Cellular Molecular Life Sciences manuscript No.
(will be inserted by the editor)

The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints

Daan H. de Groot · Julia Lischke · Riccardo Muolo · Robert Planqué · Frank J. Bruggeman · Bas Teusink

Received: date / Accepted: date

Abstract Living cells can express different metabolic pathways that support growth. The criteria that determine which pathways are selected in which environment remain unclear. One recurrent selection is overflow metabolism: the simultaneous usage of an ATP-efficient and -inefficient pathway, shown for example in *Escherichia coli*, *Saccharomyces cerevisiae* and cancer cells. Many models, based on different assumptions, can reproduce this observation. Therefore, they provide no conclusive evidence which mechanism is causing overflow metabolism. We compare the mathematical structure of these models. Although ranging from Flux Balance Analyses to self-fabricating Metabolism and Expression models, we can rewrite all models into one standard form. We conclude that all models predict overflow metabolism when two, model-specific, growth-limiting constraints are hit. This is consistent with recent theory. Thus, identifying these two constraints is essential for understanding overflow metabolism. We list all imposed constraints by these models, so that they can hopefully be tested in future experiments.

Keywords Overflow metabolism · Growth rate maximization · Genome-scale modeling · Metabolism and Expression · Elementary Flux Modes · Elementary Growth Modes

DHdG and BT acknowledge NWO VICI grant 865.14.005, JL thanks EraCoBiotech grant YogurtDesign (NWO grant 053.80.733) and RM acknowledges ITN SynCrop (grant agreement No 764591) for financial support

Daan H. de Groot
Systems Bioinformatics, AIMMS, Vrije Universiteit Amsterdam, NL-1081HZ
E-mail: d.h.de.groot@vu.nl

Julia Lischke
Systems Bioinformatics, AIMMS, Vrije Universiteit Amsterdam, NL-1081HZ

Riccardo Muolo
Systems Bioinformatics, AIMMS, Vrije Universiteit Amsterdam, NL-1081HZ

Robert Planque
Department of Mathematics, Vrije Universiteit Amsterdam, NL-1081HV

Frank J. Bruggeman
Systems Bioinformatics, AIMMS, Vrije Universiteit Amsterdam, NL-1081HZ

Bas Teusink
Systems Bioinformatics, AIMMS, Vrije Universiteit Amsterdam, NL-1081HZ
E-mail: b.teusink@vu.nl

1 Introduction

Many cells show overflow metabolism: the simultaneous metabolism of nutrients by an energy-efficient and a less energy-efficient pathway. For example, *Escherichia coli*, *Saccharomyces cerevisiae* and cancer cells fully oxidize carbon sources to CO_2 when growing slowly. Above a species-specific critical growth rate, a partial oxidation pathway kicks in, resulting in the production of overflow metabolites: acetate, ethanol and lactate respectively [1–3]. *Lactococcus lactis* shows a similar metabolic shift from mixed-acid fermentation (3 ATP per glucose) to lactic-acid fermentation (2 ATP per glucose) under anaerobic conditions [4]. Besides overflow metabolism that starts at high growth rates, *Escherichia coli* even produces overflow products at low growth rates when growing in ammonium-limited conditions [5].

Overflow metabolism seems wasteful because two metabolic pathways are used that independently sup-

port growth, and one of them is more efficient (it has a higher ATP yield per glucose molecule) than the other. Since cells need energy for growth, efficient usage of nutrients is expected to be favourable. One would therefore expect that cells using the efficient growth strategy exclusively would be selected during evolution.

The counterintuitive occurrence of overflow metabolism is in many studies explained using constraint-based optimization approaches. These approaches assume that cellular growth is constrained by physical and chemical limits, and that cells are driven towards these limits when evolutionary fitness is maximized. Accordingly, the behavior of cells results from maximizing their growth rate given a set of constraints.

Since many models reproduce the experimental data while using different biological assumptions, it is unclear what exactly causes overflow metabolism. Therefore, we need a way to analyze and compare these different models to find the cause of overflow metabolism.

Minimal, growth-supporting metabolic modes are characterized mathematically by identifying the smallest subnetworks of the entire metabolic network that can support growth. Such subnetworks are called Elementary Flux Modes (EFMs) in metabolic models [6], and Elementary Growth Modes (EGMs) in self-fabrication models [7] (see SI1 for a short introduction and comparison of EFMs and EGMs). The gradual transition from the usage of one metabolic subnetwork to the mixed usage of two subnetworks that is observed in overflow metabolism indicates the simultaneous usage of two different Elementary Modes.

In recent theoretical work [7,8] we derived that cells that maximize their growth rate will only use two Elementary Modes if they are confronted with at least two constraints. The identification of these constraints is therefore an important step towards finding the mechanistic cause of overflow metabolism. In this review, we use this theory to compare the various models of overflow metabolism by making the growth-limiting constraints explicit.

Although the models range from relatively simple Flux Balance Analyses to genome-scale self-fabrication models, we will show that they can all be written in the same concise standard form. Thus, the models are highly similar: (a proxy for) the cellular growth rate is maximized subject to two constraints. However, the biological assumptions underlying the imposed constraints differ between those models. Hence, the success of these models is dependent on the existence of two constraints and not on the precise biological interpretation of those constraints. Finding the causes of overflow metabolism therefore amounts to identifying the two active growth-limiting constraints and experimentally testing them. We shall conclude that the models each offer a hypothesis that needs to be tested in falsification experiments in the future.

2 A standard form for overflow metabolism models

We will show that, to our knowledge, all existing models that use growth rate maximization to explain overflow metabolism, can be rewritten in a standard form.

We will assume that a cell adapts its state to grow as fast as possible whenever it encounters a new environment. The cellular state is specified by *optimization variables*, for example the reaction rates ($\boldsymbol{v}$) or the enzyme concentrations ($\boldsymbol{e}$). We will denote the optimization variables by the vector $\boldsymbol{x}$, the ith entry of which is denoted by x_i. The growth rate is modeled as a linear function: *the objective function*:

$$\text{objective}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i,$$

where w_i is the weighting factor of variable i. The growth rate maximization of the cell is modeled mathematically by searching for the set of optimization variables that maximizes the objective function, given constraints to be specified later. Because the objective function is linear, there is a certain direction in the space of optimization variables in which the objective always increases. The optimal solution is the set of optimization variables that is as far in that direction as possible.

