Elementary vectors and autocatalytic sets for computational models of cellular growth

Stefan Müller[1*], Diana Széliová[2], Jürgen Zanghellini[2,3*]

1 Faculty of Mathematics, University of Vienna, Austria **2** Department of Analytical Chemistry, University of Vienna, Austria **3** Austrian Centre of Industrial Biotechnology, Vienna, Austria

* st.mueller@univie.ac.at (SM), juergen.zanghellini@univie.ac.at (JZ)

Abstract

Traditional models of cellular growth involve an approximative biomass "reaction" which specifies biomass composition in terms of precursor metabolites (such as amino acids and nucleotides). On the one hand, biomass composition is often not known exactly and may vary drastically between extreme conditions; on the other hand, the predictions of computational models crucially depend on biomass. Even elementary flux modes (EFMs) depend on the biomass reaction. (To be specific: not just the numerical values of the EFMs, but also their supports and their number.)

To better understand cellular phenotypes across conditions, we introduce and analyze new classes of elementary vectors for more comprehensive models of cellular growth, involving explicit synthesis reactions for all macromolecules. *Growth modes* (GMs) are given by stoichiometry, and *elementary* growth modes (EGMs) are GMs that cannot be decomposed without cancellations. Unlike EFMs, EGMs need not be support-minimal. Most importantly, every GM can be written as a sum of EGMs. In models with additional (capacity) constraints, *growth vectors* (GVs) and *elementary* growth vectors (EGVs) also depend on growth rate. In any case, EGMs/EGVs do not depend on the biomass composition. In fact, they cover all possible biomass compositions and can be seen as unbiased versions of elementary flux modes/vectors (EFMs/EFVs) used in traditional models.

To relate the new concepts to other branches of theory, we define *autocatalytic* GMs and the corresponding autocatalytic sets of reactions. Further, we illustrate our results in a small model of a self-fabricating cell, involving glucose and ammonium uptake, amino acid and lipid synthesis, and the expression of all enzymes and the ribosome itself. In particular, we study the variation of biomass composition as a function of growth rate. In agreement with experimental data, low nitrogen uptake correlates with high carbon (lipid) storage.

1 Introduction

A major characteristics of life is self-fabrication, involving self-maintenance and self-replication. Cellular self-fabrication requires the acquisition and transformation of nutrients, not only to maintain the cell, but also to replicate, that is, to grow. During one cycle, a cell needs to duplicate all its building blocks and to cover the related costs. In constant environments, this process is balanced and leads to exponential growth. Indeed, exponential growth means that all cellular components are synthesized in

proportion to their abundance. Clearly, metabolic activity depends on growth rate since higher growth rates imply higher synthesis rates which in turn require more ribosomes that produce the additional enzymes (and the ribosomes themselves). Thus, growth can be seen as a process that allocates cellular resources, limited by environmental and physico-chemical constraints.

Arguably, some of the most successful approaches to study (microbial) growth processes are rooted in constraint-based modeling [1]. At their heart sits a (genome-scale) metabolic model, which captures all possible (cell-specific) biochemical transformations in an annotated and mathematically structured form. The resulting stoichiometric matrix, coupled with environmental and physico-chemical constraints, is already sufficient to study (fundamental aspects of) growth. Various biased and unbiased methods have been developed to predict (steady-state) metabolic phenotypes [2]. However, first-generation genome-scale metabolic models do not involve the expression of proteins and hence do not account for the individual enzyme costs of metabolic fluxes. These models rather use a fixed biomass composition that represents the average costs of growth. Recent efforts focus on the development of next-generation genome-scale metabolic models that also account for the enzyme demands of individual reactions. As one prominent example, we mention resource balance analysis (RBA) [3]. Although specific approaches differ in their degree of mechanistic detail and mathematical formulation, the central concept in the analysis of next-generation models is the (optimal) allocation of resources. However, we currently lack a theoretical understanding of all feasible allocations in next-generation metabolic models.

In traditional, first-generation models, elementary flux modes and vectors (EFMs [4, 5] and EFVs [6, 7]) allow an unbiased study of metabolic pathways, in particular, an interpretation of any *feasible* flux distribution in terms of unique functional units. On the other hand, flux balance analysis (FBA) enables the computationally efficient identification of *optimal* fluxes [8]. In this respect, RBA can be viewed as a generalization of FBA that also accounts for enzyme costs. It represents the corresponding tool to identify optimal fluxes (and corresponding optimal allocations) in comprehensive, next-generation models. However, the analogues of EFMs and EFVs have not been identified, yet. In [9], elementary growth states have been introduced for (semi-)kinetic models of cellular self-fabrication, but for constraint-based models, the corresponding theoretical concepts are still missing.

In this work, we introduce elementary growth modes and vectors (EGMs and EGVs) that are given by stoichiometry and irreversibility (EGMs) as well as by additional constraints and growth rate (EGVs). In fact, we develop a mathematical theory that enables an unbiased characterization of all feasible flux distributions in constraint-based models of cellular growth. In analogy to EFMs and EFVs, EGMs and EGVs can be interpreted as unique functional units of self-fabrication. Even more importantly, they provide an unbiased understanding of all possible biomass compositions. In a small example of a self-fabricating cell, we highlight that the experimentally observed lipid accumulation upon nitrogen starvation is a general feature of a comprehensive, next-generation metabolic model and not the result of active regulation.

Finally, we relate EGMs and EGVs to other branches of theory. Obviously, purely stoichiometric models do not reflect implications from autocatalysis (namely, that all catalysts of active reactions need to be synthesized) and kinetics (namely, that all species involved in active reactions need to be present with nonzero concentrations). We observe that additional (capacity) constraints often ensure that EGVs are autocatalytic.

2 Results

In this section, we introduce and analyze *elementary growth modes* and *vectors* for computational models of cellular growth. Further, we relate our approach to the theory of autocatalytic sets. In order to motivate the new concepts and to illustrate our results, we use a running example (a small model of a self-fabricating cell). In Section 3 (Methods), we summarize the mathematical notation used throughout this work and provide an introduction to elementary vectors in polyhedral geometry. Finally, in Section 4 (Discussion), we elaborate on the biological interpretation of our results for the running example.

In Supporting Information A, we derive the dynamic model of cellular growth studied in this work; in Supporting Information B, we determine the membrane constraints for the small model of a self-fabricating cell; and in Supporting Information C, we provide further figures and tables to illustrate our results for the running example.

2.1 Growth model

The dynamic model of cellular growth studied in this work is given by the mass balance equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = N v(x) - \mu\, x \tag{1a}$$

and, additionally, by the dry weight constraint

$$\omega^T x = 1. \tag{1b}$$

Here, x and ω are the vectors of concentrations and molar masses (of the molecular species), respectively, v is the vector of reaction rates, N is the stoichiometric matrix which captures the net effect of the chemical reactions (and other processes), and μ is the growth rate. We summarize the fundamental objects and quantities in Table 1 and provide a minimal derivation of the set of Eqs (1) in Supporting Information A.

Table 1. Fundamental objects and quantities of cellular growth.

Symbol	**Name**	**Unit**
Mol	set of molecular species	–
Rxn	set of chemical reactions	–
$N \in \mathbb{R}^{\mathsf{Mol}\times\mathsf{Rxn}}$	(net) stoichiometric matrix	1
$\omega \in \mathbb{R}^{\mathsf{Mol}}_{>}$	molar masses	$\mathrm{g\,mol^{-1}}$
$x \in \mathbb{R}^{\mathsf{Mol}}_{\geq}$	concentrations	$\mathrm{mol\,g^{-1}}$
$v(x) \in \mathbb{R}^{\mathsf{Rxn}}$	reaction rates	$\mathrm{mol\,g^{-1}\,h^{-1}}$
$\mu \in \mathbb{R}_{\geq}$	growth rate	$\mathrm{h^{-1}}$

In the following, we also consider (dimensionless) mass fractions $(\omega \circ x) \in \mathbb{R}^{\mathsf{Mol}}_{\geq}$. Obviously, $\sum_{i\in\mathsf{Mol}} (\omega \circ x)_i = \sum_{i\in\mathsf{Mol}} \omega_i\, x_i = \omega^T x = 1$.

By Eqs (1), growth rate is given by

$$\mu = \omega^T N v. \tag{2}$$

By distinguishing between exchange and internal reactions, Eq (2) can be rewritten as

$$\mu = \omega^T N v = \omega^T \begin{pmatrix} N_{\text{exc}} & N_{\text{int}} \end{pmatrix} \begin{pmatrix} v_{\text{exc}} \\ v_{\text{int}} \end{pmatrix} = \omega^T N_{\text{exc}}\, v_{\text{exc}} \tag{3}$$

since $\omega^T N_{\text{int}} = 0$, by mass conservation. In biological terms, growth rate is determined by the exchange reactions with the environment (the uptake of nutrients and the excretion of waste products).

We highlight two observations, proven in Supporting Information A:

- *Conservation laws.* In a model of cellular growth, there cannot be any conservation laws. In mathematical terms, $\ker N^T \cap \mathbb{R}^{\mathsf{Mol}}_{\geq} = \{0\}$.
- *Dependent concentrations.* Still, there can be dependent concentrations. That is, possibly $\ker N^T \neq \{0\}$.

In the following, we assume that all dependent variables have been eliminated. In mathematical terms, $\ker N^T = \{0\}$, that is, $\operatorname{im} N = \mathbb{R}^{\mathsf{Mol}}$.

In order to motivate the new concepts in this work and to illustrate our results, we will use a running example.

2.2 Example

Consider the small model of a self-fabricating cell shown in Fig 1(a). The cell takes up glucose (G) and ammonium (N) via the catalytic reactions

$$r_{\mathsf{IG}}\colon\ 0 \xrightarrow{\mathsf{IG}} \mathsf{G} \quad \text{and} \quad r_{\mathsf{IN}}\colon\ 0 \xrightarrow{\mathsf{IN}} \mathsf{N},$$

respectively, and forms amino acids (AA) and lipids,

$$r_{\mathsf{EEA}}\colon\ \mathsf{G} + \mathsf{N} \xrightarrow{\mathsf{EAA}} \mathsf{AA}, \qquad r_{\mathsf{ELD}}\colon\ n_{\mathsf{L}} \cdot \mathsf{G} \xrightarrow{\mathsf{ELD}} \mathsf{LD}, \qquad r_{\mathsf{EL}}\colon\ \mathsf{LD} \xrightarrow{\mathsf{EL}} \mathsf{L}.$$

In fact, lipids are stored in the form of lipid droplets (LD) and transferred to the membrane (L).

Amino acids are the essential building blocks for the importers ($I = \mathsf{IG}, \mathsf{IN}$), the enzymes ($E = \mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}$), and the ribosome R,

$$s_I\colon\ n_{\mathsf{I}} \cdot \mathsf{AA} \xrightarrow{\mathsf{R}} I, \quad s_E\colon\ n_{\mathsf{E}} \cdot \mathsf{AA} \xrightarrow{\mathsf{R}} E, \quad s_{\mathsf{R}}\colon\ n_{\mathsf{R}} \cdot \mathsf{AA} \xrightarrow{\mathsf{R}} \mathsf{R}.$$

Thereby, stoichiometric coefficients are denoted by n_i with $i \in \{\mathsf{L}, \mathsf{I}, \mathsf{E}, \mathsf{R}\}$. The resulting stoichiometric matrix and the corresponding flux vector are displayed in Fig 1(b), and parameter values are given in Table 2.

In the next subsection, we introduce the basic "building blocks" of any possible flux distribution in a (purely stoichiometric) model of cellular growth.

2.3 Elementary growth modes

At steady state, Eq (1a) implies

$$N v = \mu\, x \geq 0. \tag{4}$$

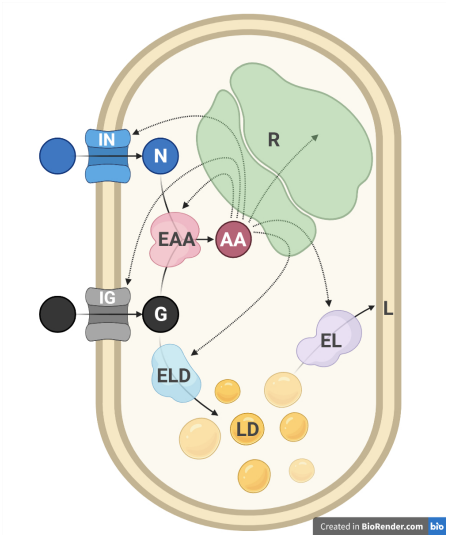
Further, in a given setting, some reactions may have a given direction, as determined by thermodynamics. That is,

$$v_{\mathcal{I}} \geq 0, \tag{5}$$

where $\mathcal{I} \subseteq \mathsf{Rxn}$ denotes the set of irreversible reactions.

The inequalities $Nv \geq 0$ and $v_{\mathcal{I}} \geq 0$ (for the fluxes v) specify a polyhedral cone which suggests the following definition.

