Enzyme economy in metabolic models

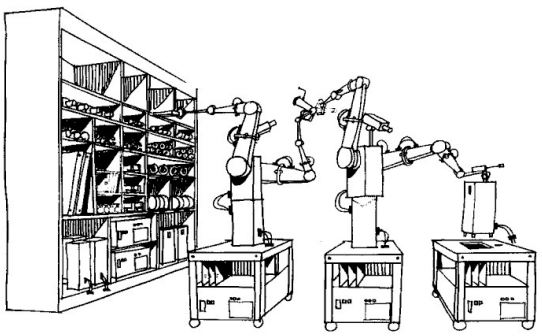
CompSysBio 2019 – Advanced Lecture Course on Computational Systems Biology

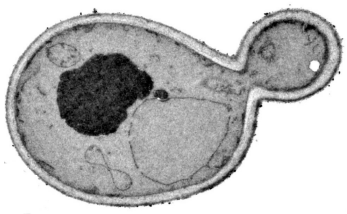
Blackboard teaching, April 1-2

Wolfram Liebermeister

Abstract

In this blackboard teaching, I will address an economic aspect of microbial metabolism: the protein cost associated with metabolic fluxes. I show how this cost can be approximated based on enzyme kinetics and on the assumption of minimal enzyme investments, and we will discuss how enzyme costs per unit flux, once they are known, can be used to predict metabolic fluxes and how simplified cost functions for Flux Balance Analysis can be derived. By considering a partitioning of protein resources between ribosomes and metabolic enzymes, predictions about enzyme cost can be translated into cellular growth rates. At the end of the course, we study constraint-based models that consider a fine-grained partitioning of the protein budget into individual enzymes. Such models predict metabolic strategies and protein investments solely from metabolic network structure, from physical and physiological constraints (such as limited cell space, protein composition, and presumable catalytic rates of enzymes), and from an assumed drive for fast growth.





1 Cell growth, resource allocation, and enzyme economy

Guiding questions for this course:

- What do living cells have to do in order to proliferate, i.e., to replicate all their components?
- How can they do so in cost-saving (or "optimal") ways, given physical and biochemical limitations?
- If cells function optimally, how will they behave and what will they look like?
- How can we describe all this by mathematical models?

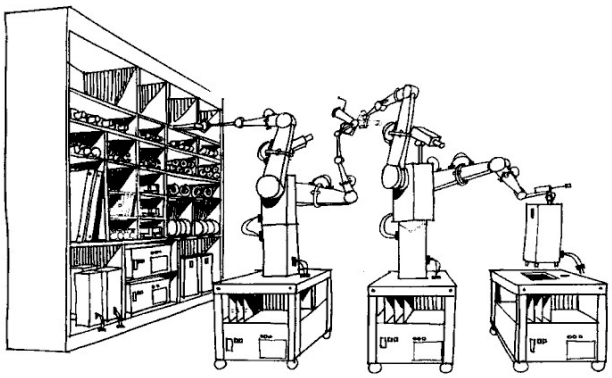
Initial questions about cell replication

- How can a living being emerge from non-living matter (sugar, water, and a couple of salts)?
- How fast can this happen?
- How much material and energy will be "wasted"?

1.1 A fundamental problem in biology: self-replicating cells

Self-replication as a "chemical reaction"

"Chemical reaction"	Inorganic compounds + cell $\longrightarrow$ 2 cells!
.. described as a "catalysed reaction"	Inorganic compounds $\xrightarrow{\text{cell}}$ cell



Source: wikipedia

Analogy: a self-replicating factory

- We consider a factory that produces all the machines it contains, and needs to replicate itself
- Knowing the types of machines and what they can produce, can we predict their numbers?
- How much material / energy will be wasted?
- Can we predict changes depending on energy supply (or energy cost), the need to produce extra products for the market, and possible machine failures, etc?

Self-replication of cells – some assumptions

- Self-replication only; no spontaneous generation
- No description of evolution: the entire machinery is already in place
- Assume a given environment
- Self-replication happens through continuous growth (of an existing cell! and division)
- Self-replication (and life) inevitably implies some "waste" of energy

"Waste" is inevitable: organisms feed on negentropy

- Simple inorganic molecules (in an "unordered" chemical solution) are transformed into complicated macromolecules (in a structured cell)
- The entropy decreases, and this must be compensated by producing even *more* entropy in the environment
- Also to *maintain* a living cell (e.g. to repair damage), entropy must be exported

$\Rightarrow$ Living systems feed on negentropy [1]

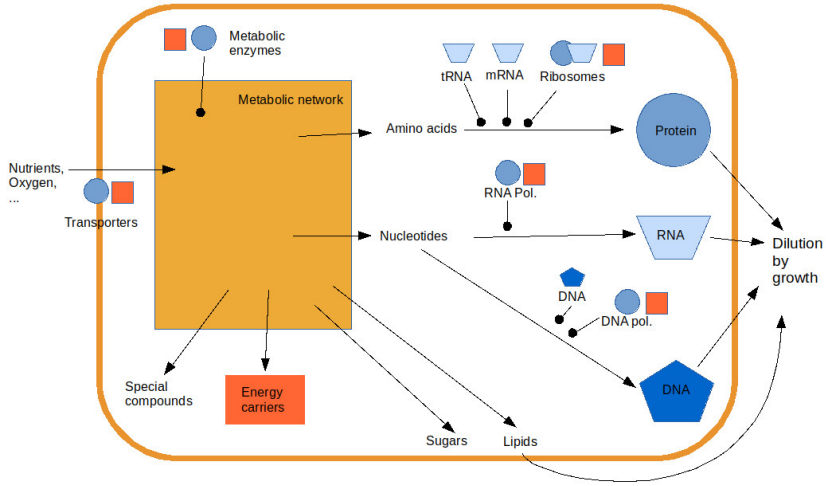
Requirements for cell survival

- Cells are alive and fragile – they need to be able to replicate (almost) all of their components
- Cells are dynamic – they need to coordinate and regulate all these processes
- Cells compete – they need to replicate faster, and be more efficient or resilient than others

1.2 Cell growth

What do you know about cell growth?

- How does growth happen in cells? What biological processes are involved?
- Exponential growth
- On what conditions in the environment does the growth rate depend?
- How would you define "efficient growth"?



Requirements for self-replication

Each component of a cell must be replicated!
⇒ For each of its components, a cell must also contain the machine that makes (or imports) this component

- What are the components in a cell (to be replicated)?
- By what machines are the different components replicated?
- Are the replication steps also possible without machines? Under what conditions? At what disadvantage?
- How are bigger structures assembled (e.g. how are cell shape and organelles maintained)?

Fast growth or substrate-efficien growth?

Evolutionary success ("fitness") depends on:
(i) *long-term survival* in a (ii) *specific (possibly changing) environment*, which may be (iii) *shaped by the organism*.

Growing "efficiently" can mean

- wasting little material or energy from the environment
- growing fast, thus being able to outcompete others!

These are not the same things (selection advantage for high cell count, *vs* selection advantage for fast replication)

Why is fast replication a relevant task (i.e. interesting for modelling)?

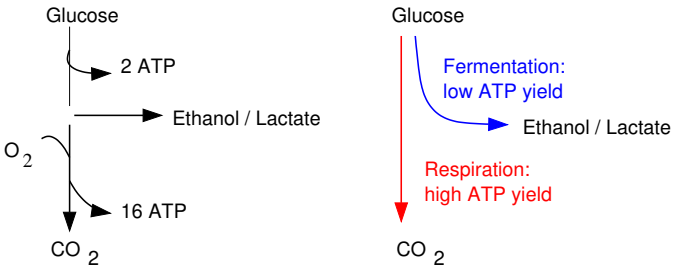
- Fast growth is not the only "natural" task of all cells.
- Selection may favour fast growth, high cell yield, withstanding harsh conditions, or many other things!

What reasons do you see for studying fast growth?

1.3 Metabolic strategies

Choice between metabolic pathways

A typical situation: a cell "can choose" between an efficient and an inefficient pathway ("metabolic strategy"). Where should it invest its material (e.g. protein)?



But what do we mean by "efficiency"?

(i) Substrate efficiency (substrate demand)

- Metabolism: e.g. ATP (or biomass precursors) produced per unit of glucose consumed
- Cell: new cells produced per per unit of glucose consumed ("yield on substrate")

(ii) Enzyme efficiency (enzyme demand)

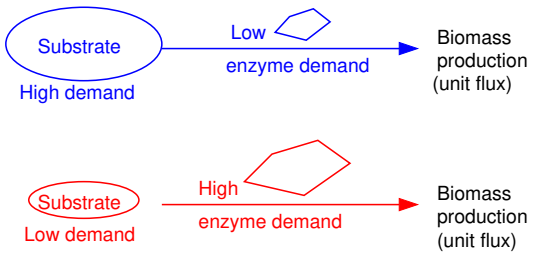
- Metabolism: Production flux (ATP, biomass precursors) per total amount of metabol ic enzyme)
- Cell: Any self-replicator should use *a small amount of catalyst* to grow fast

Why is enzyme efficiency relevant for fast growth? Assumptions:

1. Constant metabolic enzyme budget, i.e. amount per cell (or: enzyme density in cell volume)
2. Maximal growth requires maximal biomass production (amount per time)
3. Whole metabolic flux can be proportionally increased by increasing the enzyme amount;

$\Rightarrow$ Maximal growth requires maximal biomass production *per* total amount of metabolic enzyme!