Not all combinations of optimization variables can be chosen due to *constraints*, for example a limited uptake rate, or a limited available area for membrane proteins. These constraints are formalized by inequalities acting on a weighted sum of the variables:

$$\sum_{j=1}^{n} a_j x_j \leq b,$$

where a_j determines the 'cost' of increasing the jth variable. In the special case that $a_j = 0$, x_j is not bounded by this constraint. In general, we could have several, say m, constraints. These constraints can be collected in an $m \times n$ matrix A, where the ith row captures the ith constraint. All constraints can then be written together as:

$$A \cdot \boldsymbol{x} \leq \boldsymbol{b}.$$

The constraints can be viewed as planes bounding the feasible combinations of variables (see Figure 1). After all constraints have been implemented, we are left with

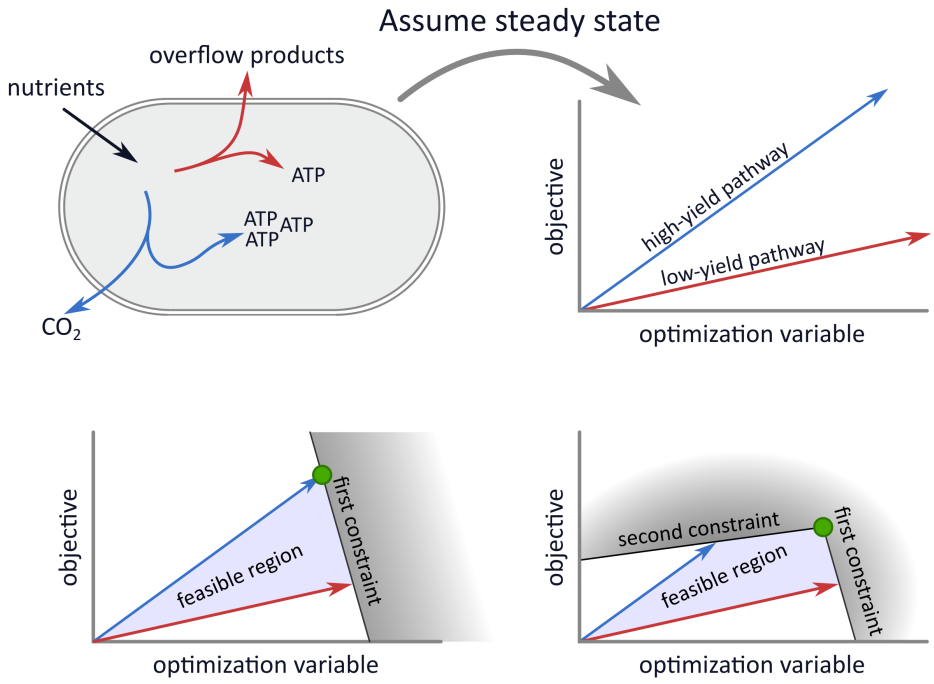

Fig. 1 A general view on overflow metabolism and how it is modeled. In general, overflow metabolism is the simultaneous usage of two independent growth-supporting subnetworks with different substrate yields. In the **top left** subfigure, the blue pathway produces more energy equivalents per gram nutrient than the red pathway. Together with the non-depicted rest of the metabolic network, the blue and red pathway can separately lead to steady state growth. In the **top right** subfigure, we illustrate that imposing *homogeneous constraints*, in this case a steady state assumption, gives rise to relations between optimization variables. The optimization variables can for example be reaction rates or enzyme concentrations, but for simplicity, we only show one variable here. The model objective is here visualized along the y-axis, so that the combination of variables that gives the highest y-coordinate is optimal. In the **bottom** figures, we add *inhomogeneous constraints* on the optimization variables. These affect which combination of variables is optimal. Under one constraint, exclusive usage of the high-yield pathway is optimal. Adding the second constraint leads to the optimality of a combination of the two pathways.

an angular space called *the solution space*. Solving the optimization problem amounts to selecting the point in this space that maximizes the objective function. It can be shown that there is always a corner point of the solution space (called vertex) in which this optimum is attained.

In this review, we will use these concepts to extract the mathematical cores of all overflow metabolism models (that we could find) and rewrite them in the following standard form.[1]

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \sum_{i=1}^{n} w_i x_i \\ \text{subject to} \quad & A \cdot \boldsymbol{x} \leq \boldsymbol{b} \\ & x_i \geq 0. \end{aligned} \tag{1}$$

The constraints in A can be *homogeneous* and *inhomogeneous*, see Figure 1 and its caption. A constraint is called homogeneous when the corresponding weighted sum of the variables equals zero, i.e., $b = 0$, and inhomogeneous otherwise. Examples of homogeneous constraints that we will define later in this review are the steady state constraint and irreversibility constraints. When all constraints of A are homogeneous, the solution space is unbounded; it can be visualized as an infinitely stretched angular cone. The optimization problem will in this case not have a finite maximum. Inhomogeneous constraints can make the cone bounded; this will be especially important in modeling overflow metabolism. We will therefore highlight them by presenting them separate from the rest of the constraints.

[1] We use a different standard form than the standard form that is used in Linear Programming. We found that our form better serves our purposes, but in SI2 we show that the two forms are equivalent.

3 Current explanations of mixed behaviour and their mathematical background

Next, we will discuss published models made to explain overflow metabolism that use growth rate maximization. We will start with the modeling approaches that

are the easiest to understand, and gradually build up complexity, ending with self-fabrication models.

3.1 Flux Balance Analysis models

Flux Balance Analysis studies the sets of reaction rates (fluxes) through a metabolic network (say with m metabolites and r reactions) that can reach a *steady state*. A steady state is attained when the net rate of production of each metabolite is equal to the net rate of its consumption. The stoichiometry of all reactions is described by the stoichiometric matrix, N, which has m rows and r columns. Each row corresponds to the mass balance of a metabolite and contains the stoichiometric coefficients of this metabolite in all reactions. When N is multiplied by the rate vector, we get the 'mass-balance constraints':

$$N \cdot \boldsymbol{v} = \mathbf{0}.$$

In this review we will consider all reactions to be irreversible; we can always split up a reversible reaction into one forward and one backward reaction, resulting in:

$$v_i \geq 0.$$

These steady state and irreversibility constraints are the homogeneous constraints. As mentioned before, the space of flux vectors that satisfy these constraints is unbounded.

In addition, flux bounds can be imposed. Upper bounds, denoted ub_i, are for example imposed to model a limited capacity of the cell for the corresponding reaction. Lower bounds, denoted lb_i, are for instance used to model the production of ATP for non-growth associated maintenance. This gives

$$lb_i \leq v_i \leq ub_i.$$

In FBA, we are mostly interested in those flux vectors, $\boldsymbol{v}$, that maximize some proxy for the growth rate. For this, the so-called *biomass reaction*, v_{BM} [9], is added to the model: a phenomenological reaction that produces all cellular compounds in the right proportions, and thereby approximates the demands for cell synthesis.

The full problem can now be written as

$$\begin{aligned} \underset{\boldsymbol{v}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \mathbf{0} \\ & v_i \geq 0 \\ & \\ & lb_i \leq v_i \leq ub_i, \end{aligned} \tag{2}$$

which is equivalent to the standard form that we introduced in Equation (1) (see the Supporting Information for the appropriate choice of $\boldsymbol{w}, A, \boldsymbol{x}, \boldsymbol{b}$).

FBA models have been used to explain overflow metabolism, mathematically capturing the reasoning of Andersen and von Meyenburg [10] in 1980: "If, however, respiration is limited, by-product formation can lead to extra ATP production and to faster growth, provided the by-product can be generated with a net gain of ATP." The imposed flux bounds differ between the models, although all models consider a limited uptake rate for the carbon source. For example, Majewski and Domach [11] further propose that *E. coli* might have a limited electron transfer capacity, while Varma and Palsson [12] assume that oxygen uptake is limited, and that a certain amount of ATP should be produced even if the cell is not growing. This leads to the following FBA problem:

$$\begin{aligned} \underset{\boldsymbol{v}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \mathbf{0} \\ & v_i \geq 0 \\ & \\ & v_{\mathrm{O_2,uptake}} \leq 15 \\ & v_{\mathrm{Glc,uptake}} \leq 10.5 \\ & v_{\mathrm{ATP,maintenance}} \geq 7.6. \end{aligned} \tag{3}$$

Carlson and Srienc [13,14] also model growth rate maximization under glucose- and oxygen-limitation, but take a different approach. Instead of finding only the optimal solution, they characterize the whole steady-state solution space by enumerating the EFMs of a coarse-grained *E. coli* network (see SI1 for an explanation about EFMs). Using their acquired knowledge of all possible solutions, the authors select four Elementary Flux Modes. Under any level of glucose and oxygen limitation, two of these EFMs together form the optimal solution. The simultaneous usage of these EFMs leads to overflow metabolism.

