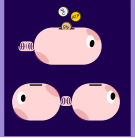
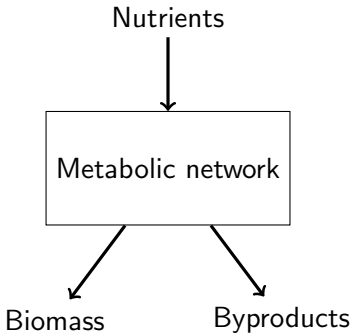


Optimization of metabolic fluxes

Steffen Waldherr

INRAE Département MathNum EURIP LEARNING PLANET INSTITUTE

A whole-cell perspective on metabolism



Properties that a model can try to describe

- Exchange fluxes / biomass production under given environmental conditions
- What is the internal network state to achieve certain exchange fluxes?
- How do the exchange fluxes / the internal network state react to external / internal perturbations?

Dimensions of metabolism

General overview

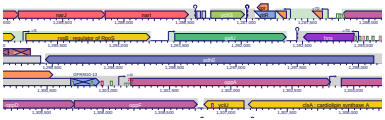
- **19 090** known biochemical compounds (KEGG COMPOUND database)
- **11 911** known biochemical reactions (KEGG REACTION database)
- **8 423** known enzymes (BRENDA database)

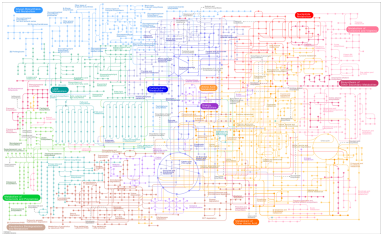
Organism specific view (`biocyc.org`)

Organism	# of reactions	# of metabolites
Escherichia coli	2 201	2 967
Saccharomyces cerevisiae	1 650	1 160
Homo sapiens	2 900	2 121
Arabidopsis thaliana	3 193	2 777

Reconstruction of metabolic networks from genome data

Genome ⟶ Metabolic network



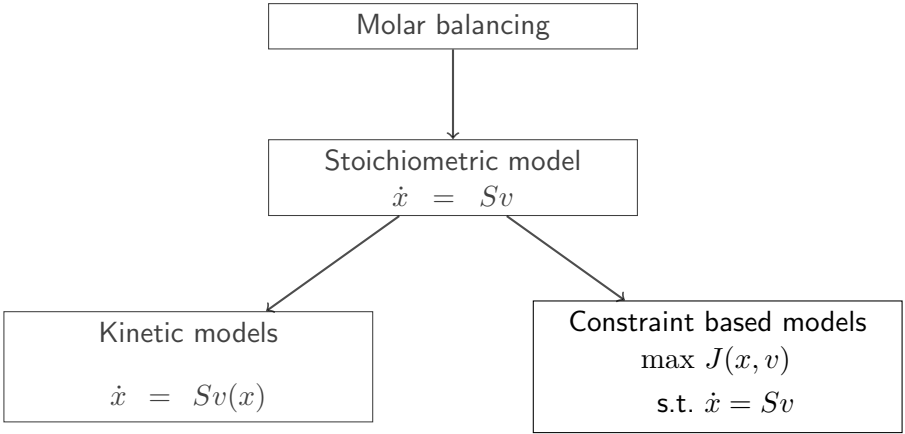


(kegg.jp)

1. Identify genes with enzymatic function (annotation / sequence homology)
2. Find matching reactions in reaction database
3. Add a biomass reaction (metabolic building blocks + energy (ATP) turnover)

Genome-scale metabolic network model

From metabolic networks to models



Flux balance analysis

Constraints applied to the network

- Intermediate / intracellular metabolites are assumed to be in a quasi-steady state:

  flux of producing reactions = flux of consuming reactions

- "Irreversible" reactions can only have flux in one direction
- Maintenance / housekeeping reactions can be constrained to have a minimum flux value (empirical)
- Nutrient uptake (exchange) reactions are constrained according to availability of nutrients in the considered environment

Optimization principle

- **Hypothesis**: Cells regulate fluxes within constraints to achieve an "optimal" configuration from an evolutionary perspective.
- In many applications, network solutions that **maximize flux through the biomass reaction** are taken