(a)



(b)

$$
\begin{array}{c}
 \\ \mathsf{G} \\ \mathsf{N} \\ \mathsf{AA} \\ \mathsf{LD} \\ \mathsf{L} \\ \mathsf{IG} \\ \mathsf{IN} \\ \mathsf{EAA} \\ \mathsf{ELD} \\ \mathsf{EL} \\ \mathsf{R}
\end{array}
\begin{array}{c}
\begin{array}{ccccccccccc}
r_{\mathsf{IG}} & r_{\mathsf{IN}} & r_{\mathsf{EAA}} & r_{\mathsf{ELD}} & r_{\mathsf{EL}} & s_{\mathsf{IG}} & s_{\mathsf{IN}} & s_{\mathsf{EAA}} & s_{\mathsf{ELD}} & s_{\mathsf{EL}} & s_{\mathsf{R}}
\end{array} \\
\begin{pmatrix}
1 & 0 & -1 & -n_{\mathsf{L}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & -n_{\mathsf{I}} & -n_{\mathsf{I}} & -n_{\mathsf{E}} & -n_{\mathsf{E}} & -n_{\mathsf{E}} & -n_{\mathsf{R}} \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
= N,
$$

$$
(v_{\mathsf{IG}},\, v_{\mathsf{IN}},\, v_{\mathsf{EAA}},\, v_{\mathsf{ELD}},\, v_{\mathsf{EL}};\, w_{\mathsf{IG}},\, w_{\mathsf{IN}},\, w_{\mathsf{EAA}},\, w_{\mathsf{ELD}},\, w_{\mathsf{EL}},\, w_{\mathsf{R}})^T = v.
$$

Fig 1. A small model of a self-fabricating cell. (a) The cell has two types of importers, taking up glucose (G) and ammonium (N) from the environment, and three types of metabolic enzymes, synthesizing amino acids (AA) from glucose and ammonium, forming lipid droplets (LD) from glucose, and transferring lipids (L) to the membrane. Amino acids are used by the ribosome (R) to synthesize importers (IN, IG), metabolic enzymes (EAA, ELD, EL), and the ribosome itself. **(b)** The resulting stoichiometric matrix and the corresponding flux vector. (Thereby v is used for import and enzymatic reactions and w for synthesis reactions.)

Definition 1. *Growth modes* (GMs) for the dynamic growth model (1) are elements of the *growth cone*

$$
C_{\mathrm{g}} = \left\{ v \in \mathbb{R}^{\mathsf{Rxn}} \mid Nv \geq 0,\, v_{\mathcal{I}} \geq 0 \right\}.
$$

A GM $v \in C_{\mathrm{g}}$ has an *associated* growth rate $\mu(v) = \omega^T N v \geq 0$ and, if $\mu(v) > 0$, an *associated* concentration vector $x(v) = Nv/\mu(v) \in \mathbb{R}^{\mathsf{Mol}}_{\geq}$.

Elementary growth modes (EGMs) are conformally non-decomposable GMs.

At this point, we refer the reader to Subsection 3.1 for an introduction to elementary vectors in polyhedral geometry (in particular, for the definitions of polyhedral cones and conformal non-decomposability). Still, we also try to provide an intuitive understanding of Definition 1:

- GMs are fluxes given by stoichiometry and irreversibility; in particular, they do not depend on concentrations or growth rate. Still, GMs have an associated growth rate and an associated concentration vector, as given by Eqs (2) and (4).
- A growth cone C_{g} is a general polyhedral cone. In contrast, a flux cone $C_{\mathrm{f}} = \left\{ v \in \mathbb{R}^{\mathsf{Rxn}} \mid Nv = 0,\, v_{\mathcal{I}} \geq 0 \right\}$, arising also in traditional models of cellular growth, is a special cone lying in a subspace (the nullspace of the stoichiometric matrix). For visualizations of a special and a (more) general polyhedral cone, see Figs 5 and 6 in Subsection 3.1.
- Whereas the elementary vectors of a flux cone are support-minimal, the elementary vectors of a growth cone are conformally non-decomposable. Thereby,

a conformally non-decomposable vector cannot be written as a sum of other vectors *without cancellations* [10].

- In applications, the growth cone (as the flux cone) is often pointed (due to irreversibility).

GMs are scalable, but the associated concentrations are scale invariant.

Proposition 2. *For a GM $v \in C_{\mathrm{g}}$ with associated concentration $x(v)$ and $\lambda > 0$, it holds that $x(\lambda v) = x(v)$.*

Proof.

$$x(\lambda v) = \frac{N(\lambda v)}{\mu(\lambda v)} = \frac{N(\lambda v)}{\omega^T N(\lambda v)} = \frac{Nv}{\omega^T N v} = x(v).$$

□

Further, we can characterize the case of zero growth rate.

Proposition 3. *For a GM $v \in C_{\mathrm{g}}$, $\mu(v) = 0$ is equivalent to $Nv = 0$.*

Proof. Obviously, $Nv = 0$ implies $\mu(v) = \omega^T N v = 0$. Conversely, $\mu(v) = \omega^T N v = 0$ with $\omega \in \mathbb{R}^{\mathsf{Mol}}_{>}$ and $Nv \in \mathbb{R}^{\mathsf{Rxn}}_{\geq}$ implies $Nv = 0$. □

GMs v with $\mu(v) = 0$ are flux modes (FMs), that is, elements of the flux cone

$$C_{\mathrm{f}} = \left\{ v \in \mathbb{R}^{\mathsf{Rxn}} \mid Nv = 0,\, v_{\mathcal{I}} \geq 0 \right\}.$$

EGMs v with $\mu(v) = 0$ are elementary flux modes (EFMs), that is, support-minimal elements of C_{f}. They are not support-minimal elements of C_{g}, in general.

Finally, we apply the general theory of elementary vectors and state the main result of this subsection.

Theorem 4. *Every nonzero GM is a conformal sum of EGMs.*

Proof. By Theorem 11 in Subsection 3.1. The growth cone C_{g} is a general polyhedral cone, and its elementary vectors are the conformally non-decomposable vectors, that is, the EGMs. □

In fact, EGMs with nonzero growth rate can be scaled to have the same growth rate as the given GM.

Corollary 5. *Let v be a nonzero GM with associated growth rate $\mu(v) =: \mu$. Then, there exist (possibly empty) finite sets E_0 and E_μ of EGMs with associated growth rates 0 and $\mu > 0$, respectively, such that*

$$v = \sum_{e \in E_0} e + \sum_{e \in E_\mu} \lambda_e\, e \quad \textit{with}\ \operatorname{sign}(e) \leq \operatorname{sign}(v),$$

$\lambda_e \geq 0$, and $\sum_{e \in E_\mu} \lambda_e = 1$. Moreover, if $\mu > 0$, then $x(v) = \sum_{e \in E_\mu} \lambda_e\, x(e)$.

Proof. By Theorem 4, there exist (possibly empty) finite sets E_0 and $E_>$ of EGMs (with associated growth rates 0 and >0, respectively) such that

$$v = \sum_{e \in E_0} e + \sum_{e \in E_>} e \quad \text{with}\ \operatorname{sign}(e) \leq \operatorname{sign}(v).$$

In particular, $\mu = \mu(v) = \sum_{e \in E_>} \mu(e)$. If $\mu > 0$, that is, $E_> \neq \emptyset$, then

$$\sum_{e \in E_>} e = \sum_{e \in E_>} \frac{\mu(e)}{\mu} \frac{\mu}{\mu(e)}\, e = \sum_{e' \in E_\mu} \lambda_{e'}\, e',$$

where $e' = \frac{\mu}{\mu(e)}\, e$ with $\mu(e') = \mu$ and $\lambda_{e'} = \frac{\mu(e)}{\mu}$ with $\sum_{e' \in E_\mu} \lambda_{e'} = 1$.

Finally, if $\mu > 0$, then $x(v) = Nv/\mu = \sum_{e \in E_\mu} \lambda_e\, Ne/\mu = \sum_{e \in E_\mu} \lambda_e\, x(e)$. □

In Corollary 5, we actually fix growth rate which turns the growth cone into a polyhedron. Hence, the result is also an instance of Theorem 12 in Subsection 3.1.

Example (EGMs). Recall the small model of a self-fabricating cell given in Fig 1 and note that all reactions are assumed to be irreversible, $\mathcal{I} = \mathsf{Rxn}$. As it turns out, there are 11 EGMs (up to scaling), corresponding to the 11 molecular species, that is, $e^i \in \mathbb{R}^{\mathsf{Rxn}}_{\geq}$ with $i \in \mathsf{Mol} = \{\mathsf{G}, \mathsf{N}, \mathsf{AA}, \mathsf{LD}, \mathsf{L}, \mathsf{IG}, \mathsf{IN}, \mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}, \mathsf{R}\}$. Explicitly,

$$\begin{aligned}
e^{\mathsf{G}} &= \mu/\omega_{\mathsf{G}} \cdot (1,0,0,0,0;0,0,0,0,0,0)^T,\\
e^{\mathsf{N}} &= \mu/\omega_{\mathsf{N}} \cdot (0,1,0,0,0;0,0,0,0,0,0)^T,\\
e^{\mathsf{AA}} &= \mu/\omega_{\mathsf{AA}} \cdot (1,1,1,0,0;0,0,0,0,0,0)^T,\\
e^{\mathsf{LD}} &= \mu/\omega_{\mathsf{LD}} \cdot (n_{\mathsf{L}},0,0,1,0;0,0,0,0,0,0)^T,\\
e^{\mathsf{L}} &= \mu/\omega_{\mathsf{L}} \cdot (n_{\mathsf{L}},0,0,1,1;0,0,0,0,0,0)^T,\\
e^{\mathsf{IG}} &= \mu/\omega_{\mathsf{IG}} \cdot (n_{\mathsf{I}},n_{\mathsf{I}},n_{\mathsf{I}},0,0;1,0,0,0,0,0)^T,\\
e^{\mathsf{IN}} &= \mu/\omega_{\mathsf{IN}} \cdot (n_{\mathsf{I}},n_{\mathsf{I}},n_{\mathsf{I}},0,0;0,1,0,0,0,0)^T,\\
e^{\mathsf{EAA}} &= \mu/\omega_{\mathsf{EAA}} \cdot (n_{\mathsf{E}},n_{\mathsf{E}},n_{\mathsf{E}},0,0;0,0,1,0,0,0)^T,\\
e^{\mathsf{ELD}} &= \mu/\omega_{\mathsf{ELD}} \cdot (n_{\mathsf{E}},n_{\mathsf{E}},n_{\mathsf{E}},0,0;0,0,0,1,0,0)^T,\\
e^{\mathsf{EL}} &= \mu/\omega_{\mathsf{EL}} \cdot (n_{\mathsf{E}},n_{\mathsf{E}},n_{\mathsf{E}},0,0;0,0,0,0,1,0)^T,\\
e^{\mathsf{R}} &= \mu/\omega_{\mathsf{R}} \cdot (n_{\mathsf{R}},n_{\mathsf{R}},n_{\mathsf{R}},0,0;0,0,0,0,0,1)^T.
\end{aligned}$$

Thereby, we introduced the factors μ/ω_i to obtain correct units ($\mathrm{mol\,g^{-1}\,h^{-1}}$) and equal growth rate μ for all EGMs. We note the following points:

- Every EGM "produces" exactly one molecular species, as indicated by its name. (For every EGM, there is exactly one molecular species with nonzero associated concentration.) Formally, $Ne^i = (\mu/\omega_i)\, u^i$, where u^i is the ith unit vector in $\mathbb{R}^{\mathsf{Mol}}$. That is, e^{G} produces G (glucose), e^{N} produces N (ammonium), …, and e^{R} produces R (ribosome).

  This special situation arises from the fact that the stoichiometric matrix N is square. If N has more columns than rows (or vice versa), then there may be more than one (or zero) EGMs for the production of the individual molecular species.

- EGMs e^{G} and e^{N} are support-minimal, but all other EGMs are not.

- For the associated growth rates, we obtain $\mu(e^i) = \omega^T N e^i = \mu/\omega_i \cdot \omega^T u^i = \mu/\omega_i \cdot \omega_i = \mu$. (This is just the consequence of introducing the factors above.) For the associated concentrations, we obtain $x(e^i) = Ne^i/\mu(e^i) = (\mu/\omega_i)\, u^i/\mu = u^i/\omega_i$.

- There are no EGMs with zero growth rate, that is, there are no EFMs.

To give an example of a conformal sum, we consider the GM

$$v' = \begin{array}{c} \begin{array}{ccccccccccc} r_{\mathsf{IG}} & r_{\mathsf{IN}} & r_{\mathsf{EAA}} & r_{\mathsf{ELD}} & r_{\mathsf{EL}} & s_{\mathsf{IG}} & s_{\mathsf{IN}} & s_{\mathsf{EAA}} & s_{\mathsf{ELD}} & s_{\mathsf{EL}} & s_{\mathsf{R}} \end{array} \\ \left(\begin{array}{ccccccccccc} *, & *, & *, & 0, & 0; & *, & *, & 0, & 0, & 0, & 0 \end{array} \right) \end{array}$$

with support in the import/enzymatic reactions $r_{\mathsf{IG}}, r_{\mathsf{IN}}, r_{\mathsf{EAA}}$ and in the synthesis reactions $s_{\mathsf{IG}}, s_{\mathsf{IN}}$. In particular, v' is a GM without lipid synthesis. Now, let $\mu(v') = \mu$. By Theorem 4 (and Corollary 5), v' is a conformal (and convex) sum of EGVs,

$$v' = \lambda_{\mathsf{G}}\, e^{\mathsf{G}} + \lambda_{\mathsf{N}}\, e^{\mathsf{N}} + \lambda_{\mathsf{AA}}\, e^{\mathsf{AA}} + \lambda_{\mathsf{IG}}\, e^{\mathsf{IG}} + \lambda_{\mathsf{IN}}\, e^{\mathsf{IN}}$$

with $\lambda_{\mathsf{G}}, \lambda_{\mathsf{N}}, \lambda_{\mathsf{AA}} \geq 0$, $\lambda_{\mathsf{IG}}, \lambda_{\mathsf{IN}} > 0$ and $\lambda_{\mathsf{G}} + \lambda_{\mathsf{N}} + \lambda_{\mathsf{AA}} + \lambda_{\mathsf{IG}} + \lambda_{\mathsf{IN}} = 1$. Note that v' produces IG, IN (since $\lambda_{\mathsf{IG}}, \lambda_{\mathsf{IN}} > 0$), however, it produces $\mathsf{G}, \mathsf{N}, \mathsf{AA}$ if and only if also $\lambda_{\mathsf{G}}, \lambda_{\mathsf{N}}, \lambda_{\mathsf{AA}} > 0$.