3.2 FBA models with thermodynamic constraints

FBA models can be refined by adding thermodynamic constraints [15, 16]. The laws of thermodynamics dictate that a chemical reaction can only have a positive rate if the summed Gibbs free energy of the reaction substrates is higher than of the reaction products, i.e., if the free energy change due to the reaction is negative, denoted by: $\Delta_r G' < 0$. By using this, one for example excludes cycles like $A \rightarrow B \rightarrow C \rightarrow A$ from carrying a positive flux, since such a cycle has zero thermodynamic driving force [17]. The free energy change

due to a reaction depends on the concentrations of the involved metabolites, but these are usually not modeled in FBA approaches. Most thermodynamic FBA approaches thus need some way to estimate either these metabolite concentrations, or the $\Delta_r G'$-values directly. There are also some methods where this estimation step can be avoided, at the expense of the thermodynamic constraints becoming less restrictive (see [16] for an overview of thermodynamic FBA methods).

Recently, Niebel et al. combined growth rate maximization and a thermodynamic constraint to describe overflow metabolism [18]. In their approach, the metabolite concentrations and the reaction rates are free variables, although the metabolite concentrations are provided with an upper and lower bound based on experimental measurements. The authors search for the optimal concentrations and rates so that the biomass production rate is maximized. This search is constrained by the second law of thermodynamics, implying that the free energy change induced by an active reaction should be negative: $\Delta_r G'_j(\boldsymbol{c}) < 0$ for reactions with $v_j > 0$. These homogeneous constraints take the place of the irreversibility constraints that were used in FBA models, where the directionality is now based on the sign of the Gibbs free energy change.

If we add up all these free energy changes induced by the chemical reactions, we get the total dissipated Gibbs energy per unit time: $g^{\text{diss}} = -\sum_j \Delta_r G'_j(\boldsymbol{c}) v_j$. The authors observed in experiments that this dissipation function appears to have a maximum at the onset of overflow metabolism. Therefore, they propose that the dissipated energy might be limited by an upper bound,

$$g^{\text{diss}} \leq g^{\text{diss}}_{\text{lim}}. \tag{4}$$

In addition, another constraint is imposed ensuring that the free energy dissipated by internal reactions equals the free energy that is extracted from external nutrients. However, after a careful examination of the mathematics used in [18] (see SI4), we believe that this constraint should be equivalent to the steady state assumption, so that we could ignore it here. If it turns out that we are wrong, the constraint can be added to the problem below, without affecting the conclusions of this review.

This modeling approach is no longer linear in the variables because the $\Delta_r G'_j(\boldsymbol{c})$-values can depend nonlinearly on the metabolite concentrations. However, for any fixed set of metabolite concentrations, $\boldsymbol{c} = \boldsymbol{c}_0$, the model reduces to a Linear Program that can be written in our standard form (see SI3.2 for the appropriate choice of the variables $\boldsymbol{w}, A, \boldsymbol{x}, \boldsymbol{b}$ used in (1)).

$$\begin{aligned} \underset{\boldsymbol{v}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \boldsymbol{0} \\ & v_j \geq 0 \\ & \Delta_r G'_j(\boldsymbol{c}) < 0 \quad \text{for reactions } v_j \neq 0 \\ & \\ & v_{\text{Glc,uptake}} \leq b_{\text{Glc}} \\ & -\sum_j \Delta_r G'_j(\boldsymbol{c}) v_j \leq g^{\text{diss}}_{\text{lim}}. \end{aligned} \tag{5}$$

3.3 Resource-allocation models

Reaction rates can almost always be increased by increasing the concentration of the catalyzing enzyme [19]. A constraint on a reaction rate can therefore not reflect the mechanistic cause of metabolic phenomena: if a cell would be confronted with such a constraint, the concentration of the corresponding enzyme could be increased, unless the enzyme concentration itself is constrained. In that case however, it is the constraint on enzyme concentrations that is the mechanistic cause.

In the past decade, many researchers shifted perspective by taking enzyme concentrations as the optimization variables instead of the reaction rates. These models are called *resource allocation* models [20–31].

Resource allocation models also maximize the biomass reaction rate, v_{BM}, while metabolism is at steady state: $N \cdot \boldsymbol{v} = \boldsymbol{0}$. The rate of the objective reaction can thus only be increased if the rates of all reactions in a complete growth-supporting subnetwork are increased. Unlike in FBA models however, each reaction rate is now coupled to the concentration of a catalyzing enzyme,

$$v_i = e_i k_{\text{cat},i} f_i(\boldsymbol{x}), \tag{6}$$

where v_i is the ith reaction rate and e_i is the concentration of the corresponding enzyme. The activity of an enzyme is determined by its catalytic rate $k_{\text{cat},i}$, and the 'saturation' of the enzyme $f_i(\boldsymbol{x})$ with its substrates $\boldsymbol{x}$. This saturation term is in reality a nonlinear function of the metabolite concentrations $\boldsymbol{x}$, that also includes product inhibition. However, we split reversible reactions, product inhibition is almost always ignored, and $f_i(\boldsymbol{x})$ is often simplified to be constant, such that $v_i = e_i k_{\text{cat},i}$ where $k_{\text{cat},i}$ is now an effective rate constant. The only way to increase the reaction rates is then to increase the enzyme concentrations. However, resources are limited: various limits on enzyme concentrations exist, which take the form of (weighted) sums that are bounded:

$$\sum_i c_i e_i \leq ub, \tag{7}$$

where c_i is a weighting factor, and should not be confused with the metabolite concentrations that were used earlier. All enzymes for which the weighting factor is nonzero, $c_i > 0$, contribute to the sum. These weighting factors can be adjusted to capture various constraints. For example, if the membrane area is constrained, the weight c_i would reflect the area taken up by one unit of protein i, and c_i would thus be zero for all non-membrane proteins. Since the sum is bounded, an increase in the concentration of protein i must be compensated by a decrease in the concentration of others. The available resources should thus be carefully allocated in order to maximize the biomass production rate. These approaches can be written in a form equivalent to our standard form (see SI3.3 for the appropriate choice of $\boldsymbol{w}, A, \boldsymbol{x}, \boldsymbol{b}$ in (1))

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{e}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \boldsymbol{0} \\ & v_i = e_i k_{\text{cat},i} \\ & e_i \geq 0 \\ & \sum_i c_i^1 e_i \leq ub_1 \\ & \vdots \\ & \sum_i c_i^n e_i \leq ub_n. \end{aligned} \tag{8}$$

With the help of Equation (6) this problem can be solved with the reaction rates or with the enzyme concentrations as the optimization variables, as we show in SI3.3.