Constraints on fluxes

1. Steady state constraint

$$Sv = 0$$

   - Fluxes constrained to **subspace**

2. Irreversibility constraints on some fluxes (from thermodynamics/heuristics/empirical evidence)

$$v_i \geq 0, \qquad i \text{ irreversible}$$

   - Fluxes constraint to **flux cone**

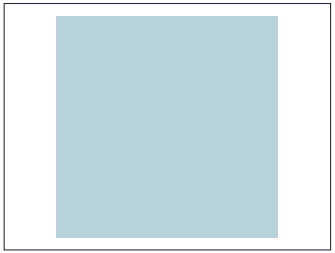
3. Flux bounds from capacity constraints, maintenance, ...

$$v_{i,min} \leq v \leq v_{i,max}$$

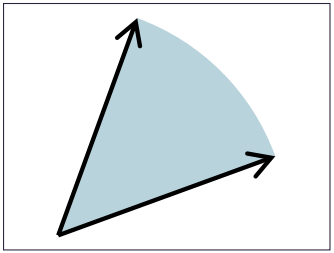
   - Fluxes constraint to **convex polytope**

Geometric illustration

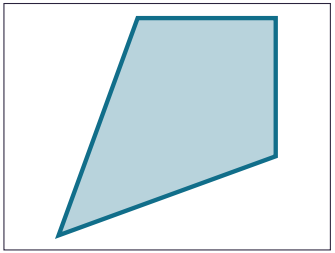
Flux space



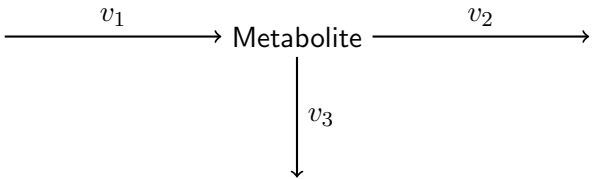
Flux cone



Flux polytope



Flux space $\rightarrow$ cone $\rightarrow$ polytope example



Construct the ...

- flux space;
- flux cone assuming $v_2,\ v_3 \geq 0$;
- flux polytope assuming $v_1 \leq 0.5$.

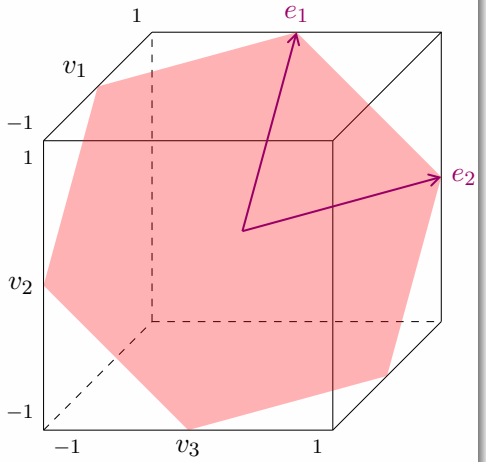
Molar balancing:

$$\dot{x} = \begin{pmatrix} 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Flux space from $Sv = 0$