Since GMs are given by stoichiometry (and irreversibility), they do not reflect constraints implied by autocatalysis and kinetics.

Example (autocatalysis and kinetics). Obviously, the GM v' above involves catalytic reactions. In particular, it involves r_{EAA} (amino acid synthesis) as well as $s_{\mathsf{IG}}, s_{\mathsf{IN}}$ (the expression of the importers). However, it is not catalytically closed in the sense that reactions r_{EAA} and $s_{\mathsf{IG}}, s_{\mathsf{IN}}$ carry fluxes, but the corresponding catalysts, the enzyme EAA and the ribosome R, are not expressed. Now, EGV e^{EAA} involves s_{EAA} (the expression of EAA), and analogously e^{R} involves s_{R}. Hence, we form a convex sum

$$v'' = (1 - \lambda_{\mathsf{EAA}} - \lambda_{\mathsf{R}})\, v' + \lambda_{\mathsf{EAA}}\, e^{\mathsf{EAA}} + \lambda_{\mathsf{R}}\, e^{\mathsf{R}} = (*, *, *, 0, 0; *, *, *, 0, 0, *)^T$$

with $\lambda_{\mathsf{EAA}}, \lambda_{\mathsf{R}} > 0$, to obtain a GM that is catalytically closed.

Further, the EGM e^{G} is kinetically consistent in the sense that it involves the reaction r_{IG} and hence the species G *and* it has a corresponding nonzero associated concentration $x_{\mathsf{G}}(e^{\mathsf{G}}) > 0$, as implied by kinetics. By the discussion above, the GM v' is kinetically consistent if and only if $\lambda_{\mathsf{G}}, \lambda_{\mathsf{N}}, \lambda_{\mathsf{AA}} > 0$. In this case, all species involved in active reactions have nonzero associated concentrations.

In the next subsection, we elaborate on the concepts of catalytic closure and kinetic consistency.

2.4 Implications from autocatalysis and kinetics

Cellular growth is autocatalytic in the sense that the cell fabricates itself (thereby exchanging substrates/products with the environment). One needs to distinguish this notion of "network autocatalysis" from "autocatalytic subnetworks" [11,12].

Consider the overall reaction corresponding to (the flux through) a subnetwork. If a molecular species appears on both the educt and product sides, in particular, with a larger stoichiometric coefficient on the product side than on the educt side, then it is *formally autocatalytic* (cf. [13]). In fact, there are several competing notions of autocatalytic species and subnetworks (cf. [11–13]).

In this work, we consider network autocatalysis. Before we state possible definitions, we distinguish two modeling approaches.

- **Detailed models** (without individual catalytic reactions):
  In this approach, catalysis occurs on the level of (small) subnetworks. In particular, individual reactions are not catalytic. For example, a simple catalytic mechanism (involving enzyme E, substrate S, and product P) is given by $\mathsf{E} + \mathsf{S} \leftrightarrow \mathsf{ES} \leftrightarrow \mathsf{EP} \leftrightarrow \mathsf{E} + \mathsf{P}$.

- **Coarse-grained models** (with individual catalytic reactions):
  In this approach, catalysis occurs on the level of individual reactions. For example, the catalytic mechanism above is written as $\mathsf{E} + \mathsf{S} \leftrightarrow \mathsf{E} + \mathsf{P}$ or $\mathsf{S} \overset{\mathsf{E}}{\leftrightarrow} \mathsf{P}$. Due to coarse-graining, catalysis cannot be identified from the stoichiometric matrix. Hence, for every catalytic reaction, the corresponding catalyst is specified explicitly.

For **detailed models**, one may call a growth mode autocatalytic if it contains an autocatalytic species or subnetwork and it is catalytically closed. Formal definitions and their comparison are beyond the scope of this work. For **coarse-grained models** (like the small model of a self-fabricating given in Fig 1), we give a formal definition of network autocatalysis.

Definition 6. For a coarse-grained model, let $\mathsf{Cat} \subseteq \mathsf{Rxn}$ be the set of catalytic reactions. A GM $v \in C_{\mathrm{g}}$ is *basically catalytic* (BC) if there is a catalytic reaction $r \in \mathrm{supp}(v) \cap \mathsf{Cat}$. Further, a GM $v \in C_{\mathrm{g}}$ is *catalytically closed* (CC) if, for every catalytic reaction $r \in \mathrm{supp}(v) \cap \mathsf{Cat}$, it holds that $(Nv)_s > 0$ for the corresponding catalyst $s \in \mathsf{Mol}$. Finally, a GM $v \in C_{\mathrm{g}}$ is *autocatalytic* (AC) if it is BC and CC.

A subset of reactions $S \subseteq \mathsf{Rxn}$ is *autocatalytic* (AC) if there exists an autocatalytic GM $v \in C_{\mathrm{g}}$ with $S = \mathrm{supp}(v)$. A nonempty subset of reactions is *minimally* autocatalytic (MAC) if it is AC and inclusion-minimal.

In the literature, a closure condition is also crucial in the definitions of "reflexive autocatalysis" [14–16] and "chemical organizations" [17–19].

We note the following points:

- AC is implied by two conditions: BC guarantees that there is at least one active catalytic reaction, and CC ensures that all active catalysts are produced.
- Catalytic closure is defined for fluxes, but it refers to concentrations via $x(v)_s = (Nv)_s/\mu(v)$. In particular, $(Nv)_s > 0$ implies $x(v)_s > 0$.

In addition to autocatalysis, also kinetics implies constraints on growth modes. In particular, a growth mode is kinetically consistent if all species involved in active reactions (not necessarily all species in the model) have nonzero associated concentrations.

Definition 7. A GM $v \in C_{\mathrm{g}}$ is *kinetically consistent* if, for every $r \in \mathrm{supp}(v) \subseteq \mathsf{Rxn}$ and $s \in \mathsf{Mol}$, $N_{sr} \neq 0$ implies $(Nv)_s > 0$.

We note the following points:

- Like catalytic closure, kinetic consistency is given by stoichiometry.
- Interestingly, kinetic consistency implies formal autocatalysis for all (not necessarily catalytic) species involved.

In the next subsection, we consider additional (capacity) constraints which often ensure catalytic closure.

2.5 Constraint-based models

For many systems, kinetic models are not yet available, and constraint-based models are used. Steady-state reaction rates (fluxes) v are considered as independent variables, that is, the non-linear dependence of the kinetics on the concentrations x is neglected.

Most importantly, catalytic processes imply linear capacity constraints for x and v. Additional constraints can be formulated for processes that are not catalytic (in the given model), e.g. lower bounds for concentrations or fluxes. In compact form, linear constraints can be can be written as

$$Ax + Bv \geq b$$

with $A \in \mathbb{R}^{m \times \mathsf{Mol}}$, $B \in \mathbb{R}^{m \times \mathsf{Rxn}}$, and $b \in \mathbb{R}^m$.

Altogether, *constraint-based* growth models involve steady-state, irreversibility, dry weight, and additional linear constraints,

$$\begin{aligned} &Nv = \mu\, x \geq 0, \quad v_{\mathcal{I}} \geq 0, \\ &\omega^T x = 1, \\ &Ax + Bv \geq b. \end{aligned} \tag{6}$$

Example (additional constraints). Recall the small model of a self-fabricating cell given in Fig 1. In addition to steady-state, irreversibility, and dry weight constraints, we consider capacity constraints for all catalysts (importers, enzymes, and the ribosome) and membrane constraints.

Let k_{cat} be the turnover number of the importers IG, IN and enzymes $\mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}$ and k_{el} be the elongation rate of the ribosome R. The resulting capacity constraints are given by

$$v_I \leq k_{\text{cat}}\, x_I \quad \text{and} \quad v_E \leq k_{\text{cat}}\, x_E \tag{7a}$$

for $I = \mathsf{IG}, \mathsf{IN}$ and $E = \mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}$ and

$$n_{\mathsf{I}}\, w_{\mathsf{IG}} + n_{\mathsf{I}}\, w_{\mathsf{IN}} + n_{\mathsf{E}}\, w_{\mathsf{EAA}} + n_{\mathsf{E}}\, w_{\mathsf{ELD}} + n_{\mathsf{E}}\, w_{\mathsf{EL}} + n_{\mathsf{R}}\, w_{\mathsf{R}} \leq k_{\text{el}}\, x_{\mathsf{R}}. \tag{7b}$$

The cell membrane area is formed by lipids L and importers IG and IN, leading to the (equality) constraint

$$A_{\mathsf{L}}\, x_{\mathsf{L}} + A_{\mathsf{I}}\, (x_{\mathsf{IG}} + x_{\mathsf{IN}}) = \frac{r}{\rho\, N_A}, \tag{7c}$$

where A_{L} and A_{I} denote the areas of lipids and importers, respectively, r denotes the surface-to-volume ratio, ρ denotes cell density, and N_A is Avogadro's number.

Additionally, we require that a minimum fraction α of the surface area is formed by lipids, leading to the (inequality) constraint

$$(1-\alpha)\, A_{\mathsf{L}}\, x_{\mathsf{L}} \geq \alpha\, A_{\mathsf{I}}\, (x_{\mathsf{IG}} + x_{\mathsf{IN}}). \tag{7d}$$

For the derivation of the membrane constraints, see Supporting Information B. As stated above, the additional constraints (7) can be summarized as $Ax + Bv \geq b$.

All constraints are based on realistic data. The parameter values for the small model of a self-fabricating cell are given in Table 2.

Table 2. Parameter values for the small model of a self-fabricating cell.

Symbol	Name	Value	Unit	Source*
n_{L}		7**	1	BNID 109367, [20]
n_{I}		646	1	MC: CPLX-157
n_{E}		325	1	BNID 108986
n_{R}		22608***	1	BNID 110218
ω_{G}	Molar mass glucose	180	$\text{g}\,\text{mol}^{-1}$	
ω_{N}	Molar mass ammonium	18	$\text{g}\,\text{mol}^{-1}$	
k_{cat}	Turnover number	79	s^{-1}	[21]
k_{el}	Elongation rate	8	$\text{AA}\ \text{s}^{-1}$	BNID 111689
A_{L}	Lipid area	0.5	nm^2	BNID 106993
A_{I}	Importer area	48	nm^2	BNID 114685
r	Surface-to-volume ratio	$9.3 - 2.8\,\mu\,\text{h}$	μm^{-1}	[22]
ρ	Cell density	290	$\text{g}\,\text{L}^{-1}$	BNID 104443
α	Minimum lipid fraction	0.5	1	[20]

* For values from BioNumbers [23] and MetaCyc [24] the database IDs are provided (BioNumbers: BNID; MetaCyc: MC).
** Estimated from the number of carbons in a typical *E. coli* lipid.
*** Number of AA in the ribosome multiplied by three to take into account RNA content.

In the final subsection, we introduce the basic "building blocks" of any possible flux distribution in a constraint-based model of cellular growth.

2.6 Elementary growth vectors

For given $\mu > 0$ and flux vector v, the concentration vector $x = Nv/\mu$ is not an independent variable. In particular, we have $Nv = \mu x \geq 0$ and $\omega^T Nv = \omega^T(\mu x) = \mu$. Hence, the constraint-based growth model (6) is equivalent to constraints just in terms of v,

$$\begin{aligned} &Nv \geq 0, \quad v_{\mathcal{I}} \geq 0, \\ &\omega^T Nv = \mu, \\ &(AN + \mu B)\, v \geq \mu\, b. \end{aligned}$$

The inequalities above (for the fluxes v) specify a polyhedron which suggests the following definition.

Definition 8. Let $\mu > 0$. *Growth vectors* (GVs) for the constraint-based growth model (6) are elements of the *growth polyhedron*

$$\begin{aligned} P_{\mathrm{g}}(\mu) = \big\{ v \in \mathbb{R}^{\mathsf{Rxn}} \,|\, & Nv \geq 0,\, v_{\mathcal{I}} \geq 0, \\ & \omega^T Nv = \mu, \\ & (AN + \mu B)\, v \geq \mu\, b \,\big\}. \end{aligned}$$

A GV $v \in P_{\mathrm{g}}(\mu)$ has an *associated* concentration vector $x(v) = Nv/\mu \in \mathbb{R}^{\mathsf{Mol}}_{\geq}$.