Basan et al. [27] made a core model that shows overflow metabolism in *E. coli* by dividing the proteome into three fractions: φ_f, φ_r and φ_{BM}, that thus sum up to one. These denote the fractions of the proteome catalyzing a fermentation, respiration and cell synthesis reaction, respectively, according to the relations: $v_f = \epsilon_f \varphi_f$, $v_r = \epsilon_r \varphi_r$, and $v_{\text{BM}} = \frac{1}{b}(\varphi_{BM} - \varphi_0)$. We will not define all unknown symbols in these and the following relations, since their interpretation is not relevant for this review. Note that the relation between the biomass reaction and the associated proteome fraction is non-standard, to include a non-growth associated maintenance term. Further, reactions for the uptake of a carbon source and the excretion of acetate[2] are included, but these do not have an associated proteome fraction. This gives the following steady state assumption

$$N \cdot \boldsymbol{v} = \begin{array}{r} \text{carbon} \\ \text{energy} \\ \text{acetate} \end{array} \begin{bmatrix} 1 & -1 & -1 & 0 & -\beta \\ 0 & n_r & n_f & 0 & -\sigma \\ 0 & 0 & S_{ac} & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_{\text{uptake}} \\ v_r \\ v_f \\ v_{\text{excretion}} \\ v_{BM} \end{bmatrix} = \boldsymbol{0}.$$

In addition, the authors set the uptake rate of nutrient: $v_{\text{uptake}} = c_{\text{uptake}}$. Together, this yields a set of equations with only one solution; the variables ($\boldsymbol{v}$ and $\boldsymbol{\phi}$) can be directly calculated, and no optimization is required. The uniqueness of the solution is due to the small size of the model and because the constraints on uptake rate and the total proteome are modeled as equalities instead of inequalities. We show in SI3.4 that, for the appropriate choice of $\boldsymbol{w}, A, \boldsymbol{x}, \boldsymbol{b}$, this is equivalent to our standard form (1) in which the biomass production rate is maximized and the constraints are treated as inequalities:

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{\varphi}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \boldsymbol{0} \\ & v_{\text{f}} = \epsilon_f \varphi_f, \; v_{\text{r}} = \epsilon_r \varphi_r, \; v_{\text{BM}} = \frac{1}{b}(\varphi_{BM} - \varphi_0) \\ & v_i \geq 0 \\ & v_{\text{uptake}} \leq c_{\text{uptake}} \\ & \varphi_f + \varphi_r + \varphi_{BM} \leq 1. \end{aligned} \tag{9}$$

The authors assume that the yield of energy per carbon molecule is higher for respiration than for fermentation: $n_r > n_f$[3], but that fermentation is more proteome-efficient: $\epsilon_f > \epsilon_r$. The enzyme cost of a certain reaction is the protein fraction necessary to attain one unit flux. We thus see that the enzyme costs of respiration are higher than fermentation: $\frac{1}{\epsilon_f} < \frac{1}{\epsilon_r}$. Because of this trade-off between yield and enzyme costs, it becomes optimal from a certain critical rate of carbon uptake to use the respiration and fermentation reactions simultaneously, so that the model shows overflow metabolism.

Vazquez et al. [24] responded to the explanation of Basan et al. by adding to the model that there is a maximum to the macromolecular density of a cell. They argue that the enzyme costs, as defined in the previous paragraph, should be proportional to the enzyme mass divided by its catalytic rate. The model that is used to explain overflow metabolism is thus the same, but with a different mechanistic underpinning of the ϵ-parameters. This reasoning was implemented earlier by the same authors in a genome-scale formalism called

[2] The acetate excretion reaction was not explicitly mentioned in [27], but must have been included. We have made it explicit to be able to write a consistent stoichiometry matrix.

[3] The authors originally used e_r, e_f to denote these stoichiometric fractions, but we have renamed them to avoid confusion with enzyme concentrations.

FBA with Macromolecular Crowding (FBAwMC), with which they already explained overflow metabolism in *E. coli* [23]. This formalism was later also used to model *S. cerevisiae* [29].

Another hypothesis is offered by Zhuang et al. [28] in which overflow metabolism is explained using a membrane occupancy constraint. The authors introduce parameters, m_i, that capture the membrane area that is occupied per mol/liter of enzyme i. Assuming that there is only a limited plasma membrane budget, B_{cyt}, this introduces the constraint $\sum_i m_i e_i \leq B_{\text{cyt}}$. Together with the steady state assumption, and a limited glucose uptake rate, this gives:

$$\begin{aligned} \underset{\boldsymbol{v},\boldsymbol{e}}{\text{maximize}} \quad & v_{BM} \\ \text{subject to} \quad & N \cdot \boldsymbol{v} = \boldsymbol{0} \\ & v_i = e_i k_{\text{cat},i} \\ & e_i \geq 0 \\ \\ & v_{\text{uptake}} \leq c_{\text{uptake}} \\ & \sum_i m_i e_i \leq B_{\text{cyt}}. \end{aligned} \tag{10}$$

This hypothesis is supported by some quantitative evidence collected by Szenk et al. [22].

Note that the above resource allocation approaches differ in the mechanistic nature of the last constraint that is added, but that the optimization function, the steady state assumption and the limited substrate uptake are all similar. Shlomi et al. used a similar approach with a total proteome constraint to describe the Warburg effect in cancer cells [31].

A modeling approach that should be set slightly apart is Constrained Allocation FBA (CAFBA) [30]. The authors use only a total proteome constraint, and no direct limit on substrate uptake. Instead, substrate limitation is modeled by increasing a parameter w_C that captures the protein fraction needed for carbon catabolism to sustain one unit of carbon influx: $\phi_C = \phi_{C,0} + w_C v_C$. If the concentration of external nutrient decreases, w_C increases, and a larger fraction of the proteome is thus needed in the carbon catabolic sector. Because the sum of the proteome fractions needs to be one, this reduces the available proteome fraction for other sectors. As such, a change in nutrient concentration leads to a re-allocation of the proteome. The genome-scale model of *E. coli* can reproduce a switch from pure respiration to acetate secretion, but it does so with small discrete jumps. The gradual switch that is usually associated with overflow metabolism can only be found when the results are averaged over many different models created by choosing random parameters.

3.4 Self-fabrication models

In the previously described modeling approaches, the demand for cell components was approximated using a virtual biomass reaction. However, this approximation ignores an important nonlinear aspect of self-fabrication. A self-fabricating cell should produce two daughters identical to itself. The proportions in which the cell should produce cellular components thus depend on its own interior. If the cell reallocates resources to meet this demand for cellular components, its interior changes and therefore also the demand. The allocation of resources thus both depends on, and determines, the demand reaction.

Another inherent nonlinearity of cellular growth arises because cellular components dilute by growth: if a compound is not produced while the volume grows, its concentration drops. This dilution rate is equal to the growth rate of the cellular volume, so in steady state the net synthesis rate of all molecules should be equal to the growth rate. In turn, the same synthesis rates of all molecules determine how much volume is produced per unit time, and thus how fast the cell grows. The synthesis rate thus both depends on, and determines, the growth rate.

A small number of modeling approaches incorporate these two nonlinearities [7, 32–37]. The demand for cell synthesis components is calculated by the models instead of imposed on the models, and the growth rate can only be found after solving a nonlinear problem – or by solving a large number of linear problems in which the growth rate is treated as a parameter, as we will see. To keep our treatment of these complex models as accessible as possible, we will first describe the essential ingredients only. Then we will, referring to SI6 for most of the mathematical derivations, derive a set of relations that enables us to compare these self-fabrication models to the previously described models. After that, we will shortly discuss the various extensions that describe overflow metabolism.