- Plane defined by

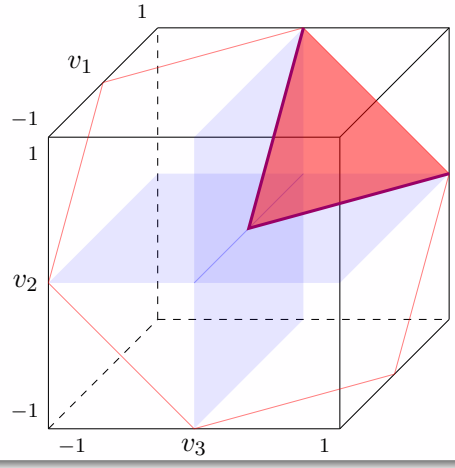
$$v_1 - v_2 - v_3 = 0$$



Flux cone

- Add irreversibility

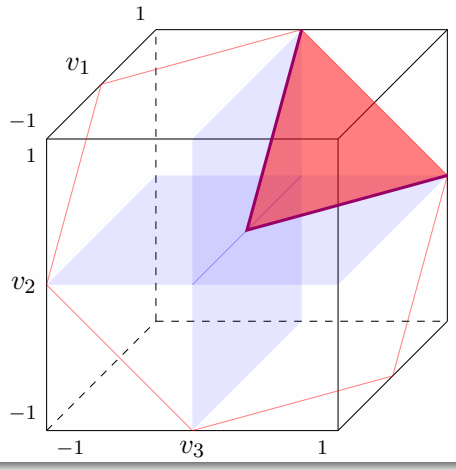
$$v_2,\ v_3 \geq 0$$



Flux cone

- Add irreversibility

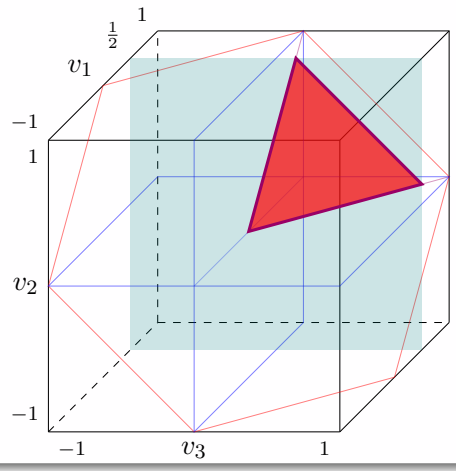
$$v_2,\ v_3 \geq 0$$



Flux polytope

- Add upper bound(s)

$$v_1 \leq 0.5$$



Setting up the constraint based model (CBM)

Constraint based model useful if non-trivial steady state fluxes exist

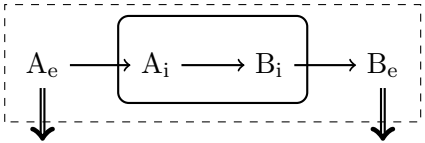
- The steady state equation
$$Sv = 0$$
should have a non-zero solution $v \Rightarrow$ non-trivial steady state flux space
- We need $\operatorname{rank} S < m$; most models have more reactions than metabolites anyway.

Metabolite / flux units

- In CBMs, metabolites are usually considered in **molar amounts per dry biomass**: $\mathrm{mmol/g}$
- Fluxes are then in $\mathrm{mmol/(gh)}$

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

Biomass composition

E. coli biomass composition

Compound	Proportion [% g/g DW]
Protein	72
DNA	4
RNA	10
Lipids	9
Polysaccharides	2.5
Mureine	2.5

Chassagnole *et al.* 2002, via bionumbers.hms.harvard.edu, ID 108705
Varies depending on environmental conditions (nutrients, aerobic/anaerobic, growth rate, ...)

Break down to metabolites

- 20 proteinogenic amino acids
- 8 D/R nucleotides
- phospholipids
- cofactors / vitamins
- ATP hydrolysis required for biomass assembly ("**growth associated maintenance**" GAM)

Biomass reaction

- Biomass reaction formalizes consumption of metabolites to generate biomass

$$v_{bio}: \qquad \sum_{i=1}^{n} c_i X_i \rightarrow 1 \text{ g dry biomass}$$

- Based on pre-determined constant biomass composition
- Coefficients c_i commonly in mmol / g dry biomass
- Unit of v_{bio} becomes $1/\mathrm{h}$: interpretable as dry biomass growth rate μ!

Maintenance

- "**Non-growth associated maintenance**" (NGAM):
  - membrane voltage gradients and osmolarity (ion pumps)
  - movement (flagella)
  - macromolecule (RNA/protein/carbohydrates) turnover
- Energy demand is commonly represented by a single ATP hydrolysis reaction

$$v_{maint}: \qquad \mathrm{ATP} + \mathrm{H_2O} \rightarrow \mathrm{ADP} + \mathrm{Pi} + \mathrm{H^+}$$

- Put as constraint into constraint based model
  - $v_{maint} \geq \alpha$ [mmol / (h · g biomass)]
  - NGAM rate estimates: *E. coli* 8.4 mmol/g/h; *S. cerevisiae* 1.0 mmol/g/h

Optimization principle

Constraint based model

$$Sv = 0$$

$$v_{i,min} \leq v_i \leq v_{i,max}$$

- **Underdetermined** system of equalities / inequalities: flux polytope
- How do we determine fluxes v that we expect to occur in nature?