Elementary growth vectors (EGVs) are convex-conformally non-decomposable GVs and conformally non-decomposable elements of the recession cone

$$\begin{aligned} R = \big\{ v \in \mathbb{R}^{\mathsf{Rxn}} \,|\, & Nv = 0,\, v_{\mathcal{I}} \geq 0, \\ & Bv \geq 0 \,\big\}. \end{aligned}$$

For the definitions of polyhedra, convex-conformal non-decomposability, and the recession cone, we refer the reader to Subsection 3.1. Still, we also try to provide an intuitive understanding of Definition 8:

- GVs are fluxes given by stoichiometry, irreversibility, additional constraints, *and* growth rate. (If growth rate was not fixed, then the problem would become nonlinear, and the resulting set would not be a polyhedron.) In particular, GVs do not depend on concentrations. Still, GVs have an associated concentration vector.
- Every polyhedron can be written as the sum of a polytope and the recession cone. Moreover, every polyhedron has a unique set of conformal generators, and these are the convex-conformally non-decomposable vectors (generating a polytope) and the conformally non-decomposable elements of the recession cone. Thereby, a convex-conformally non-decomposable vector cannot be written as a convex sum of other vectors without cancellations [10].
- The recession cone of the growth polyhedron is a subset of the flux cone. In applications, the recession cone (as a subset of the flux cone) is often pointed (due to irreversibility).

EGMs and EGVs can be computed directly, by algorithms for vertex enumeration (such as `lrs` [25]), or indirectly (after transforming a system of inequalities to a system of equalities), by algorithms for EFM enumeration (such as `efmtool` [26]).

Again, we apply the general theory of elementary vectors and state the main result of this subsection.

Table 3. EGVs for the small model of a self-fabricating cell.

	EGV 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
mb G	>								>								>									
mb LD		>								>								>								
mb N			>								>								>							
mb AA				>								>								>						
cap IG									>	>	>	>	>	>	>	>									>	
cap IN																	>	>	>	>	>	>	>	>		>
cap EAA					>								>								>					
cap ELD						>								>								>				
cap EL							>								>								>			
cap R								>								>								>		
memb L	>	>	>	>	>	>	>	>																	>	>
G	*								*								*									
LD		*								*								*								
N			*								*								*							
AA				*								*								*						
μ									L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	H	H

There are 26 EGVs. Every EGV is characterized by its set of inactive inequality constraints. For every EGV, two constraints are inactive (that is, fulfilled with >), and all other constraints are active (fulfilled with =). For the list of inequality constraints, see Table S1 in Supporting Information C. Further, for every EGV, at most one of the associated concentrations of the metabolites {G, LD, N, AA} is nonzero (*). Finally, EGVs 1–8 exist for all growth rates, whereas EGVs 9–24 exist for low (L) growth rates and EGVs 25, 26 exist for high (H) growth rates. At the critical growth rate, EGVs 9–16 merge into EGV 25, and EGVs 17–24 merge into EGV 26.

Theorem 9. *Let v be a GV for growth rate $\mu > 0$. Then, there exist finite sets $E_0 \subseteq R$ and $E_\mu \subseteq P_{\mathrm{g}}(\mu)$ of EGVs such that*

$$v = \sum_{e \in E_0} e + \sum_{e \in E_\mu} \lambda_e \, e \quad \textit{with} \ \operatorname{sign}(e) \leq \operatorname{sign}(v),$$

$\lambda_e \geq 0$, and $\sum_{e \in E_\mu} \lambda_e = 1$. Moreover, $x(v) = \sum_{e \in E_\mu} \lambda_e \, x(e)$.

Proof. By Theorem 12 in Subsection 3.1. The growth polyhedron $P_{\mathrm{g}}(\mu)$ is a general polyhedron, and its elementary vectors are the convex-conformally non-decomposable vectors and the conformally non-decomposable vectors of its recession cone. □

Note that every GV is also a GM, and hence every nonzero GV is also a conformal sum of EGMs.

Example (EGVs). Recall the small model of a self-fabricating cell given in Fig 1 and consider the additional constraints (7). For convenience, the essential constraints are summarized in Table S1 in Supporting Information C. We compute EGVs as functions of growth rate, and we observe two distinct regimes. From zero to a critical growth rate $\mu_{\mathrm{crit}} = 1.21\,\mathrm{h}^{-1}$, we find 24 EGVs. From μ_{crit} to the maximum growth rate $\mu_{\mathrm{max}} = 1.27\,\mathrm{h}^{-1}$, we find 10 EGMs. Eight of these EGVs exist for all growth rates, 16 only in regime L ($0 < \mu < \mu_{\mathrm{crit}}$) and two only in regime H ($\mu_{\mathrm{crit}} < \mu < \mu_{\mathrm{max}}$). At μ_{crit}, two sets of eight EGVs (existing in regime L) merge into two EGVs (in regime H). At μ_{max}, all EGVs (existing in regime H) merge. EGVs and their characteristic properties are summarized in Table 3.

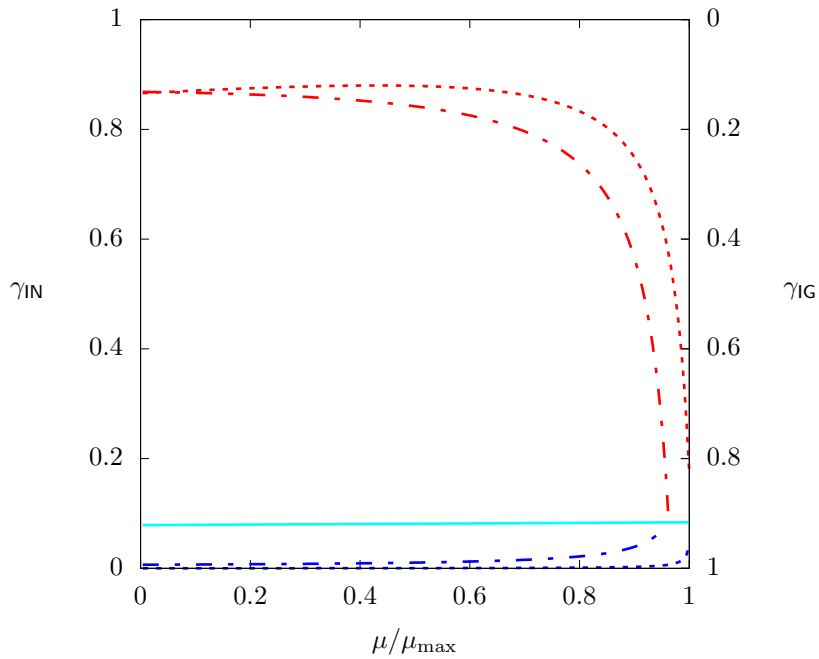


Fig 2. Relative contribution of ammonium uptake to growth, $\gamma_{\mathsf{IN}} = \frac{\omega_{\mathsf{N}} v_{\mathsf{IN}}}{\mu}$, as a function of growth rate for all EGVs (for the small model of a self-fabricating cell). Note that $\gamma_{\mathsf{IG}} = 1 - \gamma_{\mathsf{IN}}$. Colors correspond to classes of EGVs described in the main text.

Totally, there are 26 EGVs, and every EGV has full support. This is a consequence of the additional (capacity and membrane) constraints which ensure that all reactions are active. As a consequence, every EGV is autocatalytic, and the set of all reactions is the only autocatalytic set of reactions.

In contrast to EFMs, EGVs are not determined by their supports, but they are *characterized* by their sets of *inactive* inequality constraints. For every EGV, two constraints are inactive, and all other constraints are active. Further, for every EGV, the associated concentrations of lipids L and macromolecules $\{\mathsf{IG}, \mathsf{IN}, \mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}, \mathsf{R}\}$ are nonzero. However, at most one of the associated concentrations of the metabolites $\{\mathsf{G}, \mathsf{LD}, \mathsf{N}, \mathsf{AA}\}$ is nonzero. That is, no EGV is kinetically consistent.

In the example, the cell takes up only glucose and ammonium (and does not excrete any species). Hence, Eq (3) takes the form $\mu = \omega_{\mathsf{G}}\, v_{\mathsf{IG}} + \omega_{\mathsf{N}}\, v_{\mathsf{IN}}$, and we introduce the relative contributions of the import fluxes to growth rate,

$$\gamma_{\mathsf{IG}} = \frac{\omega_{\mathsf{G}}\, v_{\mathsf{IG}}}{\mu} \quad \text{and} \quad \gamma_{\mathsf{IN}} = \frac{\omega_{\mathsf{N}}\, v_{\mathsf{IN}}}{\mu}$$

with $\gamma_{\mathsf{IG}} + \gamma_{\mathsf{IN}} = 1$. The relative contribution of ammonium uptake to growth as a function of growth rate is visualized in Fig 2.

As it turns out, EGVs can be grouped into six classes regarding their uptake/storage behavior, namely into classes

1a (EGVs 1, 2),	2a (EGV 3),	3a (EGVs 4–8),
1b (EGVs 9, 10; 17, 18),	2b (EGVs 11; 19),	3b (EGVs 12–16; 20–24; 25, 26).

Thereby, EGVs in classes 1a, 2a, and 3a exist for all growth rates, whereas EGVs in classes 1b, 2b, and 3b exist either in regime L or H.

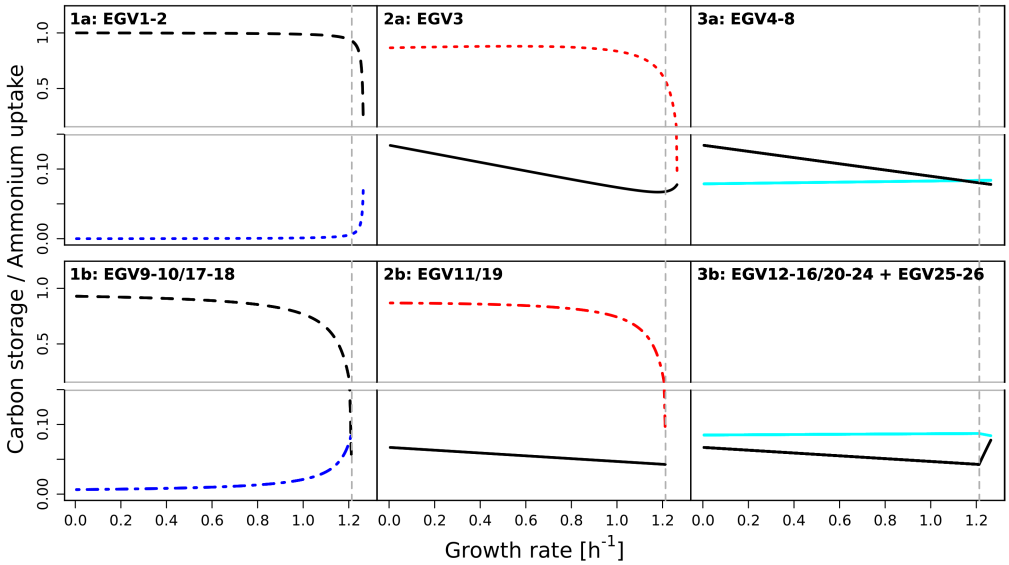


Fig 3. Ammonium uptake $\gamma_{\mathsf{IN}} = \frac{\omega_{\mathsf{N}} v_{\mathsf{IN}}}{\mu}$ and carbon storage $s_{\mathsf{C}} = \omega_{\mathsf{G}} x_{\mathsf{G}} + \omega_{\mathsf{LD}} x_{\mathsf{LD}} + \omega_{\mathsf{L}} x_{\mathsf{L}}$ as a function of growth rate for all EGVs (for the small model of a self-fabricating cell). Panels correspond to classes of EGVs described in the main text. Ammonium uptake in blue/red/turquoise (see also Fig 2). Carbon storage in black (dashed and solid: $x_{\mathsf{L}} > 0$; dashed: $x_{\mathsf{G}} = 0$, $x_{\mathsf{LD}} > 0$ or $x_{\mathsf{G}} > 0$, $x_{\mathsf{LD}} = 0$; solid: $x_{\mathsf{G}} = x_{\mathsf{LD}} = 0$.)

Fig 2 highlights the uptake behaviors of the classes of EGVs. In classes 3a and 3b (turquoise), we observe an (almost) constant ammonium uptake of $\gamma_{\mathsf{IN}} = 0.091$ (independently of growth rate). This corresponds to equal uptake rates of glucose and ammonium. In classes 1a (blue, dotted) and 1b (blue, dashed), ammonium uptake rate is smaller than glucose uptake rate, whereas in classes 2a (red, dotted) and 2b (red, dashed), it is larger. However, all EGVs approach balanced uptake for high growth rates. Classes 1b and 2b (dashed, existing in regime L) reach balanced uptake at μ_{crit}, whereas classes 1a and 2a (dotted, existing for all growth rates) reach it at μ_{max}.

For classes with unbalanced uptake behavior, there is an asymmetry between ammonium and glucose uptake at $\mu \to 0$. For smaller ammonium uptake, $\gamma_{\mathsf{IN}} \to 0$, whereas for larger ammonium uptake $\gamma_{\mathsf{IN}} \to 0.87$ (instead of $\gamma_{\mathsf{IN}} \to 1$). This asymmetry is related to the membrane equality constraint (7c).