$\Rightarrow$ minimise the total amount of metabolic enzyme *per* biomass production rate!



Substrate efficiency and enzyme efficiency may be in conflict!

"Rate/yield trade-off", claimed by some authors:

- Substrate efficiency $\leftrightarrow$ enzyme efficiency
- High biomass yield $\leftrightarrow$ High growth rate
- Classical Flux Balance Analysis (FBA) $\leftrightarrow$ FBA with bounds on total enzyme

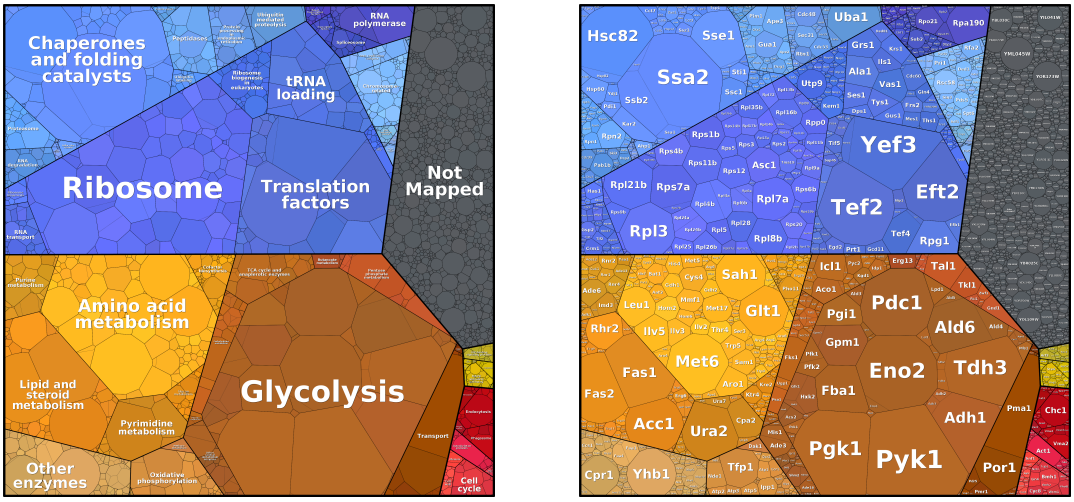
1.4 Resource allocation in cells

In which functions should an organism invest, and how much?

For example, how much should plants invest in leafs or roots?

$\Rightarrow$ Idea of "optimal resource allocation"

Yeast (S. cerevisiae) proteome



Images: www.proteomaps.net. Data: Nagaraj *et al.* (2012) [2]

What are the main protein investments in cells?

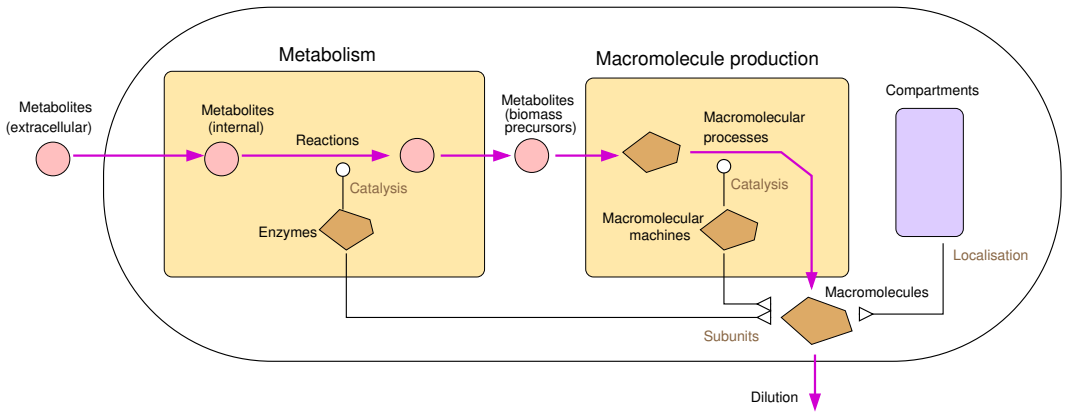
- "Central dogma": Production of DNA, RNA, and protein
- Metabolism
- Membranes and transport
- Stress reponse, repair
- and many others

Other tasks of the cell besides replication

- Basic maintenance (energy, precursors)
- Regulation
- Robustness and adaptation
- "Social behaviour" (quorum sensing, cell communication, phenotype switching, ..)
- "Superpowers" of specialised cells: movement; fighting (bacterial toxins, immune cells); producing materials such as hair or bones; special physical properties (crystallins in the eye lens)

How can we explain these pictures?

- What proteins are present?
- What are their relative amounts?
- How does the picture change between growth conditions or cell types?



Why focus on enzyme investments? (versus other types of molecules?)

- Protein is a major "economic investment" of the cell (occupying space, costly production and maintenance)

- Assuming a fixed (maximal) protein budget: minimise biomass production per total amount of overall metabolic enzyme (i.e. "enzyme efficiency for biomass production")
- Low enzyme efficiency: high enzyme demand per given production flux ⇒ high "enzyme cost"
- Overall enzyme efficiency depends on the choice of pathways (flux modes), enzymes' kinetic properties (e.g. k_{cat} values), substrate availability and concentrations, and the "choice" of intracellular metabolite and enzyme levels
- In an "enzyme-efficient" metabolic strategy, ideally every reaction should be enzyme-efficient (small enzyme investment per flux).
  This is in fact impossible, and compromises between reactions may be inevitable!
  Thus: avoid very inefficient single reactions (that would cause a very large enzyme demand)!

Optimality-based models of metabolism and entire cells

We need not only mechanistic models, but also optimality principles!

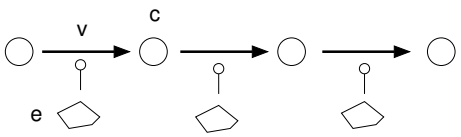
- Basic assumption: self-replication requires (or *is* more or less) macromolecule production
- Working hypothesis: cells need to maximise the biomass (precursor) production rate per total amount of metabolic enzyme
- First, focus on metabolism – how can we predict the enzyme levels?
- Later: metabolism and macromolecule production (mostly protein production)

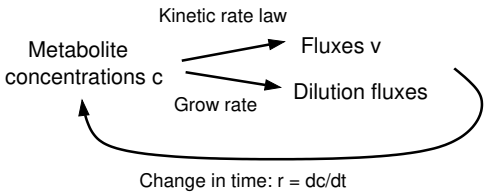
2 Kinetic and constraint-based metabolic models

What information do we need for studying the economics of self-replication?

- A blueprint of the cell in question (list of components and interactions: "topics")
- Simulation models for cellular processes (e.g., metabolism and protein production), implementing physical laws and biochemical facts ("dynamics")
- Ideas about cells *should* function, formulated as mathematical optimality problems ("economics")

2.1 The laws of kinetic metabolic models





Variables:

- Metabolite concentrations c_i
- Metabolic fluxes v_l
- Enzyme concentrations e_l
- Metabolite *net* production rates $r_i = \frac{\mathrm{d}}{\mathrm{d}t} c_i$

Three main types of laws:

1. **Mass balance** (involves stationarity condition; dilution)
2. **Reaction kinetics** (involves enzyme kinetics, thermodynamics, regulation)
3. **Space restrictions**
   – "Physiological range" (for individual compounds) $\mathbf{c}_{\min} \leq \mathbf{c} \leq \mathbf{c}_{\max}$ $0 \leq \mathbf{e}$
   – "Density constraint" (for metabolites, enzymes, or both) $\sum \alpha_i c_i + \sum \beta_j e_j \leq d$

Dynamic biochemical network models ("kinetic models")

System dynamics described by ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} c_i &= \sum_l n_{il}\, v_l \\ v_l &= v_l(\mathbf{e}, \mathbf{c}) = e_l \cdot \kappa_l(\mathbf{c}) \end{aligned} \tag{1}$$

Dilution in growing cells (with growth rate λ) leads to equation

$$\frac{\mathrm{d}}{\mathrm{d}t} c_i = \sum_l n_{il}\, v_l - \lambda\, c_i \tag{2}$$

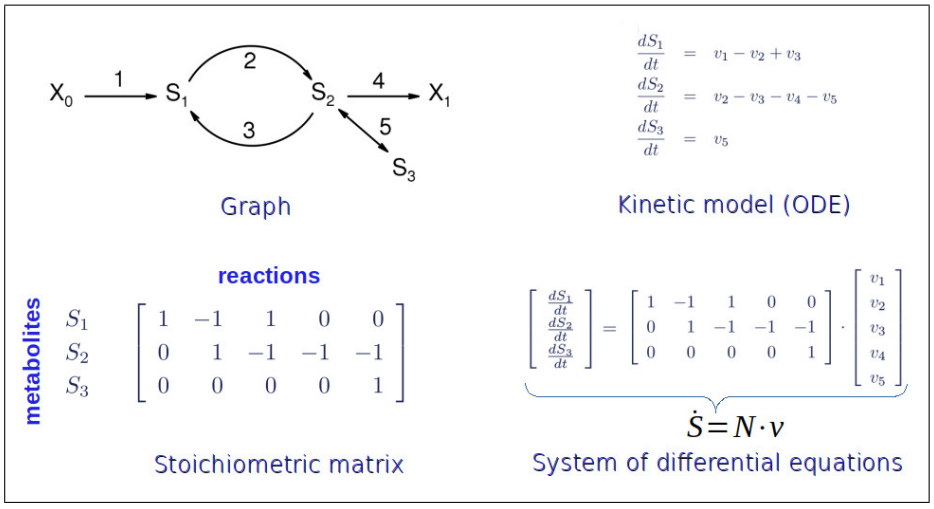
2.2 Stationary states and flux balance analysis

