks
- Machinery

The energetic inefficiency caused by plasmid

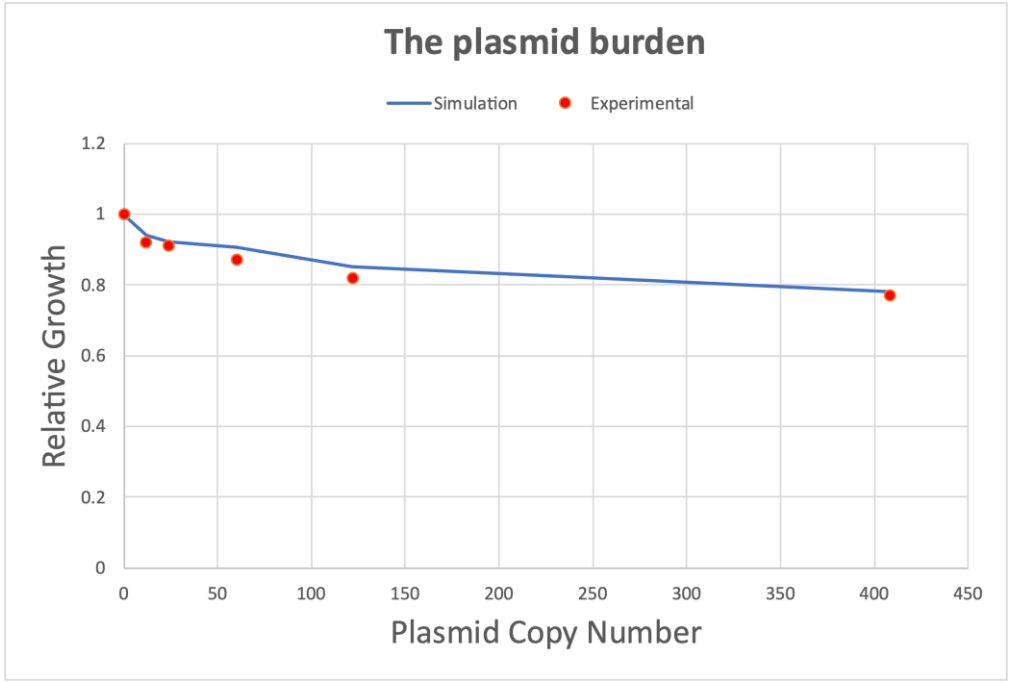
- Refitting Non-growth associated maintenance

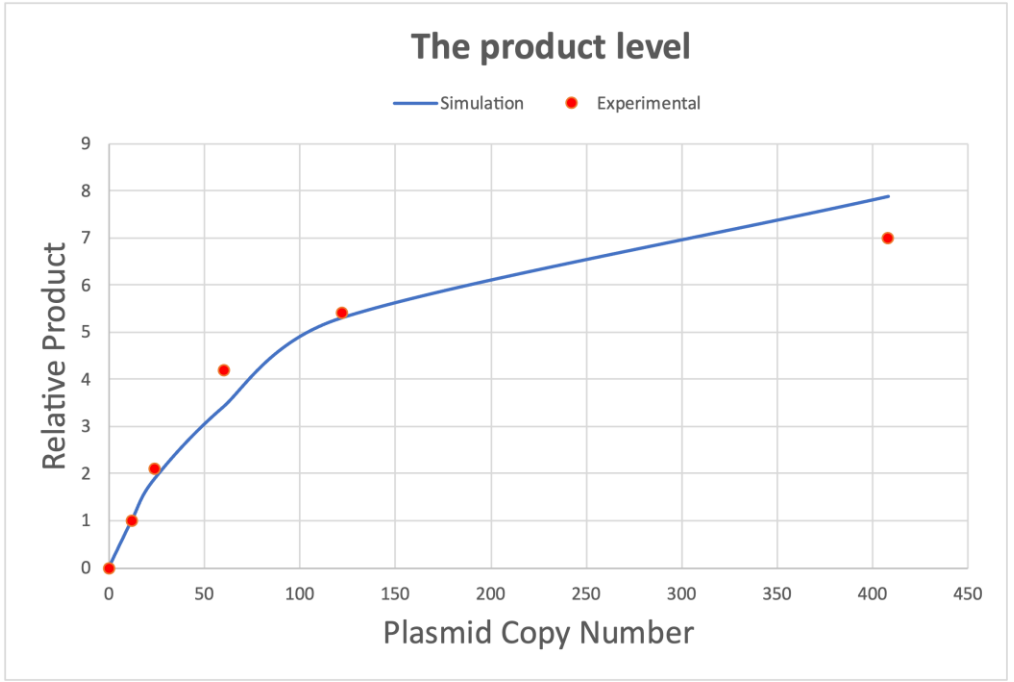




Capturing the plasmid burden

- ME-models can simulate the plasmid burden and the level of recombinant protein production





Seo, Jin-Ho, and James E. Bailey. "Effects of recombinant plasmid content on growth properties and cloned gene product formation in Escherichia coli." *Biotechnology and Bioengineering* 27.12 (1985): 1668-1674.

Capturing the plasmid burden

- ME-models can simulate the plasmid impact on exchange fluxes

Copy number	Glucose uptake (mmol gDW^{-1} h^{-1})	Growth (h^{-1})		Acetate secretion (mmol gDW^{-1} h^{-1})		Oxygen uptake (mmol gDW^{-1} h^{-1})	
		Mod.	Ex.	Mod.	Ex.	Mod.	Ex.
0	5.2	0.44	0.46	0	0	11	11
410	6.3	0.29	0.29	5.5	4.4	13.2	12.2

Z. Wang, L. Xiang, J. Shao, A. Węgrzyn, G. Węgrzyn, Effects of the presence of ColE1 plasmid DNA in Escherichia coli on the host cell metabolism. Microbial Cell Factories 5, 1-18 (2006).

Final remarks

- ETFL is an efficient formulation for ME-models
  - MILP (~ min)
  - Solvable with double precision solvers
  - Flexible
- The first ME-model to account for thermodynamics
- Easy reconstruction for different organisms