

Elementary vectors and autocatalytic sets for resource allocation in next-generation models of cellular growth

Stefan Müller^{1*}, Diana Szélioová², 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)

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	
$N \in \mathbb{R}^{\text{Mol} \times \text{Rxn}}$	stoichiometric matrix	(unit: 1)
$\omega \in \mathbb{R}_{>}^{\text{Mol}}$	molar masses	(unit: g mol ⁻¹)
$X \in \mathbb{R}_{\geq}^{\text{Mol}}$	amounts of substance	(unit: mol)
$R(\cdot) \in \mathbb{R}^{\text{Rxn}}$	reaction rates (extensive)	(unit: mol h ⁻¹)

The chemical reactions induce the dynamical system

$$\frac{dX}{dt} = NR. \quad (1)$$

We define mass,

$$M = \sum_i \omega_i X_i = \omega^T X, \quad (\text{unit: g}) \quad (2a)$$

the intensive quantities

$$x = \frac{X}{M} \in \mathbb{R}_{\geq}^{\text{Mol}}, \quad (\text{unit: mol g}^{-1}) \quad (2b)$$

$$v = \frac{R}{M} \in \mathbb{R}^{\text{Rxn}}, \quad (\text{unit: mol g}^{-1} \text{ h}^{-1}) \quad (2c)$$

and growth rate

$$\mu = \frac{1}{M} \frac{dM}{dt}. \quad (\text{unit: h}^{-1}) \quad (2d)$$

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 mol g⁻¹ (dry weight).

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

$$\frac{d}{dt} \frac{X}{M} = \frac{1}{M} \frac{dX}{dt} - \frac{X}{M^2} \frac{dM}{dt}.$$

Equations (1), (2), and the chain rule yield the dynamic model of cellular growth:

$$\frac{dx}{dt} = Nv(x) - \mu x$$

and

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

Thereby, we assume given 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: 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. [2] or [1].

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

$$\frac{d(c^T x)}{dt} = c^T Nv(x) - \mu (c^T x).$$

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, $\ker N^T \cap \mathbb{R}_{\geq}^{\text{Mol}} = \{0\}$.

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, $\frac{d(c^T x)}{dt} = \frac{d(x_1 + x_2)}{dt} = -\mu(x_1 + x_2) \leq 0$, 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, $\frac{d(c^T x)}{dt} = \frac{d(x_1 - x_2)}{dt} = -\mu(x_1 - x_2)$, 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,

$$A = A_L \cdot \#L + A_I \cdot (\#IG + \#IN),$$

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

$$\frac{A}{N_A} = A_L s_L + A_I (s_{IG} + s_{IN}),$$

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

$$\frac{A}{m N_A} = A_L x_L + A_I (x_{IG} + x_{IN}),$$

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

$$\frac{r}{\rho N_A} = A_L x_L + A_I (x_{IG} + x_{IN}).$$

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

$$A_L \cdot \#L \geq \alpha A = \alpha (A_L \cdot \#L + A_I \cdot (\#IG + \#IN)),$$

that is,

$$A_L x_L \geq \alpha (A_L x_L + A_I (x_{IG} + x_{IN})),$$

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

$$(1 - \alpha) A_L x_L \geq \alpha A_I (x_{IG} + x_{IN}).$$

C Example: figures and tables

Name	(In)equality
mass balance G	$v_{IG} - v_{EAA} - n_L v_{ELD} \geq 0$
mass balance N	$v_{IN} - v_{EAA} \geq 0$
mass balance AA	$v_{EAA} - n_I (w_{IG} + w_{IN}) - n_E (w_{EAA} + w_{ELD} + w_{EL}) - n_R w_R \geq 0$
mass balance LD	$v_{ELD} - v_{EL} \geq 0$
capacity IG	$v_{IG} \leq k_{cat} x_{IG}$
capacity IN	$v_{IN} \leq k_{cat} x_{IN}$
capacity EAA	$v_{EAA} \leq k_{cat} x_{EAA}$
capacity ELD	$v_{ELD} \leq k_{cat} x_{ELD}$
capacity EL	$v_{EL} \leq k_{cat} x_{EL}$
capacity R	$n_I (w_{IG} + w_{IN}) + n_E (w_{EAA} + w_{ELD} + w_{EL}) + n_R w_R \leq k_{el} x_R$
membrane L	$(1 - \alpha) A_L x_L \geq \alpha A_I (x_{IG} + x_{IN})$
membrane	$A_L x_L + A_I (x_{IG} + x_{IN}) = \frac{r}{\rho N_A}$
(dry) mass	$\omega^T x = 1$