Metbolic networks in stationary state

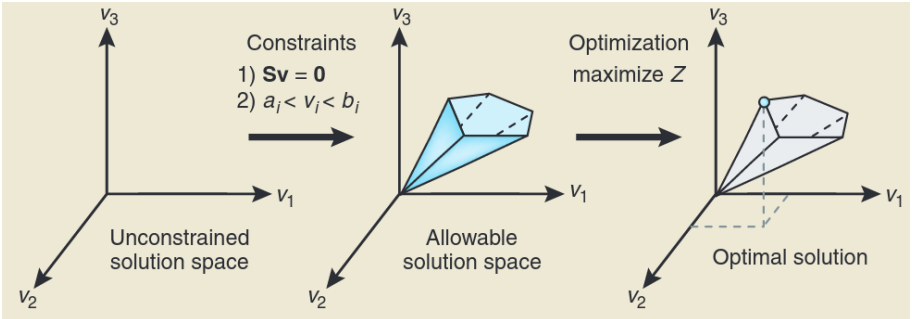
- Definition: Internal and external metabolites
- Definition: Stoichiometric matrix

Stationarity condition

$$0 = \sum_l n_{il}\, v_l = \mathbf{N}\,\mathbf{v} \tag{3}$$



- Dilution at growth rate λ: described as "loss of concentration", with dilution fluxes $v_i^{dil} = \lambda\, c_i$
- Definition: Flux mode = stationary flux distribution
- Space of flux modes: "flux cone"



Flux Balance Analysis (FBA) methods

- **Classical FBA:**
  Maximise $\mathbf{z} \cdot \mathbf{v}$ subject to stationarity and flux bounds $\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}$
- **FBA with flux minimisation** [3]
  Minimise flux sum $\sum |v_l|$ subject to FBA constraints and fixed benefit $z \cdot v$
- **FBA with molecular crowding** [4]:
  Maximise benefit $\sum_l z_l \, v_l = \mathbf{z} \cdot \mathbf{v}$ subject to FBA constraints and limited enzyme burden $\sum_l w_l \, |v_l| \leq w_{\mathrm{tot}}$

How to justify bounds on flux sums

- Assumption: (absolute) fluxes **proportional** to enzyme level
- **Limited capacity for protein** (density in cell or on membranes; capacity for production)
- Total enzyme burden should be **bounded or minimal** (leaving resources for other processes)

How bounds on flux sums change the flux predictions

- Bounds on sums enforce allocation of limited resources between metabolic pathways
- Allows for prediction of yield-inefficient (but enzyme-efficient!) metabolic strategies
- Extension to ribosomes: below this principle will be applied to resource allocation between metabolic enzymes and protein synthesis (requiring ribosomes).

2.3 Kinetic rate laws

Mass-action rate law (reversible) for reaction S $\leftrightarrow$ P

$$v(s,p) = k^+ \, s - k^- \, p \tag{4}$$

Michaelis-Menten rate law (saturable; reversible version)

$$v = e \, \frac{k^+_{\mathrm{cat}} \, s/K_{\mathrm{s}} - k^-_{\mathrm{cat}} \, p/K_{\mathrm{p}}}{1 + s/K_{\mathrm{s}} + p/K_{\mathrm{p}}} = e \, \kappa(s,p) \tag{5}$$

- Assume enzyme concentration e as prefactor
- Catalytic rate κ is smaller than catalytic constant k^+_{cat}

Simplified rate laws

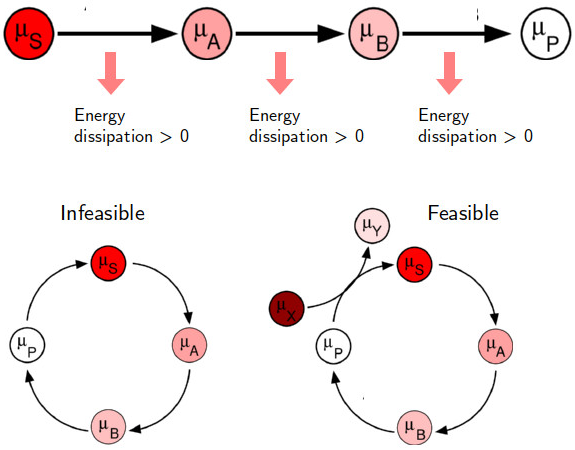
- Linear rate laws (linear in substrate concentration): $v = e \cdot c_{\mathrm{substrate}}$
- Fully saturated enzyme (rate independent of substrate): $v = k^+_{\mathrm{cat}} \cdot e$

2.4 Thermodynamic constraints on fluxes and metabolite concentrations

Kinetic rate laws are enzyme-specific – but they have some general properties due to thermodynamics

Thermodynamic constraints

- Entropy must be produced in every reaction!
- Equivalent condition (assuming constant pressure and temperature): dissipation of Gibbs free energy (GFE)
- Describe GFE difference in a reaction as a "thermodynamic driving force" θ



Chemical potentials

(Chemical potential difference $\Delta\mu_l$ is also called reaction Gibbs free energy $\Delta G'$)

$$\mu_i = \mu_i^\circ + RT \ln c_i \tag{6}$$

Thermodynamic driving force (for reaction S $\leftrightarrow$ P)

$$\theta = -\Delta\mu/RT = \ln K_{\text{eq}} - \ln \frac{p}{s} \tag{7}$$

Flux sign constraint The velocity ratio implies that

$$\text{sign}(v) = \text{sign}(\theta) = -\text{sign}(\Delta G') \tag{8}$$

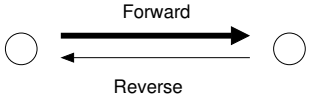
(usually used in the "weak version": $v = 0$ is always allowed, e.g. for kinetic or regulatory reasons!)

$\Rightarrow$ Positive entropy production $\sigma = \frac{v\,\theta}{T}$, in agreement with second law of thermodynamics!

Consequences of sign constraint

The driving force θ is a linear function of the log-concentrations $\ln c_i$

- Given the concentrations c_i, we obtain θ, and therefore all flux directions! (where v may also vanish)
- Given flux directions $\text{sign}(v_l)$ yields the set of possible metablite profiles (a polytope in log-metabolite space)
- If the polytope is empty: $\mathbf{v}$ is thermodynamically impossible
  (indicated by "loopless" criterion; example: perpetuum mobile)
- With concentration bounds: "thermo-physiologically impossible"



Quantitative effect of driving forces

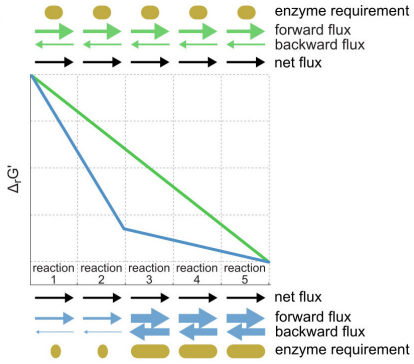
Compared to the forward microscopic one-way flux, the net flux is reduced

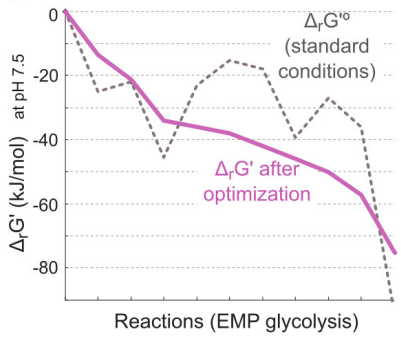
$$v = (1 - \mathrm{e}^{-\theta})\, v^+ \tag{9}$$

At given forward one-way flux: the higher the driving force, the higher the net flux!

Application: Max-min Driving Force (MDF) method [5]

Idea: Given the flux direction, choose metabolite levels such that
the *lowest thermodynamic force* among all reactions (θ_l closest to 0) is still as high as possible)





Noor *et al.* (2014) [5]

Purpose of the method

- Predict *in vivo* metabolite concentrations and thermodynamic driving forces *without* kinetics
- Find distributed bottlenecks (where enzyme efficiency is likely to be low)

Necessary input data

- Stoichiometric network
- Known flux directions
- Known equilibrium constants (or standard reaction Gibbs free energies)
- Physiological ranges for metabolite concentrations

Calculator for thermodynamics data (equilibrium constants and driving forces): equilibrator.weizmann.ac.il [6]

3 Enzyme cost minimisation

Thermodynamic force and enzyme demand

- In the MDF method, we assumed: weak thermodynamic forces make enzymes inefficient – weak thermodynamic forces must be compensated by high enzyme levels!
- Now we actually try to directly minimise the enzyme demand, and predict not only metabolite levels, but also enzyme levels.