Add an optimization objective

- **Hypothesis:** Cells regulate fluxes within constraints to achieve an "optimal" configuration from an evolutionary perspective.

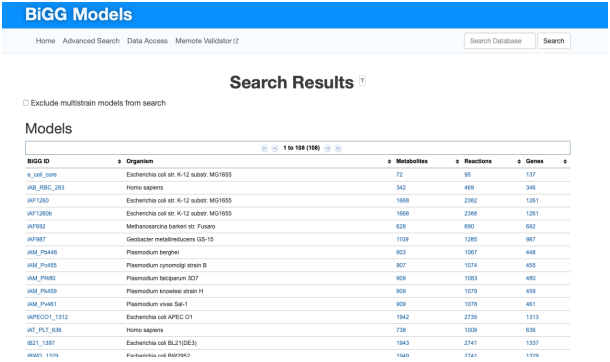
$$\begin{aligned} &\max\ J(v) \\ &\text{s.t.}\quad Sv = 0 \\ &\qquad v_{i,min} \leq v_i \leq v_{i,max} \end{aligned}$$

Useful objective functions

Type	Objective $J(v)$	Principle
Biomass yield	$\max v_{bio}$	Biomass flux at fixed max. substrate uptake
ATP yield	$\max v_{ATP}$	ATP flux at fixed max. substrate uptake
Minimal flux	$\min \|v\|^2$	Minimization of overall flux ($\sim$ enzyme usage)
Biomass flux yield	$\max v_{bio}/\|v\|^2$	Biomass yield per overall flux unit

Empirical evaluation of objective functions: Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Molecular Systems Biology, 3, 119.

Collections of constraint based models

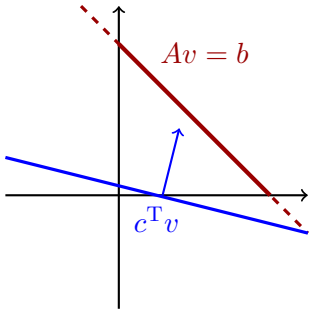


- BiGG models database: `http://bigg.ucsd.edu/models`
- ModelSEED (plant models): `https://modelseed.org/genomes/`
- BioModels database: `https://biomodels.net` (filter for "constraint-based model")

Linear programs

A linear program in standard form:

$$\begin{aligned} \max\ & c^{\mathrm{T}} v \\ \text{s.t. } & Av = b \\ & v \geq 0 \end{aligned}$$



Objective

$c^{\mathrm{T}} v$

Equality constraint

$Av = b$

Inequality constraint

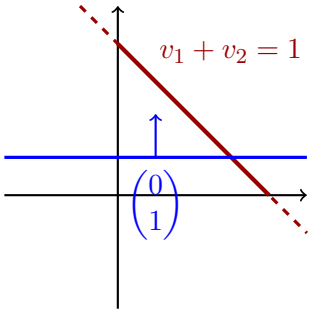
(Cone constraint) $v \geq 0$

Example

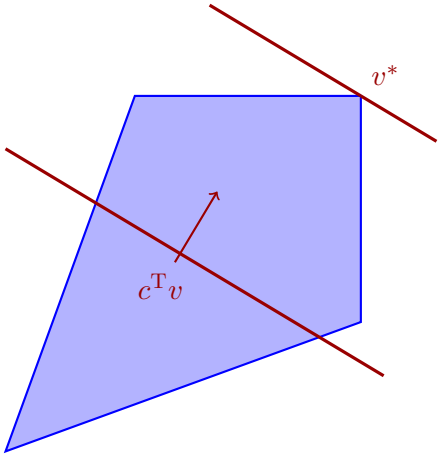
$$\max_{v_1, v_2} \; v_2$$
$$\text{s.t. } v_1 + v_2 = 1$$
$$v \geq 0$$

Thus:

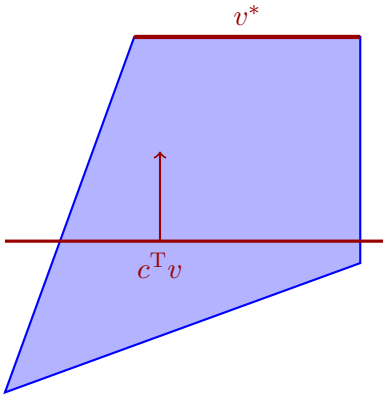
$$c^{\mathrm{T}} = \begin{pmatrix} 0 & 1 \end{pmatrix}$$
$$A = \begin{pmatrix} 1 & 1 \end{pmatrix}$$
$$b = 1$$



Generalized geometrical interpretation

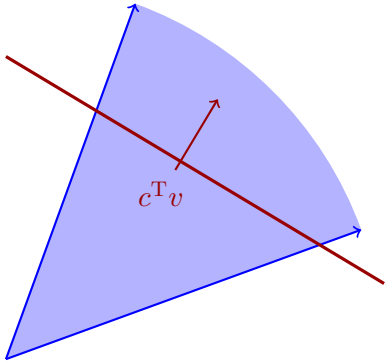


Non-uniqueness of optimal solutions



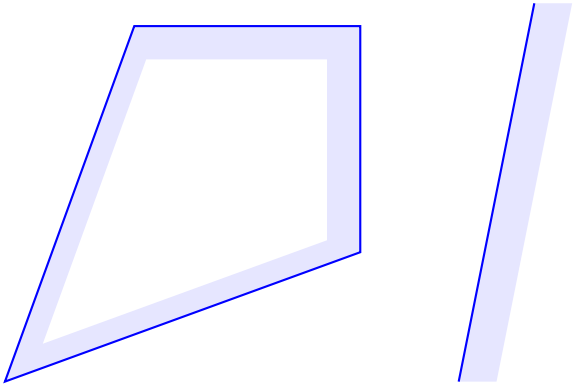
Set of optimal solutions is a **face of the polytope**

Unboundedness



Unboundedness: $\max c^{\mathrm{T}}v = \infty$

Infeasibility: Constraint set is empty



Example

$$v_1 + v_2 \leq -1$$

$$v_1, v_2 \geq 0$$

Flux balance analysis (FBA)

FBA to maximize biomass yield as LP

$$\begin{aligned} J^* = \max\ & v_{bio} \\ \text{s.t. } & Sv = 0 \\ & v_{i,min} \leq v_i \leq v_{i,max} \end{aligned}$$

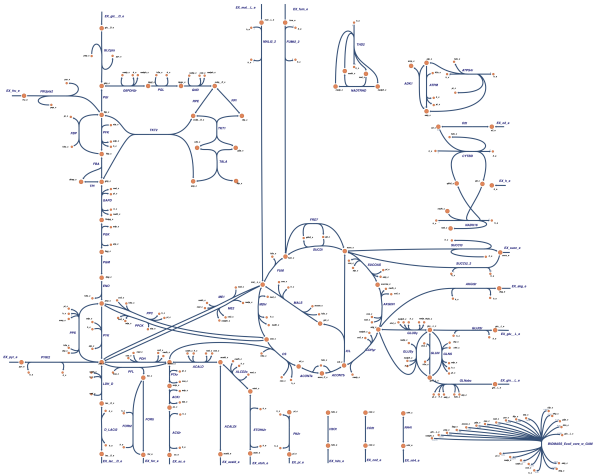
- Typical relevant constraint is glucose / oxygen uptake rate

$$-v_{e,gluc,max} \leq v_{e,gluc} \leq 0$$

- For practical reasons $v_{i,max} = M$ (10^6 mole/h/g) even if no capacity constraint
- Typically no unique optimal flux distribution v^*

FBA example: *E. coli* core

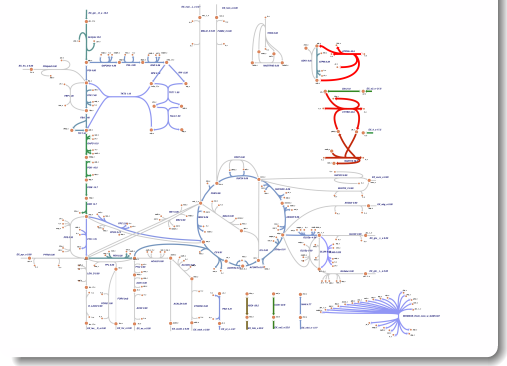
- Core carbon network from BiGG database: 72 metabolites, 95 reactions
- Network visualization from `https://escher.github.io/`