Next, we analyze the correlation between ammonium uptake γ_{IN} and the mass fraction $s_{\mathsf{C}} = \omega_{\mathsf{G}}\, x_{\mathsf{G}} + \omega_{\mathsf{LD}}\, x_{\mathsf{LD}} + \omega_{\mathsf{L}}\, x_{\mathsf{L}}$, corresponding to carbon storage in glucose and, most importantly, in lipids. As expected, when ammonium is limiting, then glucose is accumulated or rerouted towards lipid synthesis (instead of amino acid synthesis). Ammonium uptake γ_{IN} and carbon storage s_{C} as a function of growth rate are visualized in Fig 3.

In particular, Fig 3 highlights the uptake/storage behaviors of the six classes of EGVs defined above. The left panels correspond to EGVs in classes 1a and 1b. For low growth rate, they display low ammonium uptake and high carbon storage. The middle panels correspond to classes 2a and 2b, where EGVs display high ammonium uptake and low carbon storage. In fact, ammonium is accumulated. Finally, the right panels correspond to classes 3a and 3b. EGVs display balanced uptake behavior and low carbon storage.

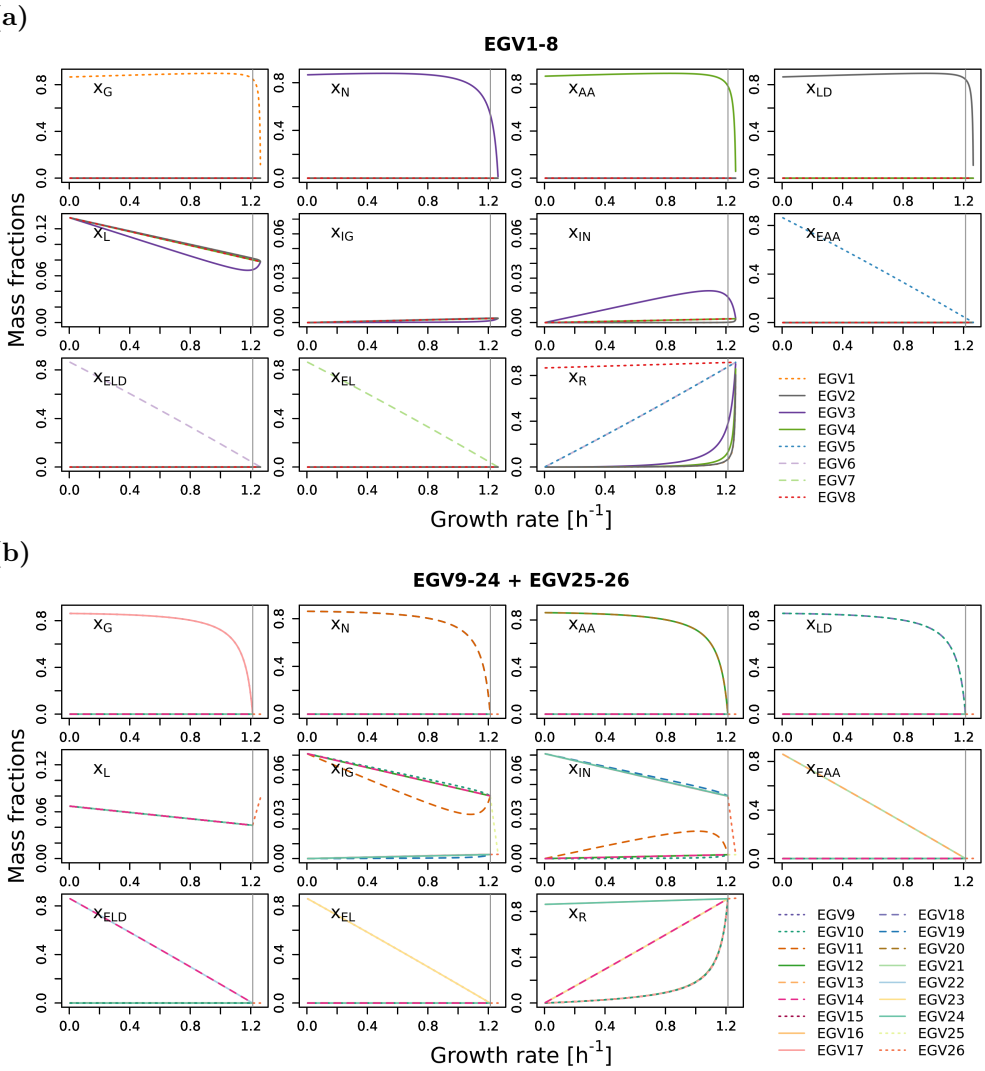


Fig 4. Mass fractions for (a) the 8 EGVs that exist for all growth rates and (b) the 16 EGVs that exist in regime L plus the 2 EGVs that exist in regime H.

In Supporting Information C, we provide details about the individual EGVs. For each of the 26 EGVs, we display the fluxes (of the 11 reactions) and the associated mass fractions (of the 11 species), as a function of growth rate. See Figs S1 and S2.

Conversely, for each of the 11 species, we show their mass fractions in the 26 EGVs, and analogously, for each of the 11 reactions, we show their fluxes in the 26 EGVs. See Fig 4 in the main text and Fig S3 in Supporting Information C.

3 Methods

3.1 Elementary vectors in polyhedral geometry

For the objects of polyhedral geometry (subspaces, cones, polyhedra), there is no *unique* minimal set of generators, in general. However, *elementary vectors* (EVs) form unique sets of *conformal* generators [10, Section 3.4]. For linear subspaces and s-cones (arising from linear subspaces and nonnegativity constraints), elementary vectors are the support-minimal (SM) vectors; for general polyhedral cones, they are the conformally non-decomposable (cND) vectors; and for general polyhedra, they are the convex-conformally non-decomposable (ccND) vectors plus the cND vectors of the recession cone [10].

Below, we summarize basic definitions and results for s-cones, general polyhedral cones, and polyhedra. Before that, we introduce the mathematical notation.

Notation. We denote the positive real numbers by $\mathbb{R}_>$ and the nonnegative real numbers by $\mathbb{R}_\geq$. Let I be an index set; often $I = \{1, \ldots, n\}$, and $x \in \mathbb{R}^I$ stands for $x \in \mathbb{R}^n$. For $x \in \mathbb{R}^I$, we write $x > 0$ if $x \in \mathbb{R}^I_>$, $x \geq 0$ if $x \in \mathbb{R}^I_\geq$, and we denote its *support* by $\operatorname{supp}(x) = \{i \in I \mid x_i \neq 0\}$. Recall that a nonzero vector $x \in X \subseteq \mathbb{R}^I$ is *support-minimal* (in X) if, for all nonzero $x' \in X$, $\operatorname{supp}(x') \subseteq \operatorname{supp}(x)$ implies $\operatorname{supp}(x') = \operatorname{supp}(x)$. For $x \in \mathbb{R}^I$, we define its *sign vector* $\operatorname{sign}(x) \in \{-, 0, +\}^I$ by applying the sign function component-wise, that is, $\operatorname{sign}(x)_i = \operatorname{sign}(x_i)$ for $i \in I$. The relations $0 < -$ and $0 < +$ on $\{-, 0, +\}$ induce a partial order on $\{-, 0, +\}^I$: for $X, Y \in \{-, 0, +\}^I$, we write $X \leq Y$ if the inequality holds component-wise. For $x, y \in \mathbb{R}^I$, we denote the component-wise product by $x \circ y \in \mathbb{R}^I$, that is, $(x \circ y)_i = x_i y_i$. For $x \in \mathbb{R}^I$ and subindex set $\mathcal{I} \subseteq I$, we write $x_\mathcal{I} \in \mathbb{R}^\mathcal{I}$ for the corresponding subvector.

S-cones

Given a linear subspace $S \subseteq \mathbb{R}^n$ and a subindex set $\mathcal{I} \subseteq \{1, \ldots, n\}$, an *s-cone* (special cone, subspace cone) is given by $C(S, \mathcal{I}) = \{x \in \mathbb{R}^n \mid x \in S,\ x_\mathcal{I} \geq 0\}$. Note that a linear subspace is an s-cone, $S = C(S, \emptyset)$.

A vector $x \in C(S, \mathcal{I})$ is *elementary* if it is SM. (For linear subspaces, the definition of elementary vectors (EVs) as SM vectors was given in [27].)

The following result is fundamental. See [10, Theorem 3] based on [27, Theorem 1].

Theorem 10. *Let $C(S, \mathcal{I})$ be an s-cone. Every nonzero vector $x \in C(S, \mathcal{I})$ is a conformal sum of EVs. That is, there exists a finite set E of EVs such that*

$$x = \sum_{e \in E} e \quad \textit{with} \ \operatorname{sign}(e) \leq \operatorname{sign}(x).$$

The set E can be chosen such that $|E| \leq \dim(S)$ and $|E| \leq |\operatorname{supp}(x)|$.

Example. Consider a minimal metabolic network, involving one internal molecular species X, where substrate is taken up, and two products are excreted, that is, $(S) \to X$, $X \to (P_1)$, $X \to (P_2)$, leading to the stoichiometric matrix

$$N = (1, -1, -1).$$

The flux cone $C_\mathrm{f} = \{v \in \mathbb{R}^3 \mid Nv = 0 \text{ and } v \geq 0\}$ is an s-cone, $C_\mathrm{f} = C(S, \mathcal{I})$ with $S = \ker N$ and $\mathcal{I} = \{1, 2, 3\}$ (since all reactions are assumed to be irreversible). The cone is generated by two (support-minimal) EVs, $e^2 = (1, 1, 0)^T$ and $e^3 = (1, 0, 1)^T$. The cone is indicated in Figure 5 below.

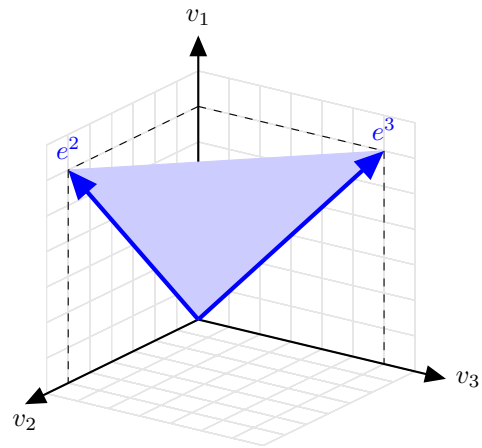


Figure 5. An s-cone.

General polyhedral cones

Let C be a polyhedral cone, that is, $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ for some $A \in \mathbb{R}^{m \times n}$. A nonzero vector $x \in C$ is *conformally non-decomposable* (cND) if, for all nonzero $x^1, x^2 \in C$ with $\operatorname{sign}(x^1), \operatorname{sign}(x^2) \leq \operatorname{sign}(x)$, the decomposition $x = x^1 + x^2$ implies $x^1 = \lambda x^2$ with $\lambda > 0$. A vector $x \in C$ is *elementary* if it is cND.

By defining EVs as cND vectors (instead of SM vectors), Theorem 10 can be extended to general polyhedral cones. See [10, Theorem 8].

Theorem 11. *Let $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ be a polyhedral cone. Every nonzero vector $x \in C$ is a conformal sum of EVs. That is, there exists a finite set E of EVs such that*

$$x = \sum_{e \in E} e \quad \textit{with} \ \operatorname{sign}(e) \leq \operatorname{sign}(x).$$

The set E can be chosen such that $|E| \leq \dim(C)$ and $|E| \leq |\operatorname{supp}(x)| + |\operatorname{supp}(Ax)|$.

Example. Consider a minimal growth model, involving three internal molecular species, a precursor P, a metabolic "enzyme" E, and the ribosome R, where the syntheses of P, E, and R are catalyzed, that is, $(S) \xrightarrow{E} P$, $P \xrightarrow{R} E$, $P \xrightarrow{R} R$, leading to the stoichiometric matrix

$$N = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The growth cone $C_{\mathrm{g}} = \{v \in \mathbb{R}^3 \mid Nv \geq 0 \text{ and } v \geq 0\}$ is a general polyhedral cone, that is, $C_{\mathrm{g}} = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ with $A = \binom{N}{\mathrm{I}_3}$ where I_3 is the identity matrix. The cone is generated by three (conformally non-decomposable) EVs, $e^1 = (1,0,0)^T$, $e^2 = (1,1,0)^T$, and $e^3 = (1,0,1)^T$. (Note that only e^1 is SM.) The cone is indicated in Figure 6 below.

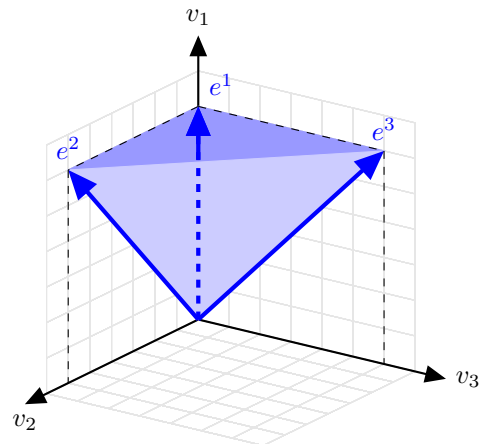


Figure 6. A polyhedral cone (that is not an s-cone).