3.1 Enzyme cost minimisation

How much enzyme is needed to support a given flux distribution? (ECM Method from Noor *et al.* 2016 [7])

Enzyme cost minimisation (ECM), basic idea

Consider a kinetic metabolic model with known fluxes

- The enzyme demand depends on metabolite levels

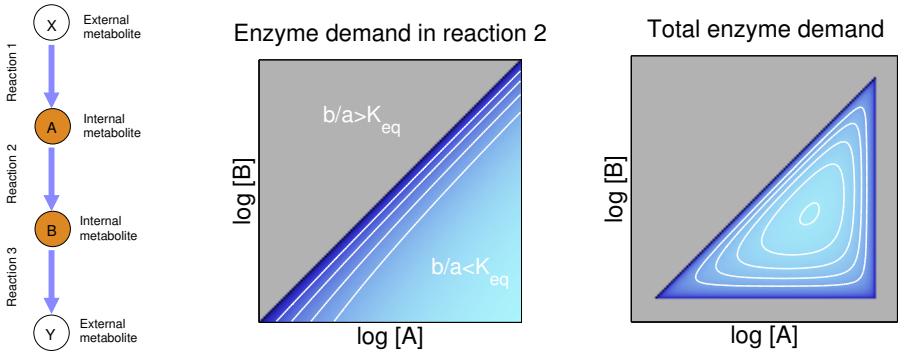
$$v = e\,\kappa(c) \Rightarrow e = \frac{v}{\kappa(c)}$$

- Poor kinetics or thermodynamics: efficiency κ becomes very small, $\Rightarrow$ high enzyme demand (to compensate low efficiency)!
- How much enzyme is needed *at least* to support the desired fluxes?
  $\Rightarrow$ Compromise between reactions! For example, the metabolite concentration profile

$$S \overset{1}{\rightarrow} HIGH \overset{2}{\rightarrow} LOW \overset{3}{\rightarrow} P$$

  implies a low demand in reaction 2, but a high demand in reactions 1 and 3.
  $\Rightarrow$ find optimal metabolite profile, allowing for a minimal enzyme demand!
- Reminder: flux directions and thermodynamics restrict the possible metabolite profiles; now we choose the *best possible* profile!

3.2 Enzyme cost minimisation: algorithm



Noor *el al.* (2016)

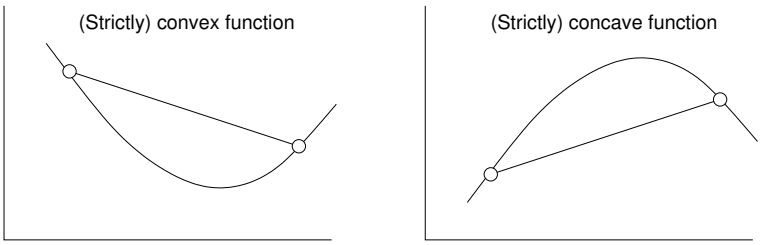
The polytope of feasible metabolite profiles is the metabolite polytope from thermodynamic FBA!
Enzyme demand at given fluxes:

$$e_{\mathrm{tot}} = e_1 + e_2 + e_3 = \frac{v_1}{\kappa_1(\mathbf{c})} + \frac{v_2}{\kappa_2(\mathbf{c})} + \frac{v_3}{\kappa_3(\mathbf{c})}$$

Similar formula for "enzymatic metabolite cost" $h(\mathbf{c}) = \sum_l h_l\, e_l(\mathbf{c})$ in metabolite space

Enzyme cost is a convex function on the metabolite polytope

Inverse catalytic rates $1/\kappa(\ln \mathbf{c})$ are convex functions (for any reversible rate laws)
$\Rightarrow$ enzyme levels and enzyme cost are convex functions, too



Why is convexity helpful?

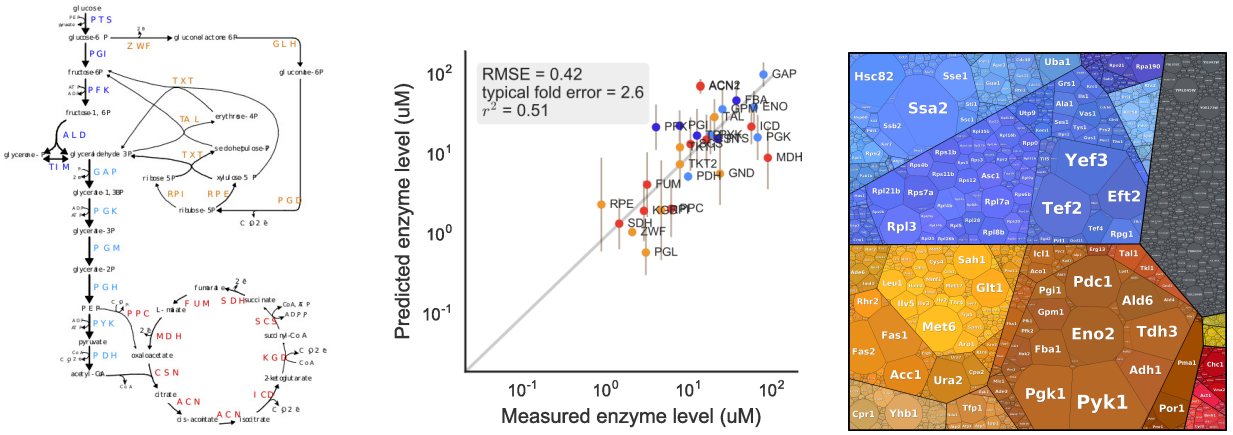
- Definition "Convex optimality problem": minimise a convex function on a convex set
- Numerically well solvable; no isolated local minima
- Solution in polynomial time
- Strict convexity (single optimum point) can be achieved by adding a regularisation term (metabolite side objective)
- Total metabolite concentration $\sum_i c_i = \sum_i \exp(\ln(c_i))$ is also convex. $\Rightarrow$ optimise the "kinetic const" (the sum of enzyme and metablite levels)

3.3 Optimal enzyme profile in Escherichia coli model

Modelling procedure for E. coli model

(i) Construct metabolic network
(ii) Determine kinetic constants (*in-vitro* literature values, completed and improved by parameter balancing [8, 9])
(iii) Use measured metabolic fluxes

$\Rightarrow$ Determine enzyme and metabolite levels, including thermodynamic forces, by ECM



Noor et al (2016) [7]

3.4 Enzyme cost, kinetics, and thermodynamics

Basic idea

- What is the relation between thermodynamic forces and enzyme demand?
- Enzyme investments are reciprocal to kinetic or thermodynamic efficiency!
- Inefficient enzyme need be be present in high amounts (for compensation)

Thermodynamic forces and enzyme levels

Separable formula for reversible rate laws (Noor *et al.* 2013 [10])
Example: reaction S $\leftrightarrow$ P with reversible Michaelis-Menten rate law

$$v = \frac{k_{\mathrm{cat}}^{+}\, s/K_{\mathrm{s}} - k_{\mathrm{cat}}^{-}\, p/K_{\mathrm{p}}}{1 + s/K_{\mathrm{s}} + p/K_{\mathrm{p}}} = k_{\mathrm{cat}}^{+} \underbrace{[1 - \underbrace{\frac{p}{s}\frac{k_{\mathrm{cat}}^{-}}{k_{+}}\frac{K_{\mathrm{s}}}{K_{\mathrm{p}}}}_{\mathrm{e}^{-\theta}}]}_{\eta^{\mathrm{rev}}} \underbrace{\frac{s/K_{\mathrm{s}}}{1 + s/K_{\mathrm{s}} + p/K_{\mathrm{p}}}}_{\eta^{\mathrm{sat}}} \tag{10}$$

- First term: catalytic constant k_{cat}^{+} (maximal enzyme efficiency)
- Efficiency terms η^{rev} and η^{sat} are between 0 and 1: they *decrease* the catalytic rate
- Omitting some terms yields simplified enzyme cost functions fewer parameters (lower bounds on actual cost!): dependence only on k_{cat}, or only on k_{cat} and driving force

Important prediction: apparent catalytic constants

- "Apparent catalytic constant", lower than k_{cat} value, can be predicted from first principles (if all model parameters are known).
- Confirms empirical observation: apparent catalytic rate decreases with lower substrate amount/quality, i.e., with growth rate

Enzyme cost scores

For ECM, we obtain the enzyme demand function

$$e(\mathbf{c}) = \frac{v}{\kappa(\mathbf{c})} = v\,\frac{1}{k_{\mathrm{cat}}}\,\underbrace{\frac{1}{\eta^{rev}(\mathbf{c})}}_{\frac{1}{1-\mathrm{e}^{-\theta}}}\,\frac{1}{\eta^{sat}(\mathbf{c})} \tag{11}$$