The cell is modeled as consisting of three types of compounds: metabolites (with concentrations $\boldsymbol{x}$[4] and possibly including macromolecules such as lipids or polynucleotides), enzymes (with concentrations $\boldsymbol{e}$), and the ribosome (with concentration r). The enzymes catalyze the conversion of metabolites into other metabolites. The ribosomes catalyze the synthesis of enzymes and ribosomes from metabolites. As before, it is assumed that the rates of the conversions scale proportionally with the concentrations of the catalysts, and kinetic

[4] The usage of $\boldsymbol{x}$ here is not related to its usage in our standard form, Equation (1)

saturation functions are again assumed constant:

$$v_i = e_i k_{\text{cat},i}, \tag{11}$$

$$v_{\text{synth},j} = r k_{\text{cat,rib}} \alpha_j, \tag{12}$$

$$v_{\text{synth,rib}} = r k_{\text{cat,rib}} \alpha_{\text{rib}}. \tag{13}$$

Here v_i is a usual metabolic reaction rate, and $v_{\text{synth},j}$ denotes the synthesis rate of enzyme j. The factor α_j is the fraction of the ribosome that is allocated to the synthesis of enzyme j, and since these are fractions we must have $\sum_j \alpha_j = 1$. It is further assumed that the concentrations of macromolecules add up to a fixed density[5]:

$$\sum_j \rho_j e_j + \rho_{\text{rib}} r = 1, \tag{14}$$

where the ρ_j are volumetric parameters.[6] In a cell that is growing exponentially with rate μ, concentrations dilute with this same rate, see SI5 for a derivation. For the metabolites, this changes the steady state assumption from $N \cdot \boldsymbol{v} = 0$ in FBA approaches to $N \cdot \boldsymbol{v} = \mu \boldsymbol{x}$. Moreover, if we explicitly model enzyme synthesis, we should also account for the metabolites that are consumed during this synthesis. Let M be the matrix that denotes how many metabolites are needed to make a specific enzyme, then we get a first set of constraints on the fluxes:

$$\begin{bmatrix} N & -M \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v}_{\text{synth}} \end{bmatrix} = \mu \boldsymbol{x}, \tag{15}$$

$$v_{\text{synth},j} = \mu e_j, \tag{16}$$

$$v_{\text{synth,rib}} = \mu r. \tag{17}$$

Equations (11) to (17) define the core ingredients of the self-fabrication models. In Supporting Information 6 we show how this system can be rewritten. In short: equations (11), (12) and (13) are used to get expressions for the concentrations $\boldsymbol{e}$ and r in terms of the fluxes, and these expressions are used in equations (14), (16) and (17) to get four relations between the fluxes $\boldsymbol{v}, \boldsymbol{v}_{\text{synth}}$ and μ. These relations are linear in the reaction rates, so that the system can be written in a form that looks like a familiar Linear Program:

$$\begin{aligned} &\underset{\boldsymbol{v},\boldsymbol{v}_{\text{synth}},\mu}{\text{maximize}} \quad \mu \\ &\text{subject to} \quad A(\boldsymbol{x},\mu) \cdot \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v}_{\text{synth}} \\ \mu \end{bmatrix} = \boldsymbol{0} \\ &\qquad\qquad v_i, v_{\text{synth},i} \geq 0. \end{aligned} \tag{18}$$

Although this looks like a Linear Program, it is more difficult since the constraint matrix is not constant: it depends on the metabolite concentrations, $\boldsymbol{x}$, and on the growth rate, μ. The $\boldsymbol{x}$-dependence is often "solved" by ignoring the dilution of small metabolites and fixing the concentrations of macromolecules based on experimental data [33,36][7]. The μ-dependence that makes the problem nonlinear is overcome by fixing the growth rate in the constraint matrix to a certain value: $A(\boldsymbol{x},\mu) \to A(\boldsymbol{x},\mu_0)$, and then add the constraint that the μ in the optimization variables should equal μ_0. Note that, since it is fixed, we can no longer maximize the growth rate. However, we can check if there is a solution that solves the system. If there is no solution, then $\mu_0 > \mu_{\max}$; if there is a solution, we can increase μ_0. The maximal growth solution is found by repeating this procedure until the problem is still feasible for $\mu_0 = \mu_{\max}$, but infeasible for all $\mu_0 > \mu_{\max}$. So we get

$$\begin{aligned} &\underset{\boldsymbol{v},\boldsymbol{v}_{\text{synth}},\mu_0}{\text{maximize}} \quad \mu_0 \\ &\text{such that} \quad \begin{bmatrix} A(\boldsymbol{x},\mu_0) & & \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v}_{\text{synth}} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \mu_0 \end{bmatrix} \\ &\qquad\qquad v_i, v_{\text{synth},i} \geq 0. \end{aligned} \tag{19}$$

Using the described mathematical core, Goelzer et al. proposed a formalism that was named Resource Balance Analysis (RBA), with which they modeled overflow metabolism in *Bacillus subtilis* [38]. In addition to the density constraint of Equation (14), the authors used a constraint on the maximal concentration of macromolecules in the membrane; in our notation: $\sum_j \sigma_j e_j \leq D_{\text{mem}}$.

In parallel, Thiele et al. [35] for *E. coli*, and Lerman et al. [34] for *Thermotoga maritima* presented the so-called Metabolism and Expression (ME) models. The mathematical basis of ME-models is equal to the basis of RBA-models (Equations (11) to (17)), but ME-models are even more comprehensive: for example, the synthesis rates of mRNA, tRNA, and RNA-polymerases are explicitly modeled. Moreover, some catalytic rates, of the ribosome for example, are no longer assumed to be independent of the growth rate; their dependence is estimated from experimental data. These extensions add many variables and constraints to the model, but we show in SI8 that these additions can still be written as relations that are linear in the reaction rates and nonlinear in the growth rate. In short, although the A-matrix of Equation (19) gets larger, ME-models

[5] In some modeling methods this density is modeled as an upper bound [32, 33, 37]. In SI7 we explain the advantages and disadvantages of doing this.

[6] Dependent on the biochemical interpretation of the ρ-parameters, some models include contributions of the metabolite concentrations $\rho_i x_i$ [7,37].

[7] One could also solve the problem for fixed metabolite concentrations [7,37], and then scan over all possible sets of concentrations, but this becomes computationally infeasible in large quantitative models.

can still be written in this form. O'Brien et al. used an ME-model to model overflow metabolism in *E. coli* [36]. The cytosolic density constraint was here supplemented with an upper bound on the substrate uptake flux.

Molenaar et al. [32] were the first to present a mechanistic model of cellular self-fabrication, a core model with 5 reactions and 3 metabolites. Because their model is so small, they could use enzyme kinetics and nonlinear optimization to directly maximize the growth rate. The optimal solutions show a discrete switch from an efficient pathway to an inefficient pathway. This is different from the gradual switch that is observed in overflow metabolism, even though, just as in [33], the authors model an upper bound on the membrane density. However, this does not effectively constrain the concentrations of membrane proteins, since the surface-to-volume ratio can be freely adjusted in the model. Therefore, only the density constraint of Equation (14) is effective. In personal correspondence, the authors confirmed that, in hindsight, it might have been more realistic to set a lower bound to the size of the cell. In this case, an additional constraint would have become active.