Polyhedra

Let P be a polyhedron, that is, $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. A vector $x \in P$ is *convex-conformally non-decomposable* (ccND) if for all $x^1, x^2 \in P$ with $\operatorname{sign}(x^1), \operatorname{sign}(x^2) \leq \operatorname{sign}(x)$ and $0 < \lambda < 1$, the decomposition $x = \lambda x^1 + (1-\lambda)x^2$ implies $x^1 = x^2$.

Let $R = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ be the *recession cone* of P. A vector $e \in P \cup R$ is *elementary* (an EV of P) if $e \in P$ is ccND or $e \in R$ is cND.

Ultimately, Theorem 10 can be extended to general polyhedral cones. See [10, Theorem 13].

Theorem 12. *Let $P = \{x \mid Ax \geq b\}$ be a polyhedron and $R = \{x \mid Ax \geq 0\}$ its recession cone. Every vector $x \in P$ is a conformal sum of EVs. That is, there exist finite sets $E_0 \subseteq R$ and $E_1 \subseteq P$ of EVs such that*

$$x = \sum_{e \in E_0} e + \sum_{e \in E_1} \lambda_e e \quad \textit{with } \operatorname{sign}(e) \leq \operatorname{sign}(x),$$

$\lambda_e \geq 0$, and $\sum_{e \in E_1} \lambda_e = 1$. (Hence, $|E_1| \geq 1$.)

The set $E = E_0 \cup E_1$ can be chosen such that $|E| \leq \dim(P) + 1$ and $|E| \leq |\operatorname{supp}(x)| + |\operatorname{supp}(Ax)| + 1$.

3.2 Software

Simulations were performed in `Python` 3.7.9 using the package `efmtool` 0.2.0 [26]. Fig 1 was created with `BioRender`. The remaining figures were created with `R` 4.0.2. All code is available on `GitHub` at https://github.com/diana-sz/EGMs.

4 Discussion

In traditional models of cellular growth, elementary flux modes and vectors (EFMs [4,5] and EFVs [6,7]) allow the unbiased characterization of all feasible metabolic phenotypes in terms of unique functional units, namely EFMs (in flux cones) and EFVs (in flux polyhedra). More specifically, a certain cellular function can only be performed if there is a corresponding EFM or EFV. However, EFMs and EFVs depend on a fixed biomass composition, which is specified as a biomass "reaction" in addition to the actual metabolic reactions. As a consequence, EFMs and EFVs are not able to describe cellular resource reallocations, in particular, upon changes of environmental conditions.

In more comprehensive models, the biomass reaction is replaced by explicit synthesis reactions for all (model-specific) macromolecules. In such next-generation models (as used in RBA [3]), biomass composition is no longer fixed, but results from the expression of all proteins (and the synthesis of other macromolecules). In fact, the concentrations $x = Nv/\mu$ of all molecular species are determined by the stoichiometric matrix N, the fluxes v, and growth rate μ, cf. Eq (4).

In this work, we developed a mathematical theory that allows the unbiased characterization not only of all feasible metabolic phenotypes, but also of all feasible biomass compositions in comprehensive, next-generation models. In analogy to EFMs and EFVs, we introduced elementary growth modes and vectors (EGMs and EGVs) which can be interpreted as unique functional units of self-fabrication.

Since EGMs are the basic "building blocks" of any possible flux distribution in *purely stoichiometric* models of cellular growth, they are typically not "self-fabricating" themselves. In particular, they do not reflect implications from autocatalysis, namely that all catalysts of active reactions are synthesized. Still, additional capacity constraints of the form $v_E \leq k_{\text{cat}}\, x_E$ ensure nonzero enzyme concentrations x_E for nonzero fluxes v_E. As a consequence, EGVs in *constraint-based* models are often autocatalytic, as in our running example.

Example. We used our theoretical concepts to study the storage behavior in the small model of a self-fabricating cell given in Fig 1. In particular, we identified 26 EGVs in three classes regarding their storage behaviors (cf. Fig 3):

1. carbon accumulates in the form of glucose or lipids;
2. nitrogen accumulates in the form of ammonium;
3. ammonium and glucose are stored together in the form of proteins.

We observed that carbon accumulation (class 1) goes hand in hand with low ammonium uptake. In fact, a common strategy to increase cellular carbon content in biotechnological applications is to limit the source of nitrogen [28–36]. Restricting nitrogen limits the synthesis of amino acids and proteins, so the remaining carbon source can be rerouted towards lipid and carbohydrate synthesis, which does not require nitrogen. Remarkably, this effect is a feature of a constraint-based metabolic model and does not require active regulation. Experimentally, nitrogen limitation sometimes leads to the accumulation of fat [28–36] or starch with [32,34] or without [37] a simultaneous lipid accumulation. All of these possibilities are covered by EGVs and agree with our observations. Of course, which of these behaviors is realized is the result of cellular regulation and hence outside the scope of our analysis.

Further, nitrogen accumulation (class 2) takes place when ammonium uptake exceeds glucose import. In fact, most of the ammonium accumulates since protein synthesis is limited by insufficient glucose import. Still, carbohydrate concentrations

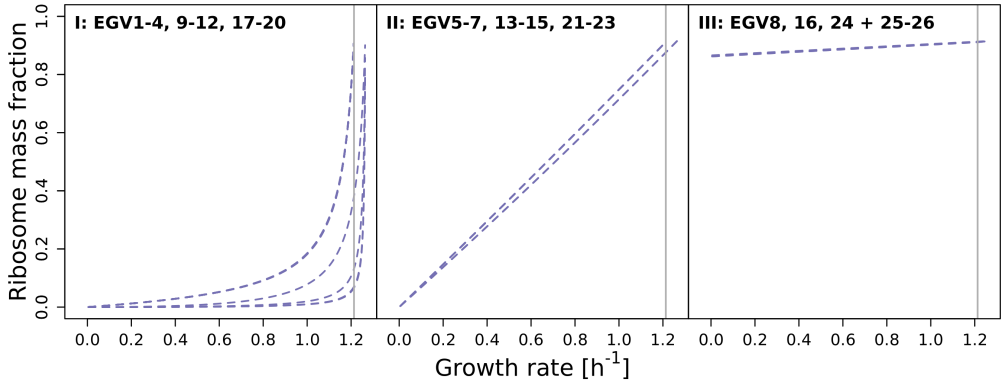


Fig 7. Ribosome mass fraction as a function of growth rate for all EGVs (for the small model of a self-fabricating cell). Panels correspond to classes of EGVs described in the main text.

remain at normal levels. Such phenotypes have been observed in species such as diatoms, foraminifers, and fungi [38].

The last behavior (class 3) occurs when ammonium and glucose uptake are balanced. The individual EGVs correspond to situations where a certain protein's abundance is larger than required by the flux level of the catalyzed reaction [39, 40]. This mirrors the fact that enzyme expression is not necessarily indicative of flux levels.

At high growth rate, storage becomes exceedingly expensive for a cell since most resources are required to support growth [39]. Consistently, all EGVs show decreasing storage with increasing growth rate. At maximum growth rate, all EGVs (that exist for high growth rate) merge. Uptake is balanced, and there is no storage (except lipids in the membrane).

Alternatively, the 26 EGVs can be grouped into three classes regarding their ribosome content, namely into classes

I: EGVs 1–4;	II: EGVs 5–7;	III: EGVs 8;
9–12; 17–20,	13–15; 21–23,	16; 24; 25, 26.

Figure 7 highlights the ribosome mass fraction $\omega_{\mathsf{R}}\, x_{\mathsf{R}}$ (as a function of growth rate) of the three classes.

In class II, the ribosome content is linear as a function of growth rate. This behavior has been observed experimentally (however, with an offset at zero growth rate) [41] and explained theoretically (thereby assuming growth rate-independent "housekeeping" proteins) [42].

In class III, we observe an (almost) constant (and high) ribosome mass fraction of $\omega_{\mathsf{R}}\, x_{\mathsf{R}} = 0.9$ (independently of growth rate). EGVs in this class may describe exactly those ribosomes that are required for the production of growth rate-independent "housekeeping" proteins (and explain the offset at zero growth rate).

In class I, the ribosome content is non-linear at high growth rates. EGVs in this class have nonzero associated metabolite concentrations and may explain a potential non-linear behavior.

Altogether, the experimentally observed ribosome mass fraction [41] can be understood as a convex, conformal sum of EGVs from the three classes.

Elementary vectors and computation. Whereas EFMs are *support-minimal* vectors of the flux cone (a special cone), already EFVs are *conformally non-decomposable* (and not support-minimal) vectors of a flux polyhedron. Also EGMs of the growth cone (a general polyhedral cone) and EGVs of a growth polyhedron are conformally non-decomposable vectors. Thus, conformal non-decomposability (and not support-minimality) is the key feature that defines elementary vectors in computational models of cellular growth, see also [7].

In terms of computation, EGM and EGV analysis (just like EFM and EFV analysis) suffer from a combinatorial explosion in the number of elementary vectors with the size of the network. As with EFMs and EFVs, one may focus on the enumeration of subsets of EGMs and EGVs or consider extra constraints such as thermodynamic feasibility.

Acknowledgments

We thank Christoph Flamm and Ralf Steuer for discussions on a previous version of the manuscript. SM was supported by the Austrian Science Fund (FWF), project P33218. DS and JZ acknowledge support from the Austrian Center of Industrial Biotechnology (acib). The COMET center *acib – Next Generation Bioproduction* is funded by BMK, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria, and Vienna Business Agency in the framework of COMET Competence Centers for Excellent Technologies. The COMET Funding Program is managed by the Austrian Research Promotion Agency (FFG).

References

1. Klamt S, Hädicke O, Kamp Av. Stoichiometric and Constraint-Based Analysis of Biochemical Reaction Networks. In: Benner P, Findeisen R, Flockerzi D, Reichl U, Sundmacher K, editors. Large-Scale Networks in Engineering and Life Sciences. Modeling and Simulation in Science, Engineering and Technology. Springer International Publishing; 2014. p. 263–316.

2. Lewis NE, Nagarajan H, Palsson BØ. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. Nature Reviews Microbiology. 2012;10(4):291–305. doi:10.1038/nrmicro2737.

3. Goelzer A, Fromion V. Bacterial growth rate reflects a bottleneck in resource allocation. Biochimica et Biophysica Acta (BBA) - General Subjects. 2011;1810(10):978–988. doi:10.1016/j.bbagen.2011.05.014.

4. Schuster S, Hilgetag C. On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst. 1994;2:165–182. doi:10.1142/S0218339094000131.

5. Schuster S, Hilgetag C, Woods JH, Fell DA. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J Math Biol. 2002;45(2):153–181. doi:10.1007/s002850200143.

6. Urbanczik R. Enumerating constrained elementary flux vectors of metabolic networks. IET Syst Biol. 2007;1(5):274–279.

7. Klamt S, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, et al. From elementary flux modes to elementary flux vectors: Metabolic

pathway analysis with arbitrary linear flux constraints. PLoS Comput Biol. 2017;13(4):e1005409. doi:10.1371/journal.pcbi.1005409.

8. Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. Environmental Microbiology. 2002;4(3):133–140. doi:10.1046/j.1462-2920.2002.00282.x.

9. de Groot DH, Hulshof J, Teusink B, Bruggeman FJ, Planqué R. Elementary Growth Modes provide a molecular description of cellular self-fabrication. PLoS Comput Biol. 2020;16(1):e1007559.

10. Müller S, Regensburger G. Elementary vectors and conformal sums in polyhedral geometry and their relevance for metabolic pathway analysis. Front Genet. 2016;7(90):1–11. doi:10.3389/fgene.2016.00090.

11. Barenholz U, Davidi D, Reznik E, Bar-On Y, Antonovsky N, Noor E, et al. Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. eLife. 2017;6. doi:10.7554/elife.20667.

12. Blokhuis A, Lacoste D, Nghe P. Universal motifs and the diversity of autocatalytic systems. Proceedings of the National Academy of Sciences. 2020;117(41):25230–25236. doi:10.1073/pnas.2013527117.

13. Andersen JL, Flamm C, Merkle D, Stadler PF. Defining Autocatalysis in Chemical Reaction Networks. J Syst Chem. 2020;8:121–133.

14. Kauffman SA. Autocatalytic sets of proteins. Journal of Theoretical Biology. 1986;119(1):1–24. doi:10.1016/s0022-5193(86)80047-9.

15. Steel M. The emergence of a self-catalysing structure in abstract origin-of-life models. Applied Mathematics Letters. 2000;13(3):91–95. doi:10.1016/s0893-9659(99)00191-3.

16. Hordijk W, Steel M. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. Journal of Theoretical Biology. 2004;227(4):451–461. doi:10.1016/j.jtbi.2003.11.020.

17. Fontana W, Buss LW. "The arrival of the fittest": Toward a theory of biological organization. Bulletin of Mathematical Biology. 1994;56(1):1–64. doi:10.1007/bf02458289.

18. Kaleta C, Centler F, Dittrich P. Analyzing Molecular Reaction Networks: From Pathways to Chemical Organizations. Molecular Biotechnology. 2006;34(2):117–124. doi:10.1385/mb:34:2:117.