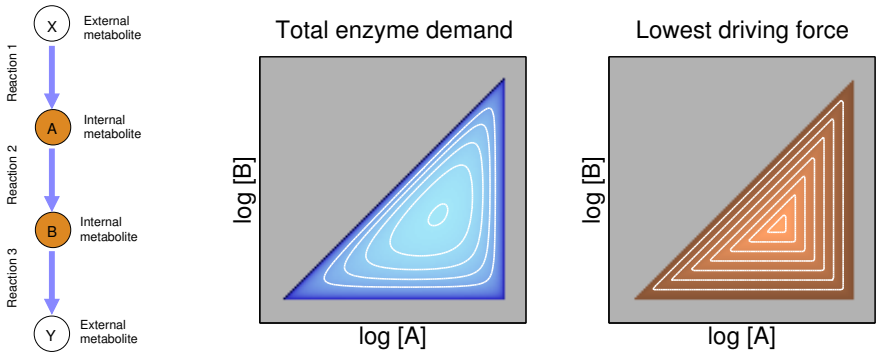
and the enzymate cost function in metabolite space

$$h(\ln \mathbf{c}) = \sum h_l \, e_l(\ln \mathbf{c}) = \sum_l \frac{h_l \, v_l}{\kappa_l(\mathbf{c})} = \sum_l h_l \, v_l \, \frac{1}{k_{cat,l}} \underbrace{\frac{1}{\eta_l^{rev}(\ln \mathbf{c})}}_{\frac{1}{1-e^{-\theta_l}}} \frac{1}{\eta_l^{sat}(\ln \mathbf{c})} \tag{12}$$

The formula shows the quantitative effect of thermodynamic forces on enzyme demand.
Again, the formula can be simplified by omitting some terms!

Comparison to MDF method

Eq. (11) shows that small driving forces lead to high enzyme cost: η^{rev} goes to infinity as a reaction approaches equilibrium (polytope boundary); outside the polytope there is no feasible solution.



Noor *et al.* 2016

MDF is a good proxy method! It avoids small driving forces, i.e., large enzyme cost!

Importance of ECM for enzyme economy of the cell?

- Predict quantitative relation between quantitative fluxes and protein levels;
- Ratio $\kappa = v/e$ "enzyme efficiency", "catalytic rate", or "apparent catalytic constant"
- These quantities are critical constraint-based (FBA or whole-cell) resource allocation methods

4 Flux cost minimisation

4.1 Simultaneous optimisation of fluxes, metabolite levels, and enzyme levels

How to optimise fluxes, metabolite levels, and enzyme levels at the same time?

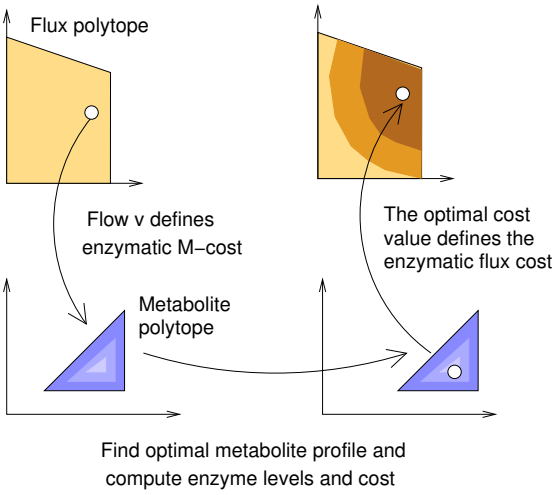
Basic idea (flux cost minimisation)

- For given flux distribution: ECM determines the minimal possible enzyme cost (optimised over all possible enzyme and metabolite profiles)
- Now: optimise the flux distribution to minimise this cost at a given flux benefit
- Same principle as in FBA with flux minimisation, but with a realistic, nonlinear flux cost function (describing the best-case enzyme cost)

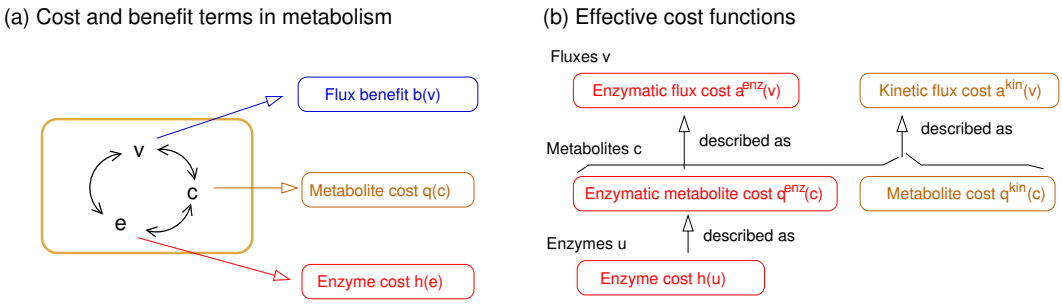
4.2 Flux cost minimisation

Enzyme cost as a flux cost function

- Instead of linear flux cost (in FBA), use *nonlinear* enzymatic cost derived from ECM
- This cost function accounts for correct kinetics and an optimisation of metabolite levels
- Metabolite cost can also be included

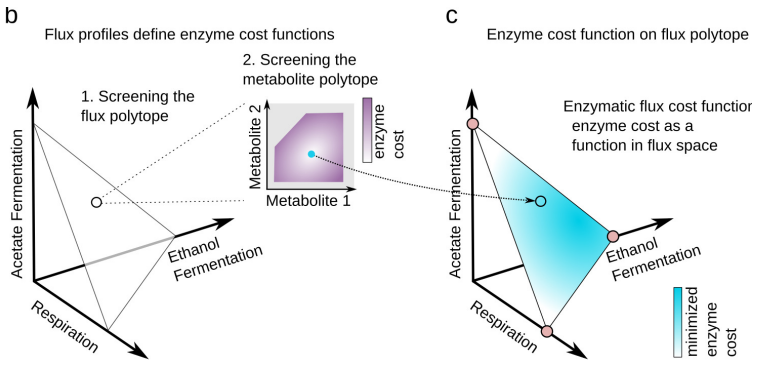


Enzyme cost as a function in enzyme, metabolite, and flux space



Flux cost minimisation: the optimal flux distribution is an elementary flux mode (EFM)

- An EFM is a "maximally sparse" stationary flux distribution:
  no reaction can be switched off without the stationary flux breaking down [11]
- The enzymatic flux cost function is concave on the flux polytope
  $\Rightarrow$ optimal points must be polytope vertices!
- Model with given flux directions and without flux bounds:
  the set of stationary flux distributions is a convex polytope, and all vertices are EFMs
- This result is equivalent to previous proofs in [12] and [13]



Wortel *et al.* (2018) [14]

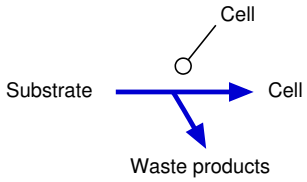
Flux cost minimisation in practice

- Using ECM, we compute the enzyme (maybe, plus metabolite) cost *at a given target flux* ("flux benefit").
- Using FCM, we can find the flux distribution (EFM) that minimises this cost.
- The method can be applied to metabolic network models with biomass production as the target flux.
- If not flux bounds are assumed: a *minimal enzyme cost per biomass production rate*

is equivalent to a *maximal biomass production rate per enzyme cost*.

4.3 Cell growth rate and cell biomass yield

Until here we considered metabolism (with enzymes as external variables) and the enzyme cost per target flux
How can we translate this into cell growth?



Two different whole-cell objectives

- **Biomass yield on substrate:** cell [biomass amount] / substrate consumed
- **Cell growth rate:** new cells / (time * cell) (or equivalently: biomass production rate per biomass amount)

.. and two optimality principles for metabolism:

- Substrate efficiency
- Enzyme efficiency

The principles for cells and for metabolism are analogous!

Consider metabolic model (biomass reaction as target reaction ⇒ biomass production rate = flux benefit)

- Substrate-efficient metabolism ⇒ High biomass yield
  Yield on substrate is a stoichiometric property! (ratios of amounts; time does not play a role)
- Enzyme-efficient metabolism ⇒ High growth rate
  Cell growth rate is a kinetic property (time and enzyme efficiency are involved)

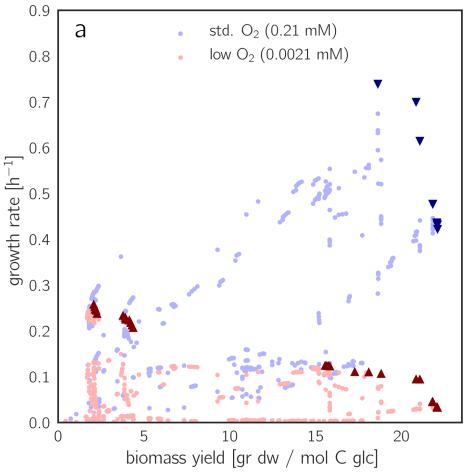
Simple assumption: assuming fixed total amount of metabolic enzyme in the cell (i.e. per biomass)

$$\frac{\text{BM rate}}{\text{enzyme cost}} \sim \frac{\text{biomass production rate}}{\text{biomass}} = \text{cell growth rate ?} \tag{13}$$

Rate/yield spectrum (for all EFMs) obtained from flux cost minimisation

- Compute achievable enzyme efficiency and biomass yield for all EFMs
- Translate the result into cell growth rates and cell biomass yield

Rate/yield spectrum for "standard" extracellular concentrations (blue) and lower oxygen level (red))



Wortel *et al.* (2018) [14]

The shape of the front determines whether there is a rate - yield tradeoff

Two possible cases:

- **Single optimal point** maximising rate and yield. A selection for growth rate and a selection for yield will result in the same strategy; a selection for growth rate will result in a yield-efficient strategy ("win-win" situation in cell population)
- **Extended Pareto front.** Growth rate and yield are maximised by different strategies. A selection for growth and a selection for yield will result in different strategies; a selection for growth rate will result in a yield-inefficient strategy ("tragedy of the commons" situation in cell population)

Example case: fermentation and respiration

- Rate/yield trade-offs and metabolic strategies
- Some examples: Catabolite repression; respiration vs fermentation; EMP vs ED glycolysis
- In each case, choice between different fluxes, protein levels, growth rates.