4 Discussion

4.1 Commonalities and differences

We reviewed many constrained optimization approaches that describe overflow metabolism, ranging from relatively simple linear Flux Balance Analysis models to complicated nonlinear Metabolism and Expression models. Some approaches use small core models, others use genome-scale networks comprising thousands of reactions. The imposed constraints are either limits on reaction rates, Gibbs dissipation limits or limits on enzyme concentrations. Despite all these differences, we managed to write all these models in a concise standard form by focusing on their mathematical essence. We conclude from this that these models must share a feature, one that must be essential for describing overflow metabolism. In the following, we will compare the models in their standard form using our recent theoretical work [7, 8] to analyze and explain this feature, using an extremum principle that governs the solutions of all the reviewed approaches.

4.1.1 A general extremum principle: Overflow metabolism is caused by two growth-limiting constraints

All reviewed approaches model a growing cell by imposing a set of homogeneous constraints: a first set that ensures a steady state, and a second set that determines the feasible direction for irreversible reactions. There are some differences in how the first set is imposed. FBA and resource allocation approaches model a system that produces cell components in the proportions captured by a constant demand reaction, the biomass reaction. The steady state assumption ensures that no intermediate metabolite accumulates. The self-fabrication models implement this assumption with a balanced growth assumption: all cellular compounds should be produced to match the rate of consumption and dilution by growth. The demand reaction is therefore dependent on the growth rate, which gives rise to nonlinear relations between the optimization variables and the growth rate. These differences are illustrated in Figure 2.

Despite these seemingly different setups, we can define Elementary Modes in both cases: growth-supporting subnetworks that form the minimal building blocks of the solution space. These are called Elementary Flux Modes in the linear models [6], and Elementary Growth Modes in the self-fabricator models [7], see SI1 for a short introduction. The defining property of these modes is that all possible solutions of the growth models can be written as a combination of these modes. In other words, EFMs are non-decomposable metabolic subnetworks, and EGMs are non-decomposable self-fabrication subnetworks. Overflow metabolism is decomposable in an energy-efficient subnetwork, and a less energy-efficient subnetwork, and is thus a combination of two Elementary Modes.

Using the concept of Elementary Modes we derived an extremum principle stating that the number of flux-carrying Elementary Modes in the optimal solution will be smaller or equal than the number of active (i.e. growth-limiting) constraints. These growth-limiting constraints are the additional constraints that are imposed after the steady state and irreversibility constraints (this is illustrated and explained in Figure 1). The extremum principle implies that only one Elementary Mode will be selected by growth rate maximization under one constraint. For example, in a model in which only one nutrient uptake rate is constrained, we will never observe a gradual switch from a high-yield metabolic mode to the combination with a low-yield mode. Overflow metabolism must thus be a result of two constraints, see Figure 2 for an illustration of this result.

The extremum principle suggests that the success of describing overflow metabolism might not lie in the details of the stoichiometric networks, or the exact choices of model parameters, but rather in the mere existence of two constraints. This implies that finding the mecha-

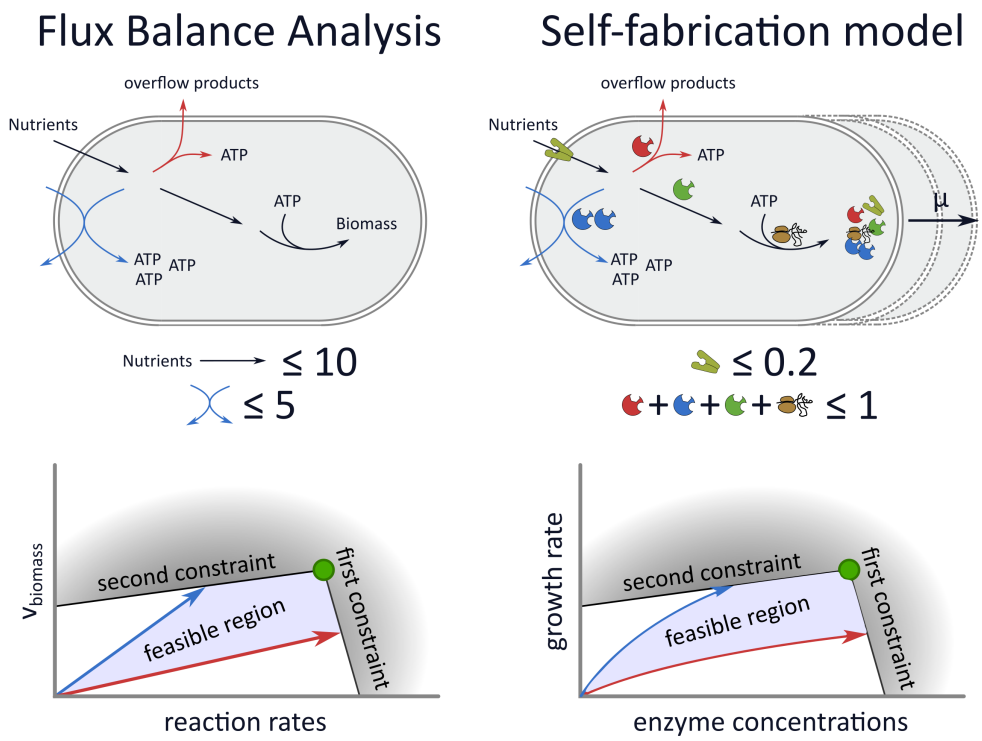

Fig. 2 FBA models and self-fabrication models lead to a similar mathematical problem. In the **top** figures we illustrate two of the reviewed approaches. FBA models consider steady state fluxes through networks of metabolic reactions with constraints on the reaction rates. A virtual biomass reaction is added as a proxy for the growth rate. Self-fabricator models make the synthesis of enzymes and the ribosome explicit, and can therefore model the growth rate as the volume increase due to the production of components. The enzyme concentrations can now be viewed as the optimization variables, so that protein concentration constraints can also be included. In the **bottom** figures we show a highly simplified illustration of the solution space of both approaches. In the linear approaches, FBA and proteome-constrained models, all quantities depend linearly on the growth rate, while there are nonlinear dependencies in the self-fabricator models. However, we showed that in both cases, overflow metabolism is caused by two growth-limiting constraints.

nistic cause of overflow metabolism amounts to finding which two constraints are actually limiting growth.

Unfortunately, the extremum principle does not predict which constraint causes overflow metabolism. It states that there should be two constraints, but does not reveal their identity. Moreover, overflow metabolism in different species might be due to completely different constraints. Thus, to find out which constraints cause overflow metabolism, we must test hypothetical constraints.

4.1.2 Specific experiments: the mechanistic cause of overflow metabolism can be found with falsification experiments

Encouraged by the conclusion that the growth-limiting constraints must be important in causing overflow metabolism, we have listed all the constraints that are used in the reviewed models (Table 1). We see that almost all models indeed use two constraints.[8] Molenaar et al. [32] and Mori et al. [30] form exceptions to this rule, using only one effective constraint. Our theory thus implies that only one EFM will be used in the optimal solutions of these models. Indeed, the models show discrete switches between EFMs when the growth rate increases. The model from Molenaar et al., containing only two EFMs, switches at once from respiration to fermentation. The genome-scale model of Mori et al. contains many different EFMs that form intermediate steps between full respiration and full fermentation. Their model therefore shows many small discrete switches, approximating a gradual switch.[9] This gives

[8] Only Varma and Palsson use more than two constraints. Their third constraint is a lower bound on an ATP-maintenance reaction, which we have left out of the table for clarity.