19. Dittrich P, di Fenizio PS. Chemical Organisation Theory. Bulletin of Mathematical Biology. 2007;69(4):1199–1231. doi:10.1007/s11538-006-9130-8.

20. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The lipid bilayer. Garland Science; 2002.

21. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry. 2011;50(21):4402–4410. doi:10.1021/bi2002289.

22. Harris LK, Theriot JA. Surface Area to Volume Ratio: A Natural Variable for Bacterial Morphogenesis. Trends in Microbiology. 2018;26(10):815–832. doi:https://doi.org/10.1016/j.tim.2018.04.008.

23. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Research. 2009;38(suppl_1):D750–D753. doi:10.1093/nar/gkp889.

24. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. Nucleic acids research. 2018;46(D1):D633–D639.

25. Avis D, Fukuda K. A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra. Discret Comput Geom. 1992;8:295–313. doi:10.1007/BF02293050.

26. Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics. 2008;24(19):2229–2235.

27. Rockafellar RT. The elementary vectors of a subspace of R^N. In: Combinatorial Mathematics and its Applications (Proc. Conf., Univ. North Carolina, Chapel Hill, N.C., 1967). Chapel Hill, N.C.: Univ. North Carolina Press; 1969. p. 104–127.

28. Calvey CH, Su YK, Willis LB, McGee M, Jeffries TW. Nitrogen limitation, oxygen limitation, and lipid accumulation in Lipomyces starkeyi. Bioresource technology. 2016;200:780–788.

29. Zhang H, Wu C, Wu Q, Dai J, Song Y. Metabolic flux analysis of lipid biosynthesis in the yeast Yarrowia lipolytica using 13C-labled glucose and gas chromatography-mass spectrometry. PloS one. 2016;11(7):e0159187.

30. Kolouchová I, Mat'átková O, Sigler K, Masák J, Řezanka T. Lipid accumulation by oleaginous and non-oleaginous yeast strains in nitrogen and phosphate limitation. Folia microbiologica. 2016;61(5):431–438.

31. Griffiths MJ, van Hille RP, Harrison ST. The effect of nitrogen limitation on lipid productivity and cell composition in Chlorella vulgaris. Applied microbiology and biotechnology. 2014;98(5):2345–2356.

32. Pancha I, Chokshi K, George B, Ghosh T, Paliwal C, Maurya R, et al. Nitrogen stress triggered biochemical and morphological changes in the microalgae Scenedesmus sp. CCNM 1077. Bioresource technology. 2014;156:146–154.

33. Jiang Y, Yoshida T, Quigg A. Photosynthetic performance, lipid production and biomass composition in response to nitrogen limitation in marine microalgae. Plant Physiology and Biochemistry. 2012;54:70–77.

34. Pruvost J, Van Vooren G, Le Gouic B, Couzinet-Mossion A, Legrand J. Systematic investigation of biomass and lipid productivity by microalgae in photobioreactors for biodiesel application. Bioresource technology. 2011;102(1):150–158.

35. Xiong W, Liu L, Wu C, Yang C, Wu Q. 13C-Tracer and Gas Chromatography-Mass Spectrometry Analyses Reveal Metabolic Flux Distribution in the Oleaginous Microalga Chlorella protothecoides. Plant Physiology. 2010;154(2):1001–1011. doi:10.1104/pp.110.158956.

36. Hsieh CH, Wu WT. Cultivation of microalgae for oil production with a cultivation strategy of urea limitation. Bioresource technology. 2009;100(17):3921–3926.

37. Pleissner D, Eriksen NT. Effects of phosphorous, nitrogen, and carbon limitation on biomass composition in batch and continuous flow cultures of the heterotrophic dinoflagellate Crypthecodinium cohnii. Biotechnology and bioengineering. 2012;109(8):2005–2016.

38. Kamp A, Høgslund S, Risgaard-Petersen N, Stief P. Nitrate storage and dissimilatory nitrate reduction by eukaryotic microbes. Frontiers in microbiology. 2015;6:1492.

39. O'Brien EJ, Utrilla J, Palsson BO. Quantification and classification of E. coli proteome utilization and unused protein costs across environments. PLoS computational biology. 2016;12(6):e1004998.

40. Yang L, Yurkovich JT, Lloyd CJ, Ebrahim A, Saunders MA, Palsson BO. Principles of proteome allocation are revealed using proteomic data and genome-scale models. Scientific reports. 2016;6(1):1–8.

41. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. Science. 2010;330(6007):1099–1102.

42. Scott M, Klumpp S, Mateescu EM, Hwa T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol Syst Biol. 2014;10:747.

43. de Jong H, Casagranda S, Giordano N, Cinquemani E, Ropers D, Geiselmann J, et al. Mathematical modelling of microbes: metabolism, gene expression and growth. Journal of The Royal Society Interface. 2017;14(136):20170502. doi:10.1098/rsif.2017.0502.

Supporting Information

A The dynamic model of cellular growth

We derive the dynamic model of cellular growth (1), studied in the main text.

We denote fundamental objects and quantities as follows:

Mol	set of molecular species	
Rxn	set of chemical reactions	
$N \in \mathbb{R}^{\mathsf{Mol}\times\mathsf{Rxn}}$	stoichiometric matrix	(unit: 1)
$\omega \in \mathbb{R}^{\mathsf{Mol}}_{>}$	molar masses	(unit: $\mathrm{g\,mol^{-1}}$)
$X \in \mathbb{R}^{\mathsf{Mol}}_{\geq}$	amounts of substance	(unit: mol)
$R(\cdot) \in \mathbb{R}^{\mathsf{Rxn}}$	reaction rates (extensive)	(unit: $\mathrm{mol\,h^{-1}}$)

The chemical reactions induce the dynamical system

$$\frac{\mathrm{d}X}{\mathrm{d}t} = NR. \tag{8}$$

We define dry weight (correctly: dry mass),

$$M = \sum_i \omega_i X_i = \omega^T X, \quad \text{(unit: g)} \tag{9a}$$

the intensive quantities

$$x = \frac{X}{M} \in \mathbb{R}^{\mathsf{Mol}}_{\geq}, \quad (\text{unit: } \mathrm{mol\,g^{-1}}) \tag{9b}$$

$$v = \frac{R}{M} \in \mathbb{R}^{\mathsf{Rxn}}, \quad (\text{unit: } \mathrm{mol\,g^{-1}\,h^{-1}}) \tag{9c}$$

and growth rate

$$\mu = \frac{1}{M}\frac{\mathrm{d}M}{\mathrm{d}t}. \quad (\text{unit: } \mathrm{h^{-1}}) \tag{9d}$$

Thereby, we use mass instead of volume to define the "concentrations" x, the (intensive) reaction rates v, and growth rate μ. In practice, cellular composition is often given in the unit $\mathrm{mol\,g^{-1}}$ (dry weight).

Finally, we recall the chain rule (of differentiation),

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{X}{M} = \frac{1}{M}\frac{\mathrm{d}X}{\mathrm{d}t} - \frac{X}{M^2}\frac{\mathrm{d}M}{\mathrm{d}t}.$$

Equations (8), (9), and the chain rule yield the dynamic model of cellular growth:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Nv(x) - \mu\, x$$

and

$$\omega^T x = 1.$$

Thereby, we assume constant cell density. Recall that reaction rates depend on (volumetric) concentrations X/V,

$$R = R(X/V) = R(M/V \cdot X/M) = R(\rho\, x)$$

with volume V (unit: L) and cell density $\rho = \frac{M}{V}$ (unit: $\mathrm{g\,L^{-1}}$). Hence,

$$v = \frac{R}{M} = \frac{R(\rho\, x)}{M} = v(\rho\, x).$$

For constant cell density ρ, $v = v(x)$ only depends on concentrations.

For alternative derivations, see e.g. [43] or [9].

By multiplying the mass balance equation with a vector $c \in \mathbb{R}^{\mathsf{Mol}}$, we obtain

$$\frac{\mathrm{d}(c^T x)}{\mathrm{d}t} = c^T N v(x) - \mu\,(c^T x).$$

We highlight two observations that hold for any model of cellular growth.

Fact (conservation laws). In a model of cellular growth, there cannot be any conservation laws. In mathematical terms, $\ker N^T \cap \mathbb{R}^{\mathsf{Mol}}_{\geq} = \{0\}$.

To see this, assume $c^T N = 0$ with $0 \neq c \geq 0$, for example, assume $c_1 = c_2 = 1$ and $c_i = 0$, otherwise. Then, $\frac{\mathrm{d}(c^T x)}{\mathrm{d}t} = \frac{\mathrm{d}(x_1+x_2)}{\mathrm{d}t} = -\mu(x_1 + x_2) \leq 0$, and $\mu > 0$ implies $x_1 = x_2 = 0$ at steady state.

Fact (dependent concentrations). In a model of cellular growth, there can be dependent concentrations. In mathematical terms, $\ker N^T \neq \{0\}$.

To see this, assume $c^T N = 0$ with $0 \neq c$, for example, assume $c_1 = 1$, $c_2 = -1$, and $c_i = 0$, otherwise. Then, $\frac{\mathrm{d}(c^T x)}{\mathrm{d}t} = \frac{\mathrm{d}(x_1-x_2)}{\mathrm{d}t} = -\mu(x_1 - x_2)$, and $\mu > 0$ implies $x_1 = x_2$ at steady state.

B Example: membrane constraints

For the small model of a self-fabricating cell studied in the main text, we derive the membrane constraints (7c) and (7d).

The cell membrane area A is formed by lipids L and importers IG and IN,

$$A = A_{\mathsf{L}} \cdot \#\mathsf{L} + A_{\mathsf{I}} \cdot (\#\mathsf{IG} + \#\mathsf{IN}),$$

where A_{L} and A_{I} denote the areas of lipids and importers, respectively, and $\#X$ denotes the number of molecule X. After division by Avogadro's number N_A, we have

$$\frac{A}{N_A} = A_{\mathsf{L}}\, s_{\mathsf{L}} + A_{\mathsf{I}}\,(s_{\mathsf{IG}} + s_{\mathsf{IN}}),$$

where $s_X = \frac{\#X}{N_A}$ denotes the amount of substance. Further, after division by cell mass m, we have

$$\frac{A}{m\,N_A} = A_{\mathsf{L}}\, x_{\mathsf{L}} + A_{\mathsf{I}}\,(x_{\mathsf{IG}} + x_{\mathsf{IN}}),$$

where $x_X = \frac{s_X}{m}$ denotes the (mass-specific) concentration. Finally, using cell volume V, the surface-to-volume ratio $r = \frac{A}{V}$, and cell density $\rho = \frac{m}{V}$, we obtain $\frac{A}{m} = \frac{A}{V}\,\frac{V}{m} = \frac{r}{\rho}$ and hence

$$\boxed{\frac{r}{\rho\,N_A} = A_{\mathsf{L}}\, x_{\mathsf{L}} + A_{\mathsf{I}}\,(x_{\mathsf{IG}} + x_{\mathsf{IN}}).}$$

Additionally, we require that a minimum fraction α of the surface area is formed by lipids,

$$A_{\mathsf{L}} \cdot \#\mathsf{L} \geq \alpha\, A = \alpha\,(A_{\mathsf{L}} \cdot \#\mathsf{L} + A_{\mathsf{I}} \cdot (\#\mathsf{IG} + \#\mathsf{IN}))\,,$$

that is,

$$A_{\mathsf{L}}\, x_{\mathsf{L}} \geq \alpha\,(A_{\mathsf{L}}\, x_{\mathsf{L}} + A_{\mathsf{I}}\,(x_{\mathsf{IG}} + x_{\mathsf{IN}}))\,,$$

where we use concentrations instead of numbers of molecules. Equivalently,

$$\boxed{(1-\alpha)\,A_{\mathsf{L}}\, x_{\mathsf{L}} \geq \alpha\, A_{\mathsf{I}}\,(x_{\mathsf{IG}} + x_{\mathsf{IN}}).}$$

C Example: figures and tables

Table S1. Essential constraints for the small model of a self-fabricating cell.