Which of the two objectives is biologically relevant?
Or in practice: What strategies (high-yield or low yield) should we expect in organisms "optimised by evolution"?
both strategies are observed, under different typical circumstances

- High oxygen: respiration *vs* no oxygen: fermentation
- High glucose: fermentation *vs* low glucose: respiration
- Normal body cells: respiration *vs* cancer cells: fermentation

Often respiro-fermentation is observed instead of pure fermentation, but the question remains:
why not simply respiration?

Rate/yield trade-offs

Let us assume a high yield as a "default" efficient behaviour.
Why (and in what cases) does nature "prefer" a lower yield?

- Assume that strategies follow from natural selection and are "optimal" in some sense.
- Why would it be optimal to waste carbon?
- Under what general conditions will low-yield strategies be preferred?

Distinguish two questions Molenaar *et al.* (2009) [15]

1. What is selected for, fast growth or high yield?
2. If growth is selected for, is growth maximised by high-yield or low-yield strategy?

But: is growth maximised by high-yield or low-yield strategies?

Two contrary claims:

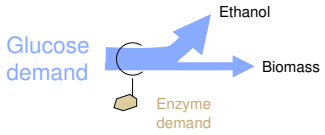
- A high yield (.. at a given glucose influx) leads to a *high* growth rate.
  This holds by definition! And it means: a waste of substrate leads to slower growth!
- A very high yield (bringing the system close to chemical equilibrium) leads to a *low* growth rate
  This means: a waste of substrate leads to faster growth!
  With such as "rate/yield trade-off", a competition for speed would automatically lead to a "non-sustainable" waste of carbon!

The rate-yield spectrum shows us that there is no general relationship between growth rate and yield
$\Rightarrow$ it clearly depends on circumstances!

Metabolic strategy depends on glucose concentration in the medium

Enzyme cost and choice of metabolic strategies depend on glucose concentration

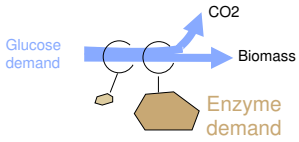
(a) Low–yield (fermentation) strategy



Substrate demand high –> Low biomass yield

Enzyme demand low –> High growth rate

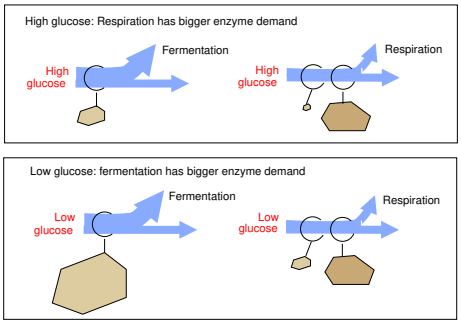
(b) High–yield (respiration) strategy



Substrate demand low –> High biomass yield

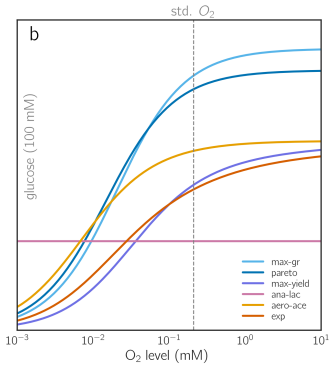
Enzyme demand high –> Low growth rate

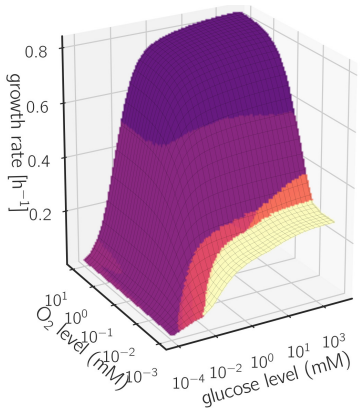
(c) Enzyme costs at high and low glucose concentrations



Parameter screens and Monod landscape

Definition "Monod curve":
Steady growth rate as a function of substrate (e.g. glucose) concentration in the growth medium





Wortel *et al.* (2018) [14]

Calculation by Flux Cost Minimisation

- Monod curve: screen an external substrate concentration (e.g. glucose or oxygen); Rate/yield spectrum may change, and trade-offs may emerge or disappear!
- "Monod landscape": growth rate as a function of glucose and oxygen level

5 Proteome partitioning and bacterial growth laws

5.1 Proteome partitioning and cell growth

Open question from metabolic models

- How can we translate biomass/enzyme efficiency into growth rate more precisely?
- Consider the varying fraction of metabolic enzymes within the proteome!
- Explanation by resource allocation (competition for resources such as space);
  similar to resource allocation between enzymes, but between total metabolic enzyme and ribosomes!

Exercise: An upper bound on cell growth rates

- Roughly estimate the maximal growth rate of a minimal cell
- Assume that *at least* ribosomes must replicate ribosomes.
- Why is it that real cells do not achieve this growth rate?

Proteome partitioning: some guiding question

High overall enzyme efficiency allows for higher cell growth rates!

- If we can compute th eoverall enzyme effiency, how can we compute the cell growth rate?

- Aim: a simple conversion formula (accounting for variable proteome partitioning)
- Constant metabolic fraction in the proteome ⇒ linear formula
- Variable metabolic fraction in the proteome ⇒ nonlinear formula

The idea of proteome partitioning yields simple linear or saturable conversion formulae.

5.2 Proteome partitioning: basic assumptions

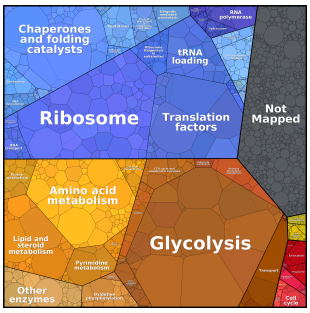
Ressource allocation in entire cell: how large are the large fractions of the proteome?

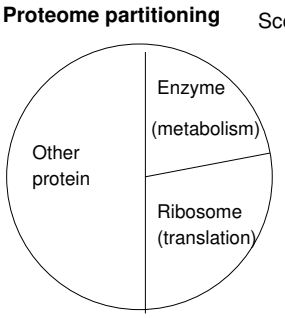
Basic idea (for varying growth rates caused by varying enzyme efficiency):

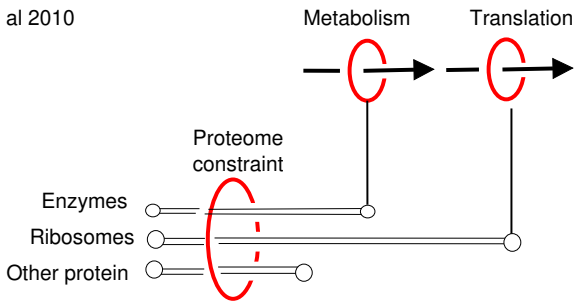
- Efficient metabolic enzymes are needed in lower amounts (at a given biomass production rate)
- Then more of the "protein budget" can be allocated to ribosomes, to speed up protein production
- Both speed-ups together allow for faster growth

But didn't we say "at a given biomass production rate" – which depends itself on the growth rate?
In fact, enzyme efficiency describes the enzyme demand *per biomass production*; so metabolism and ribosomes can "negotiate" the highest growth rate at which they can still work.

Translate this "negotiation" into a mathematical model and solve for the maximal growth rate!