[9] The authors also present a figure showing a gradual switch, but this is the average behaviour for many models with slightly different parameters. The discontinuities are then averaged out.

Table 1 An overview of the models that try to explain overflow metabolism, including which constraints were used in addition to the steady-state assumptions.

Paper	Type	Constraint 1	Constraint 2
Majewski et al. [11]	FBA	glucose uptake rate	electron transfer capacity
Varma et al. [12]	FBA	glucose uptake rate	oxygen uptake rate
Carlson et al. [14]	FBA	glucose uptake rate	oxygen uptake rate
Niebel et al. [18]	tFBA	glucose uptake rate	free energy dissipation
Basan et al. [27]	resource	glucose uptake rate	total proteome
Mori et al. [30]	resource		total proteome
Vazquez et al. [23, 24]	resource	glucose uptake rate	macromolecular density
Van Hoek et al. [29]	resource	glucose uptake rate	macromolecular density
Zhuang et al. [28]	resource	glucose uptake rate	membrane occupancy
Szenk et al. [22]	resource	glucose uptake rate	membrane occupancy
Shlomi et al. [31]	resource	glucose uptake rate	total proteome
Molenaar et al. [32]	self-fabr		macromolecular density
Goelzer et al. [38]	self-fabr	membrane density	macromolecular density
O'Brien et al. [36]	self-fabr	glucose uptake rate	macromolecular density

rise to a separate hypothesis that we cannot fully exclude. However, we find it more probable that, upon a change in the environment, gene expression continuously tunes the proportions in which two EFMs are used, than that it repeatedly shuts down one EFM to upregulate another.

Among the models that use two constraints, there is some variation in the biological underpinnings of these constraints. The question is how to find the relevant constraints. The genome-scale approach is to make an extensive model and try to quantitatively match the experimental data. One risk, however, is overfitting, because a large enough model could potentially fit any experimental data. Such an approach should therefore be backed up by independent measurements of assumed constraints. Still, it is hard to imagine how a model could distinguish the effects caused by a 'total proteome constraint' and a 'macromolecular density constraint'. For this, we need perturbation experiments. For example, to artificially perturb the proteome allocation of *E. coli*, Basan et al. [27, 39] overexpressed the nonfunctional protein LacZ in one experiment and added translation inhibitors in another. We have derived a formalism in which such perturbation experiments can be analyzed [8]. Basan et al. could provide evidence for their proposed total proteome constraint. In our opinion, this makes their proposed constraint the best-established mechanistic cause of overflow metabolism in *E. coli* up to this point. However, as we now know, there should be a second growth-limiting constraint. Basan et al. used a limit on the uptake rate of nutrients, which cannot truly be considered as a mechanistic cause of overflow metabolism, because, as described by Molenaar et al. in 2009: "... using an artificial maximal capacity constraint on substrate uptake ignores the possibility of variable investments made in substrate transport systems." The identity of the second constraint in *E. coli*, even though a constraint on transport of glucose is generally used and thus apparently accepted, remains to be established.

4.1.3 Towards a complete model of cellular self-fabrication.

We observe two directions of development towards a complete model of cellular self-fabrication in the models that we have reviewed. Along the first direction the optimization variables are moved closer to the actual regulatory space of the cell, and thereby closer to the origin of overflow metabolism. Along the second direction, more and more of the inherent nonlinearity of self-fabrication is incorporated in the models.

To explain the first direction of development, we recall that FBA models use fluxes as variables, which cannot be directly regulated by the cell. Instead, enzyme concentrations are regulated and these will, together with the metabolite concentrations, determine the fluxes. The resource allocation models switch perspective to enzyme concentrations as variables with the major advantage that constraints on enzyme concentrations can be formulated directly. These constraints can be related to physically observable quantities, such as the available membrane area or cytosolic volume, whereas flux constraints cannot. Flux constraints can only be determined ad hoc using experimental data, which limits their predictive power.[10] One can move even further towards the regulatory space of the cell because the enzyme concentrations are in fact dependent on the enzyme synthesis rates, and these are regu-

[10] In fact, many of the resource allocation models still use a flux constraint for the nutrient uptake reaction, so that these are also not entirely predictive.

lated by the allocation of the ribosomes over the different mRNAs. The reviewed self-fabricator models indeed use as variables the enzyme synthesis rates [36, 38], or the ribosome allocation fractions [32]. A final step towards the regulatory space of the cell could be to model the regulation of mRNA synthesis via gene expression directly, but we do not know of any models that have implemented this.

The second direction of development moves towards incorporating three nonlinearities that are related to self-fabrication. We already mentioned two of them: 1) the dependence of the biomass composition on the enzyme allocation, and 2) the dependence of the demanded enzyme synthesis rates on the growth rate. The incorporation of these two nonlinearities form the main improvement of self-fabrication models with respect to FBA type models. Here we want to raise attention for a third nonlinearity: the kinetic dependence of enzyme and ribosome activities on the metabolite concentrations. If a cell reallocates resources, metabolite concentrations change as well, causing changes in the saturation levels of enzymes. Including the metabolite concentrations in the model however, requires information about the enzyme kinetics of all the different enzymes in the cell. Moreover, it makes the optimization problem computationally infeasible because the problem is no longer guaranteed to have only one local optimum. Therefore, global optimization software has to be used and it is difficult to ascertain that the found solution is the actual optimum. For these reasons, enzyme kinetics can so far only be included in core models [32] and theoretical work [7, 8, 37]. The question is if there are constraints, rules and patterns in the changes in (optimal) metabolite levels that would allow us to approximate optimal solutions without global optimisation of the full kinetic model.

4.1.4 Alternatives for growth rate maximization

We have focused on growth rate maximization models, but alternative explanations of overflow metabolism cannot be fully excluded. For example, it might be that not absolute fitness is maximized, but rather relative fitness compared to competitors. For example, cells could produce overflow products to intoxicate their neighbours, or cells could maximize their uptake rate to claim the largest share of the nutrient pool. These explanations have been reviewed elsewhere [40, 41].

It might even be that cells are not completely optimized for anything. For example, it was shown that the overexpression of transcriptional regulator ArcA could increase the growth rate of *E. coli* on glycolytic substrates [42]. This shows that metabolism was not optimal in the wild type strain within the studied environmental conditions.

The sub-optimality of a population of microorganisms might be due to the high regulatory costs that would be required to steer each individual cell to the optimum. De Martino et al. [43] calculated a possible probability distribution for the metabolic states of single cells by maximizing the entropy of this distribution while the average growth rate was fixed to the measured value. This approach leads to a model of single-cell behaviour in which the least additional assumptions were made: 'the probability distribution is as general as possible'. Their predicted distribution captured the measured fluxes better than a Flux Balance Analysis approach. Subsequently, they quantified the amount of regulation that would be needed to get a higher average growth rates, showing that attaining a maximal average growth rate would bring infinite regulatory costs.