Name	**(In)equality**
mass balance G	$v_{\mathsf{IG}} - v_{\mathsf{EAA}} - n_{\mathsf{L}} v_{\mathsf{ELD}} \geq 0$
mass balance N	$v_{\mathsf{IN}} - v_{\mathsf{EAA}} \geq 0$
mass balance AA	$v_{\mathsf{EAA}} - n_{\mathsf{I}} \left(w_{\mathsf{IG}} + w_{\mathsf{IN}}\right) - n_{\mathsf{E}} \left(w_{\mathsf{EAA}} + w_{\mathsf{ELD}} + w_{\mathsf{EL}}\right) - n_{\mathsf{R}}\, w_{\mathsf{R}} \geq 0$
mass balance LD	$v_{\mathsf{ELD}} - v_{\mathsf{E}} \geq 0$
capacity IG	$v_{\mathsf{IG}} \leq k_{\mathrm{cat}}\, x_{\mathsf{IG}}$
capacity IN	$v_{\mathsf{IN}} \leq k_{\mathrm{cat}}\, x_{\mathsf{IN}}$
capacity EAA	$v_{\mathsf{EAA}} \leq k_{\mathrm{cat}}\, x_{\mathsf{EAA}}$
capacity ELD	$v_{\mathsf{ELD}} \leq k_{\mathrm{cat}}\, x_{\mathsf{ELD}}$
capacity EL	$v_{\mathsf{EL}} \leq k_{\mathrm{cat}}\, x_{\mathsf{EL}}$
capacity R	$n_{\mathsf{I}} \left(w_{\mathsf{IG}} + w_{\mathsf{IN}}\right) + n_{\mathsf{E}} \left(w_{\mathsf{EAA}} + w_{\mathsf{ELD}} + w_{\mathsf{EL}}\right) + n_{\mathsf{R}}\, w_{\mathsf{R}} \leq k_{\mathrm{el}}\, x_{\mathsf{R}}$
membrane L	$(1-\alpha)\, A_{\mathsf{L}}\, x_{\mathsf{L}} \geq \alpha\, A_{\mathsf{I}} \left(x_{\mathsf{IG}} + x_{\mathsf{IN}}\right)$
membrane	$A_{\mathsf{L}}\, x_{\mathsf{L}} + A_{\mathsf{I}} \left(x_{\mathsf{IG}} + x_{\mathsf{IN}}\right) = \frac{r}{\rho\, N_A}$
dry weight	$\omega^T x = 1$

Mass balance constraints $(Nv)_s \geq 0$ are stated only for metabolites $s \in \{\mathsf{G}, \mathsf{N}, \mathsf{AA}, \mathsf{LD}\}$, since lipids L and macromolecules $\mathsf{Mac} = \{\mathsf{IG}, \mathsf{IN}, \mathsf{EAA}, \mathsf{ELD}, \mathsf{EL}, \mathsf{R}\}$ are products of exactly one reaction each and not educts of any reaction. The corresponding constraints $v_{\mathsf{EL}} \geq 0$ and $w_s \geq 0$ for $s \in \mathsf{Mac}$ are covered by the irreversibility of all reactions, $v \geq 0$. The last two constraints are *equality* constraints (which are always active).

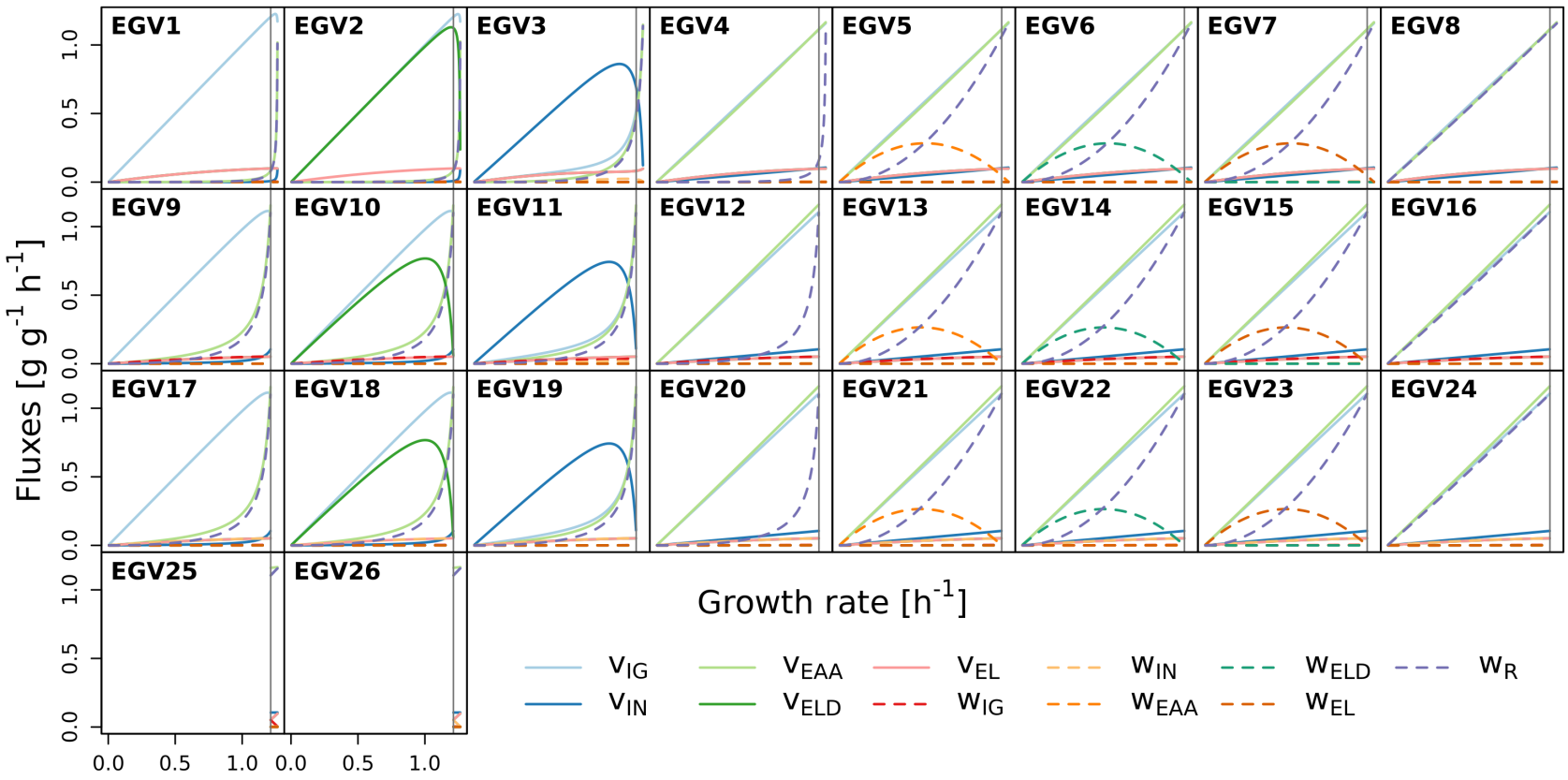


Fig S1. (Scaled) fluxes for all 26 EGVs.

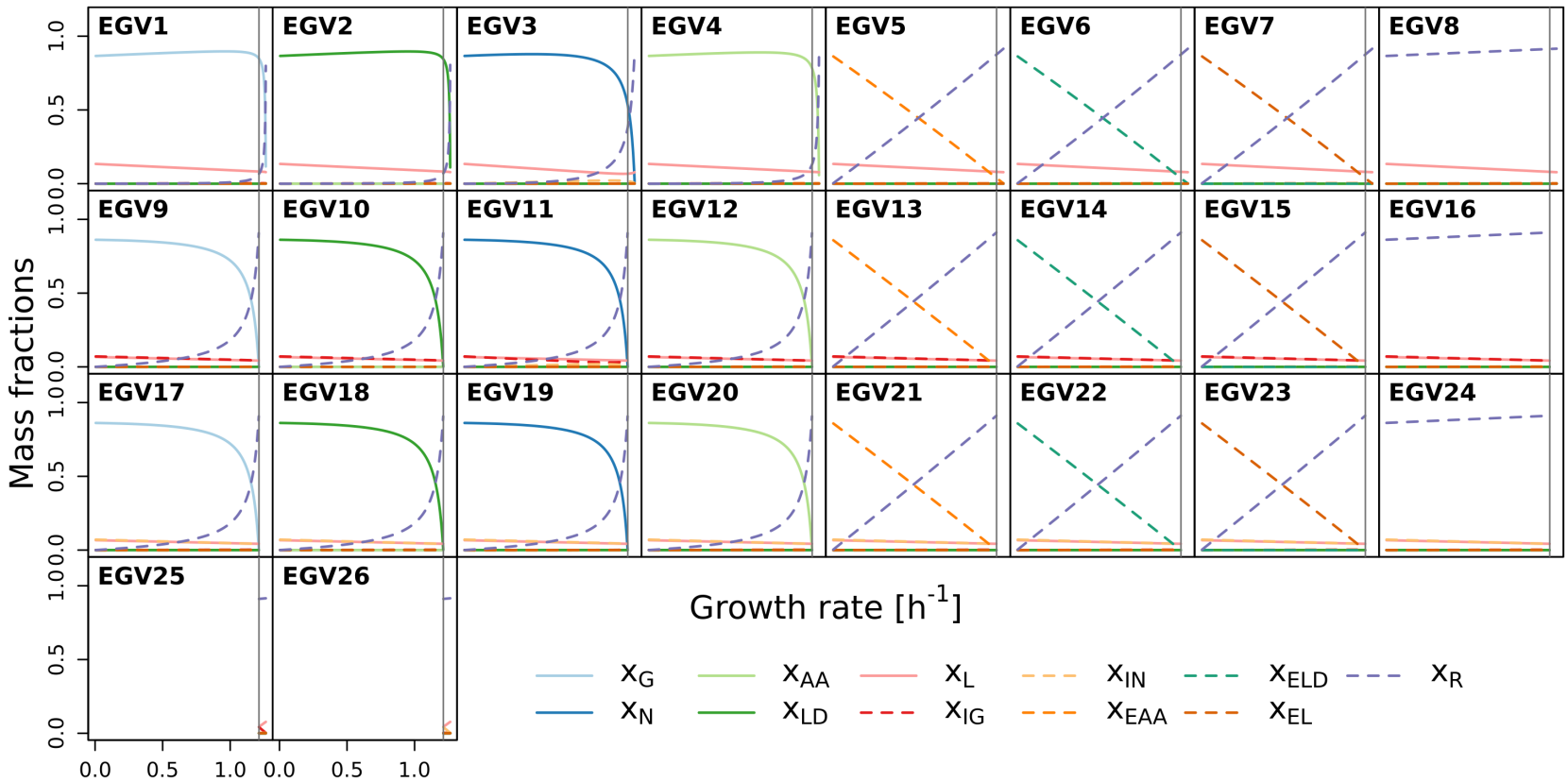


Fig S2. Mass fractions for all 26 EGVs.

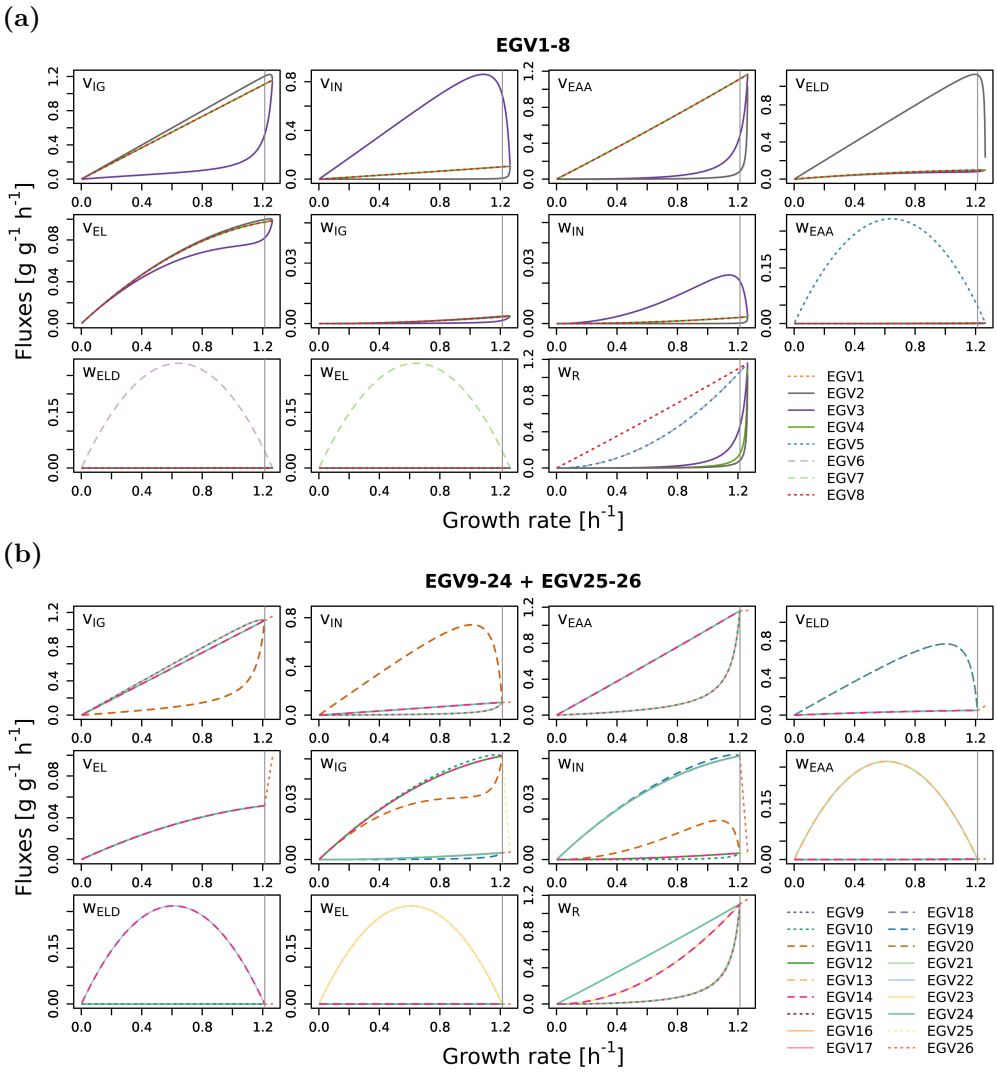


Fig S3. (Scaled) fluxes for (a) the 8 EGVs that exist for all growth rates and (b) the 16 EGVs that exist in regime L plus the 2 EGVs that exist in regime H.