Proteome partitioning model Scott *et al.* 2010 [16]
Basic assumptions

- Metabolism produces precursors, ribosomes produce proteins from precursors (other processes are neglected)
- Growth is limited by precursor production and by protein production
- For simplicity: all three are proportional, with given prefactors!
- Production rates are proportional to machine concentrations: enzymes and ribosomes work at fixed catalytic rates ("efficiencies")
- There is a common maximal "protein budget" for metabolic enzymes, ribosomal proteins, and other proteins (due to space restrictions?)
- "Other proteins" fraction is assumed to be constant

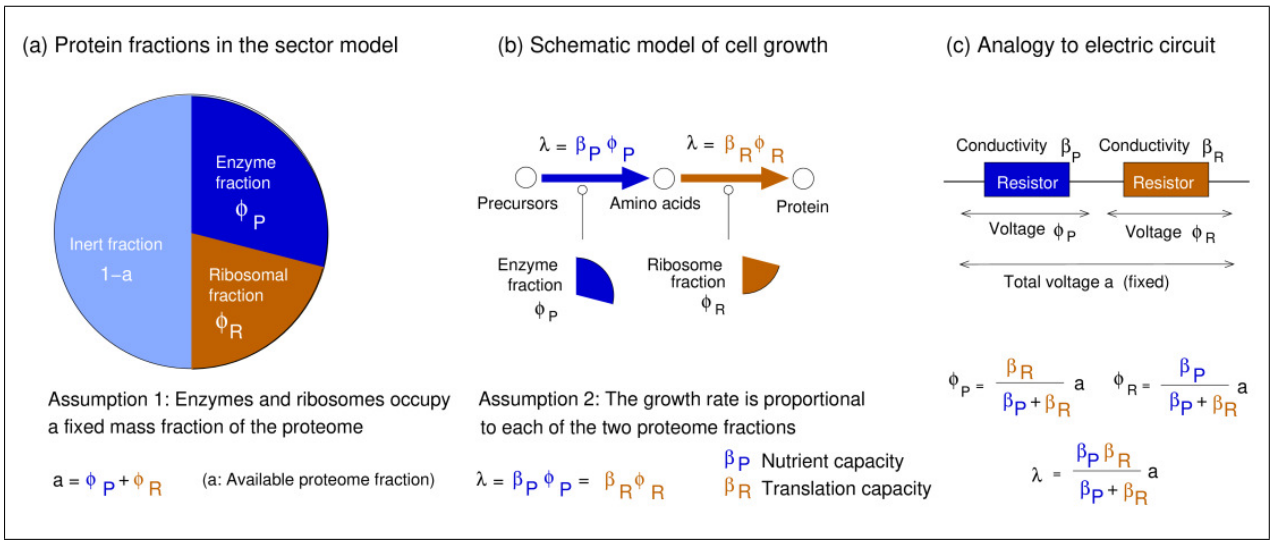
5.3 Proteome partitioning: model and bacterial growth laws

Observations:

- Cell growth rate depends on carbon sources / ribosome inhibition
- Variable proportions between ribosomes and metabolic enzymes

In particular:

- Growth changes due to poor carbon source: slow growth, *higher* investment in metabolic enzymes
- Growth changes due to ribosome inhibition: slow growth, *lower* investment in metabolic enzymes



Constraint-based proteome partitioning model

With appropriate units:

$$\lambda = \underbrace{v_{\text{met}}}_{\sim \Phi_{\text{met}}} = \underbrace{v_{\text{rib}}}_{\sim \Phi_{\text{rib}}} \tag{14}$$

Linear relationship between fluxes and "machines"' proteome fractions:

$$\lambda = \beta_{\text{met}} \, \Phi_{\text{met}} \quad = \quad \beta_{\text{rib}} \, \Phi_{\text{rib}} \tag{15}$$

The parameter β_{met} is the overall biomass/enzyme efficiency (to be computed e.g. from ECM).

The "variable proteome fractions" $\Phi_{\text{met}} + \Phi_{\text{rib}}$ are constrained to a fixed sum:

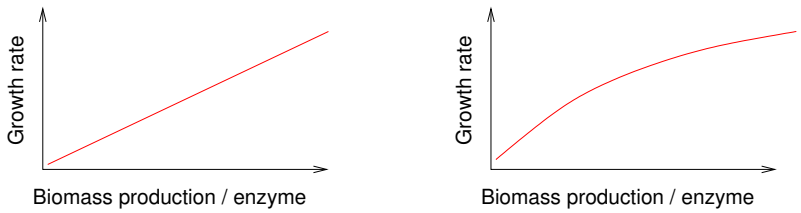
$$\Phi_{\text{met}} + \Phi_{\text{rib}} \quad = \quad \Phi_{\text{fix}} \tag{16}$$

Solving for the cell growth rate (trick: analogy to electric circuit – β as conductivities!)

$$\Rightarrow \lambda = \frac{1}{\frac{1}{\beta_{\text{met}}} + \frac{1}{\beta_{\text{rib}}}} \Phi_{\text{fix}} = \frac{\beta_{\text{met}} \, \beta_{\text{rib}}}{\beta_{\text{met}} + \beta_{\text{rib}}} \Phi_{\text{fix}} \tag{17}$$

yield our conversion from metabolic efficiency β_{met} to growth rate λ!

Linear and nonlinear efficiency → growth rate conversion



Proteome partitioning ("bacterial growth laws")

$$\Phi_{\text{met}} = \frac{\lambda}{\beta_{\text{met}}} = \frac{\beta_{\text{rib}}}{\beta_{\text{met}} + \beta_{\text{rib}}} \Phi_{\text{fix}} \tag{18}$$

Maximisation of growth rate is hidden in the formulae

"Optimality model" (with inequality constraints) *vs* "fully constrained model" (with equality constraints)

- In the proteome partitioning approach, there was no optimisation – the solution was completely determined

by equalities!
- Hidden optimality assumption: instead of describing the capacity constraints, density constraints, and relation between production rates and cell growth rate as equalities, we might describe them by inequalities and require, in addition, a maximal growth rate!
- On the contrary, since we know that in the optimal state, all these inequality constraints will be active, we can replcae them by equality constraints and (formally) omit the optimality condition.

Protein partitioning has been applied to models with more proteome fractions (variable "unused" protein and several metabolic pathways)

6 What limits cell growth?

From what have we learned so far, how is cell growth limited?
There is no "local reason": the reasons are global and distributed!

Resource allocation depends on the interplay of many constraints!

- Possible states are defined by constraints:
  (i) stationarity, including dilution
  (ii) kinetics/capacity constraints
  (iii) density constraints
  (iv) other physiological bounds ("targets")
- Select an optimal state by an assumed objective (e.g. maximal growth; protein production; in chemostat, growth at a given growth rate and at minimal external substrate level)
- A change in one cell component may lead to small perturbations or global rearrangements (perturbations at bifurcation points)
- Relaxing one constraint can lead to changes all over the place!

Science fiction exercise

Imagine a hypothetical "superbug" cell that can violate a basic law of physics.

Speculate about how it may use this "superpower" (and further adapt to it) to improve its selection advantage[1].

- Choose one fundamental law of physics to be violated (inside the cell, outside the cell, or both), and imagine how the cell would optimally adapts (in its "natural environment", to be specified).
- Imagine how such an organism could function, what other restrictions it could bypass, and what other adaptations it would accumulate in its further evolution towards maximal fitness.
- What (other genetic) changes are likely to occur? Are there any existing adaptations (in other environments, or by any "tricks" of cells) that resemble this hypothetical adaptation?

[1] If you like it more extreme: instead of physics, consider a basic law of mathematics or logics.

7 Resource Balance Analysis

Can we compute growth maximisation in a kinetic, genome-scale whole-cell model?

- Huge, non-convex problem – probably impossible to find the global optimum!
- But: a linearisation (replacing kinetics by apparent catalytic rates) can lead to tractable model!

Now we consider constraint-based, growth-optimising, genome-scale whole-cell models!!

"How large are the individual enzyme levels in the proteome"?!

Constrained-allocation FBA (CAFBA): Mori *et al.* 2016 [18]

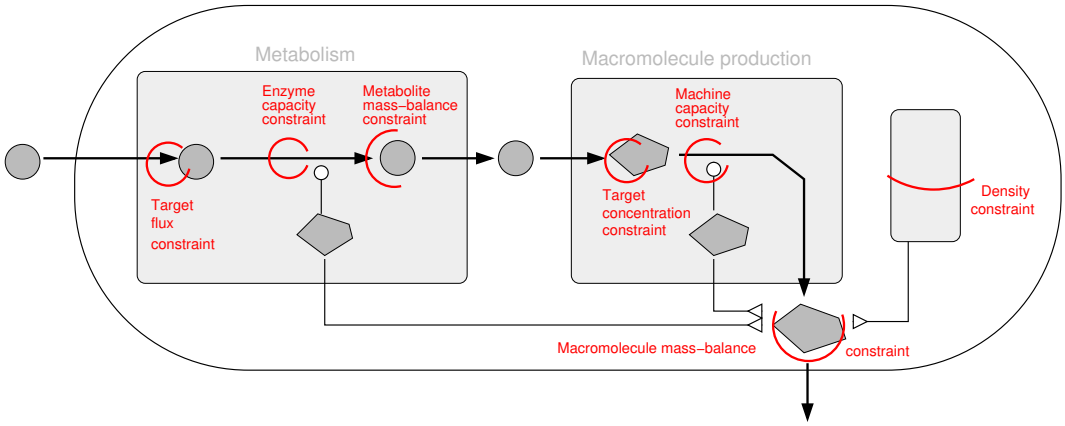
- Proteome partitioning model around a full (FBA-like) metabolic network model
- Instead of considering a few metabolic pathways as in Basan et al, include the ENTIRE metabolic network
- CAFBA can be seen as FBA with molecular crowding, but with a distinction between transporters, enzymes, and ribosomes; assumes (kinetic) relationship between external substrate and transporter demand (as in satFBA); and employs empirical linear growth laws for ribosomes.

Resource Balance Analysis (RBA) [19, 20]

RBA extends the proteome partitioning idea beyond CAFBA [21, 19]

- RBA describes optimal resource allocation in constraint-based, genome-scale whole-cell models
- Detailed model of metabolic reactions and all (potentially relevant) cell processes with major protein investments (metabolism, transport, ribosomes, chaperones)
- It describes metabolic fluxes, macromolecule production and dilution, and macromolecule concentrations.
- It includes knowledge of gene sequences, protein complexes
- It does *not* assume a given biomass composition, but *determines* this composition by necessities in the cell state in question!
- Physiological constraints (e.g., minimial amount of DNA, etc)

Replace all "local objectives" (e.g. ATP production, saving enzyme) by structural constraints!