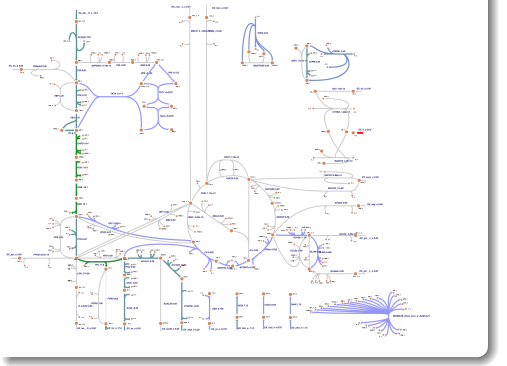
FBA results: comparing intracellular flux states

- With a graphical layout of the metabolic network is available: **graphical illustration of intracellular metabolic state**

Aerobic growth in *E. coli* core



Anaerobic growth in *E. coli* core



Made with `escher.github.io`

Dynamic FBA: general idea

- Put FBA models in a dynamic context (biomass growth, nutrient consumption)
- Starting from a mass balancing model like the Monod model:

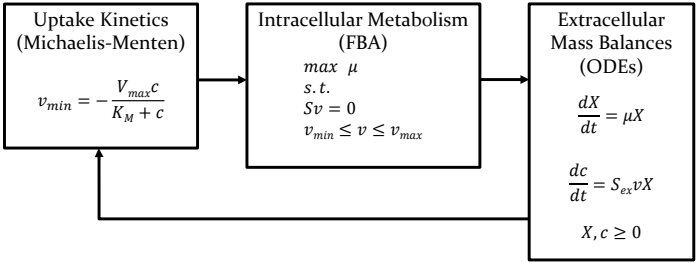
$$\frac{dX}{dt} = \mu(c)X$$
$$\frac{dc}{dt} = -\frac{\mu(c)}{Y_{X/c}}X$$

- 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)

Integrating the DFBA model parts

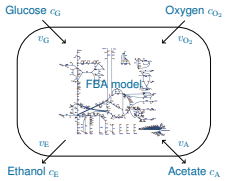


EPCB book, chapter "Optimal cell behavior in time"

- Enzyme kinetics for bounds $v_{i,min}(c)$, $v_{i,max}(c)$: usually only a couple of (uptake) reactions
- Optimal growth rate & exchange fluxes from FBA model are used in dynamic equations

DFBA: Example with *E. coli* core model

A



B

Exchange constraints

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

$$-30\,\frac{\text{mmol}}{\text{gDW h}}\,\frac{c_{O_2}}{20\,\frac{\text{mg}}{\text{L}}+c_{O_2}} \le v_{O_2} \le 0$$

$$0 \le v_E$$

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

C

Dynamic equations

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

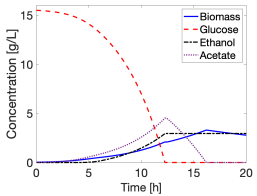
$$\dot{c}_{O_2} = v_{O_2} m_{O_2} X + k_L a(10\,\tfrac{\text{mg}}{\text{L}} - c_{O_2})$$

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

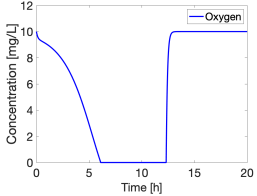
$$\dot{c}_\text{E} = v_\text{E} m_\text{E} X$$

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

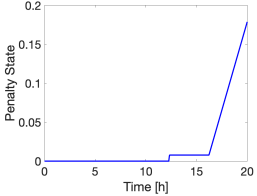
D



E



F



EPCB book, chapter "Optimal cell behavior in time"

Outlook: further extensions of FBA

- Thermodynamic FBA
- Resource allocation models:
  - ME models
  - Resource balance analysis
  - Dynamic enzyme-cost FBA

Exercise on `https://principlescellphysiology.org/book-economic-principles/index.html`

Run FBA on the carbon core model (Jupyter notebook → Google Colaboratory)