5 Conclusion

We reviewed 15 different models of overflow metabolism, ranging from Flux Balance Analyses, to nonlinear self-fabricator models such as Metabolism and Expression models. Despite the many differences between the models, we could rewrite the mathematical cores of each of them into a concise standard form. This standard form could be analyzed using an extremum principle, stating that the number of Elementary Modes at maximal growth is less or equal than the number of growth-limiting constraints. The extremum principle implies that overflow metabolism is caused by at least two growth-limiting constraints. We therefore listed all reviewed models with their proposed constraints. We hope that this list will serve as a source of hypotheses that can now be tested using falsification experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Alan J Wolfe. The acetate switch. 69(1):12–50, 2005.
2. Erik Postma, Cornelis Verduyn, W Alexander Scheffers, and Johannes P Van Dijken. Enzymic analysis of the Crabtree effect in glucose-limited chemostat cultures of *Saccharomyces-cerevisiae*. *Applied and Environmental Microbiology*, 55(2):468–477, 1989.
3. Otto Warburg. On the Origin of Cancer Cells. *Science*, 123(3191):309–314, 1956.

4. T D Thomas, D C Ellwood, and V M Longyear. Change from homo- to heterolactic fermentation by Streptococcus lactis resulting from glucose limitation in anaerobic chemostat cultures. *Journal of bacteriology*, 138(1):109–117, apr 1979.
5. James Patrick Folsom and Ross P. Carlson. Physiological, biomass elemental composition and proteomic analyses of *Escherichia coli* ammoniumlimited chemostat growth, and comparison with iron- and glucose-limited chemostat growth. *Microbiology (United Kingdom)*, 161(8):1659–1670, 2015.
6. Stefan Schuster and Claus Hilgetag. On Elementary Flux Modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 02(02):165–182, 1994.
7. Daan H de Groot, Josephus Hulshof, Bas Teusink, Frank J Bruggeman, and Robert Planqué. Elementary Growth Modes provide a molecular description of cellular self-fabrication. *bioRxiv*, page 608083, jan 2019.
8. Daan H De Groot, Coco Van Boxtel, Robert Planqué, Frank J. Bruggeman, and Bas Teusink. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLOS Computational Biology*, pages 1–25, 2018.
9. Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3):344–349, 2010.
10. K B Andersen and K von Meyenburg. Are growth rates of *Escherichia coli* in batch cultures limited by respiration? *Journal of bacteriology*, 144(1):114–123, oct 1980.
11. Michael M Domach and R A Majewski. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnology and Bioengineering*, 35(7):732–738, 1990.
12. A Varma and B O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and environmental microbiology*, 60(10):3724–31, oct 1994.
13. Ross Carlson and Friedrich Srienc. Fundamental *Escherichia coli* Biochemical Pathways for Biomass and Energy Production: Identification of Reactions. *Biotechnology and Bioengineering*, 85(1):1–19, 2004.
14. Ross Carlson and Friedrich Srienc. Fundamental *Escherichia coli* Biochemical Pathways for Biomass and Energy Production: Creation of Overall Flux States. *Biotechnology and Bioengineering*, 86(2):149–162, 2004.
15. Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical journal*, 92(5):1792–1805, mar 2007.
16. Elad Noor. Removing both internal and unrealistic energy-generating cycles in Flux Balance Analysis. *arXiv.org*, 2018.
17. Jan Schellenberger, Nathan E Lewis, and Bernhard O Palsson. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, 100(5):1381, 2011.
18. Bastian Niebel, Simeon Leupold, and Matthias Heinemann. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*, 1(1):125–132, 2019.
19. Athel Cornish-Bowden. *Principles of enzyme kinetics*. Elsevier, 2014.
20. Sheng Hui, Josh M Silverman, Stephen S Chen, David W Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular systems biology*, 11(1):784, 2015.
21. Conghui You, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim, Carl W Gunderson, Yi Ping Wang, Peter Lenz, Dalai Yan, and Terence Hwa. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500(7462):301–306, 2013.
22. Mariola Szenk, Ken A. Dill, and Adam M.R. R de Graff. Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell Systems*, 5(2):95–104, apr 2017.
23. Alexei Vazquez, Qasim K Beg, Marcio A DeMenezes, Jason Ernst, Ziv Bar-Joseph, Albert-László Barabási, László G Boros, and Zoltán N Oltvai. Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Systems Biology*, 2:7, jan 2008.
24. Alexei Vazquez and Zoltán N Oltvai. Macromolecular crowding explains overflow metabolism in cells. *Scientific Reports*, 6:31007, 2016.
25. Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.
26. Matthew Scott, Stefan Klumpp, Eduard M Mateescu, and Terence Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular systems biology*, 10(8):747, 2014.
27. Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R Williamson, and Terence Hwa. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*, 528(7580):99–104, 2015.
28. Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. Economics of membrane occupancy and respiro-fermentation. *Molecular Systems Biology*, 7(500):500, 2011.
29. Milan J A van Hoek and Roeland M H Merks. Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC systems biology*, 6(1):22, 2012.
30. Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. Constrained allocation Flux Balance Analysis. *PLoS Computational Biology*, 12(6):1–24, 2016.
31. Tomer Shlomi, Tomer Benyamini, Eyal Gottlieb, Roded Sharan, and Eytan Ruppin. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Computational Biology*, 7(3):1–8, 2011.
32. Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. Shifts in growth strategies reflect trade-offs in cellular economics. *Molecular systems biology*, 5(323):323, 2009.
33. Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218, 2011.
34. Joshua A. Lerman, Daniel R. Hyduke, Haythem Latif, Vasiliy A. Portnoy, Nathan E. Lewis, Jeffrey D. Orth, Alexandra C. Schrimpe-Rutledge, Richard D. Smith, Joshua N. Adkins, Karsten Zengler, and Bernhard O. Palsson. In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications*, 3(May):910–929, 2012.
35. Ines Thiele, Ronan M.T. Fleming, Richard Que, Aarash Bordbar, Dinh Diep, and Bernhard O Palsson. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS ONE*, 7(9), 2012.
36. Edward J. O'Brien, Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke, and Bernhard Palsson. Genome-scale

models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(693), 2013.

37. Hugo Dourado and Martin J Lercher. An Analytical Theory of Cellular Growth. *bioRxiv*, page 607374, 2019.
38. Anne Goelzer, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel, Rolf Nölker, Mahendra Mariadassou, Stéphane Aymerich, Michael Hecker, Philippe Noirot, Dörte Becher, and Vincent Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, nov 2015.
39. Markus Basan. Resource allocation and metabolism: the search for governing principles. *Current Opinion in Microbiology*, 45:77–83, 2018.
40. Anisha Goel, Meike Tessa Wortel, Douwe Molenaar, and Bas Teusink. Metabolic shifts: A fitness perspective for microbial cell factories. *Biotechnology Letters*, 34(12):2147–2160, 2012.
41. Thomas Pfeiffer and Annabel Morley. An evolutionary perspective on the Crabtree effect. *Frontiers in molecular biosciences*, 1(October):17, 2014.
42. Markus Basan, Sheng Hui, and James R Williamson. ArcA overexpression induces fermentation and results in enhanced growth rates of *E. coli*. *Scientific Reports*, 7(1):11866, 2017.
43. Daniele De Martino, Anna MC Andersson, Tobias Bergmiller, Călin C Guet, and Gašper Tkačik. Statistical mechanics for metabolic networks during steady state growth. *Nature Communications*, 9(1):2988, 2018.