RBA is based on four sets of constraints

- **Mass conservation** for compounds
  due to chemical reactions (metabolic reactions, protein synthesis, dilution)
  Steady state (with dilution for macromolecules: $v_{dil} = \lambda\, c \Rightarrow c = \frac{v_{dil}}{\lambda}$)
  $\Rightarrow$ basic FBA constraints and dilution
- **Capacity constraints** for reactions / cell processes:
  Biochemical reactions must be catalysed by cellular machines (enzymes, ribosomes, chaperones, ..)
  $\Rightarrow$ leads to an "implicit price" for each metabolic pathway.
- **Maximal density** for space in cell compartments:
  Each compartment can only contain a limited amount of molecules (maximal total density!) $\Rightarrow$ selection of most parsimonious pathways.

- **Target values** for concentrations and fluxes defining a viable cell

Protein levels result from a compromise:

- Capacity constraints (e.g., between metabolic enzyme and catalysed flux): favours *high* protein levels
- Density constraints (e.g., proteins in cytosol or on membranes): favour *low* protein levels

RBA: formulae for optimisation of growth rate and calculation of optimal state

"Inner loop": feasible (or optimal) cell state at a predefined growth rate
For each growth rate λ, solve the non-smooth linear feasibility problem (or linear programming problem)

"Outer loop": maximise the growth rate

- Fix a growth rate and decide: Can a steady (growth) state be maintained? $\Rightarrow$ linear (i.e., FBA-like) problem
- Repeat this many times; find the maximal growth rate at which the problem can be solved

There is a maximal growth rate, where the set of feasible states collapses (the volume becomes zero)
$\Rightarrow$ Calculation by a serial dichotomy

Input data required for building an RBA model

- Metabolic network
- Mapping from reactions to enzymes and further to proteins
- Protein amino acid sequences and cofactors (e.g. metal ions)
- Apparent catalytic constants of metabolic enzymes
- Types of cellular machines (e.g. ribosomes and chaperones), their apparent catalytic constants and cofactor demands
- Molecule sizes and localisations (for density constraints)
- Other constraints on concentrations or fluxes ("targets") defining a viable cell

Advantages and limitations of RBA

Main advantages

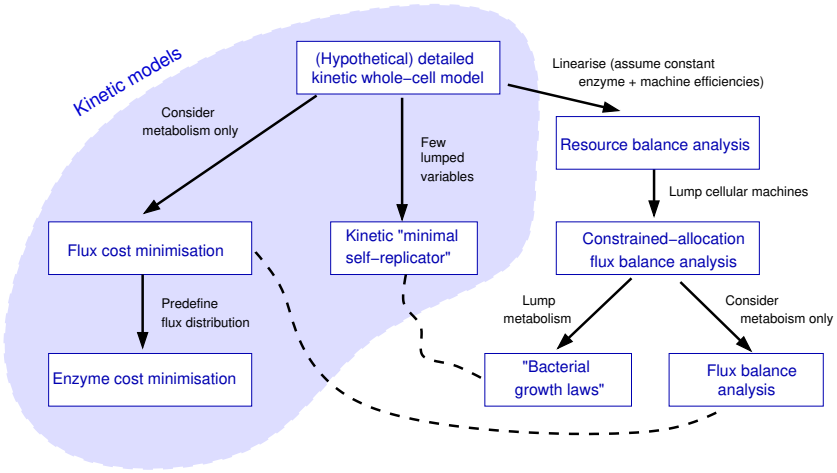
- RBA problem proven to be convex
- Series of linear optimality problems ("dichotomy search") is tractable for large models
- Predictions of complex adaptations (e.g. some proteins contain iron; under iron starvation, the cell may *increase* the import of iron, but also *avoid* using proteins that contain iron, and the pathways in whihc they operate).
  The second effect cannot be captured by FBA (which assumes a fixed biomass composition)

Main limitation (and theoretical problem): apparent catalytic rates

- Kinetics and thermodynamics are ignored (to obtain a linear problem)
- Empirical formulae for k_{app} are used
- Assumption: $v_{\mathrm{met}} = e\, k_{\mathrm{app}}$, where $k_{\mathrm{app}} = \kappa(\mathbf{c}) = k_{\mathrm{cat}}^{+} \cdot \eta^{rev}(\mathbf{c}) \cdot \eta^{sat}(\mathbf{c}) < k_{\mathrm{cat}}$
- *In vivo* k_{cat} values are unknown; in reality apparent k_{cat} values are variable (and depend on the unknown metabolic state)
- Determine empirical (growth-rate dependent) for k_{app} by comparing v (from preliminary FBA) and measured protein abundances

8 Summary of resource allocation models and enzyme economy

Overview of (kinetic and constraint-based, complex and simple) resource allocation models



Again: what limits cell growth?

- In RBA models, many constraints are hit at the same time.
- Instead of asking what (single) process limits growth? Rather ask: which of the constraints are hit?
- Relaxing any of these constraints may lead to faster growth

9 Points for discussion

1. What did we learn about modelling growth?

- What is missing in cell growth models so far?
- Growth rate as a parameter, or as an outcome?
- What are the factors (inside and outside the cell) that limit growth?

2. Open questions / problems in cell modeling

- Biomass composition
- Growth-dependent quantities
- Steady states?
- Varying environment and robustness
- Unknown environment

3. Is optimal behaviour a good assumption about cells?

- Choice of the objective function
- Is strict minimisation of enzyme levels a realistic assumption?
- Is multi-objective (Pareto) optimality a better concept?
- Dependence on the evolution scenario
- Higher levels of organisation
- Evolution may lead to non-optimality!
- What does "optimality" mean anyway?

4. Political questions

- Understanding growth and the limits of growth
- Economic metaphors in biology
- Ideological use of economic metaphors in biology; ideological use of biological metaphors in society

References

[1] E. Schrödinger. *What is life?* Cambridge University Press, 1944.

[2] N. Nagaraj, N.A. Kulak, J. Cox, N. Neuhauser, K. Mayr, O. Hoerning, O. Vorm, and M. Mann. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra hplc runs on a bench top orbitrap. *Mol Cell Proteomics*, 11:M111.013722, 2012.

[3] H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271(14):2905–2922, 2004.

[4] Q.K. Beg, A. Vazquez, J. Ernst, M.A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z.N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *PNAS*, 104(31):12663–12668, 2007.

[5] E. Noor, A. Bar-Even, A. Flamholz, E. Reznik, W. Liebermeister, and R. Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Computational Biology*, 10:e100348, 2014.

[6] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo. eQuilibrator – the biochemical thermodynamics calculator. *Nucleic Acids Research*, 40(D1):D770–D775, 2012.

[7] E. Noor, A. Flamholz, A. Bar-Even, D. Davidi, R. Milo, and W. Liebermeister. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Computational Biology*, 12(10):e1005167, 2016.

[8] T. Lubitz, M. Schulz, E. Klipp, and W. Liebermeister. Parameter balancing for kinetic models of cell metabolism. *J. Phys. Chem. B*, 114(49):16298–16303, 2010.

[9] T. Lubitz and W. Liebermeister. Parameter balancing: consistent parameter sets for kinetic metabolic models. *Bioinformatics*, 35:3857–3858, 2019.

[10] E. Noor, A. Flamholz, W. Liebermeister, A. Bar-Even, and R. Milo. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Letters*, 587(17):2772–2777, 2013.

[11] S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2):53–60, 1999.

[12] S. Müller, G. Regensburger, and R. Steuer. Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *Journal of Theoretical Biology*, 347:182–190, 2014.

[13] M.T. Wortel, H. Peters, J. Hulshof, B. Teusink, and F.J. Bruggeman. Metabolic states with maximal specific rate carry flux through an elementary flux mode. *FEBS Journal*, 281(6):1547–1555, 2014.

[14] M.T. Wortel, E. Noor, M. Ferris, F.J. Bruggeman, and W. Liebermeister. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Computational Biology*, 14(2):e1006010, 2018.

[15] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5:323, 2009.

[16] M. Scott, C.W. Gunderson, E.M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330:1099, 2010.

[17] M. Basan, S. Hui, H. Okano, Z. Zhang, Y Shen, J.R. Williamson, and T. Hwa. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature*, 528:99, 2015.

[18] M. Mori, T. Hwa, O.C.S. Martin, A. De Martino, and E. Marinari. Constrained allocation flux balance analysis. *PLoS Computational Biology*, 12(6):e1004913, 2016.

[19] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta*, 1810(10):978–988, 2011.

[20] A. Goelzer, J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nölker, M. Mariadassou, S. Aymerich, M. Hecker, P. Noirot, D. Becher, and V. Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, 2015.

[21] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47:1210–1218, 2011.

[22] R. Milo and R. Phillips. *Cell biology by the numbers*. Garland Science, 2015.

[23] E. Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems Biology - A Textbook. Second edition*. Wiley-VCH, 2015.