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

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

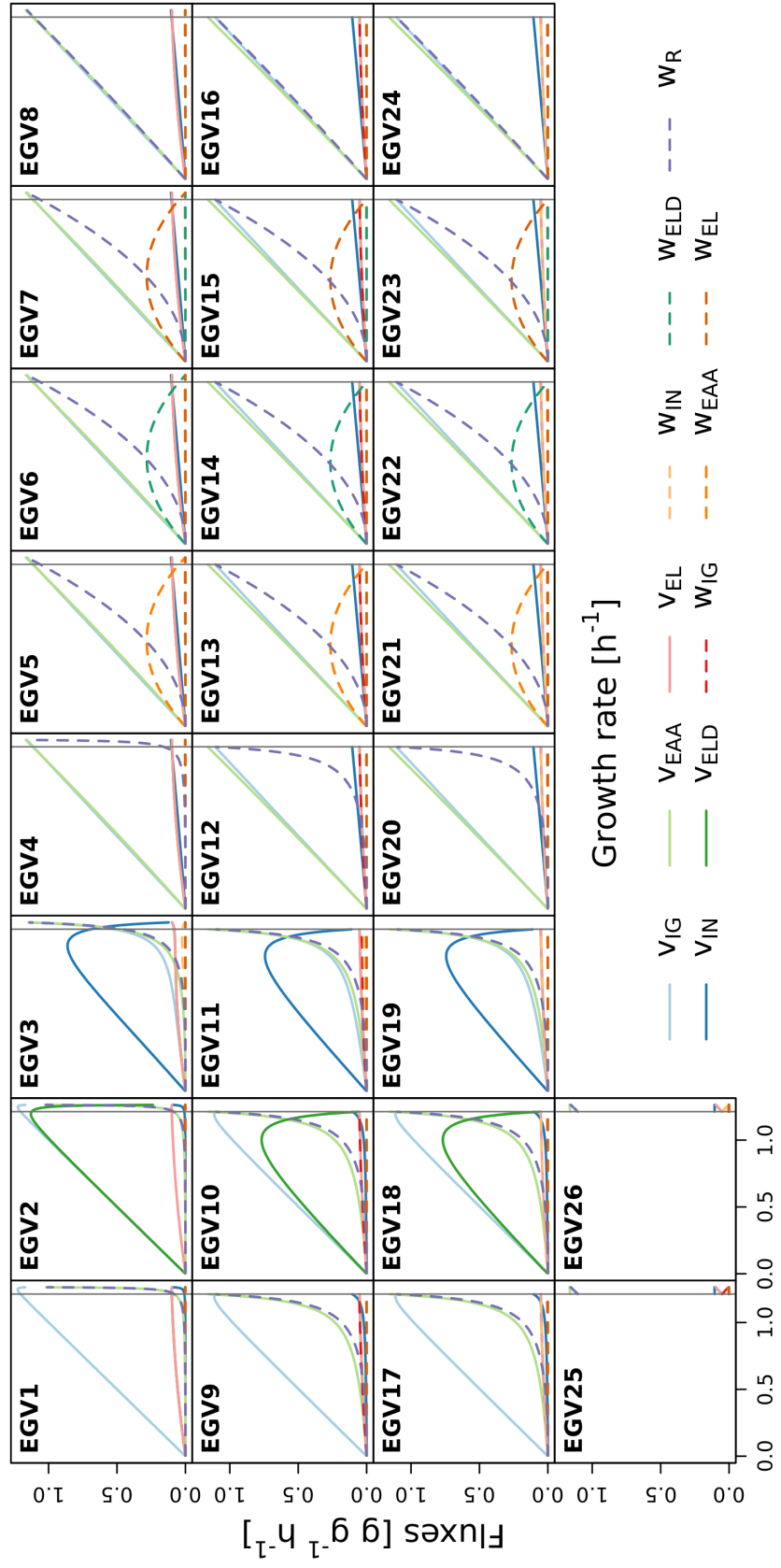


Fig A. (Scaled) fluxes for all 26 EGVs.

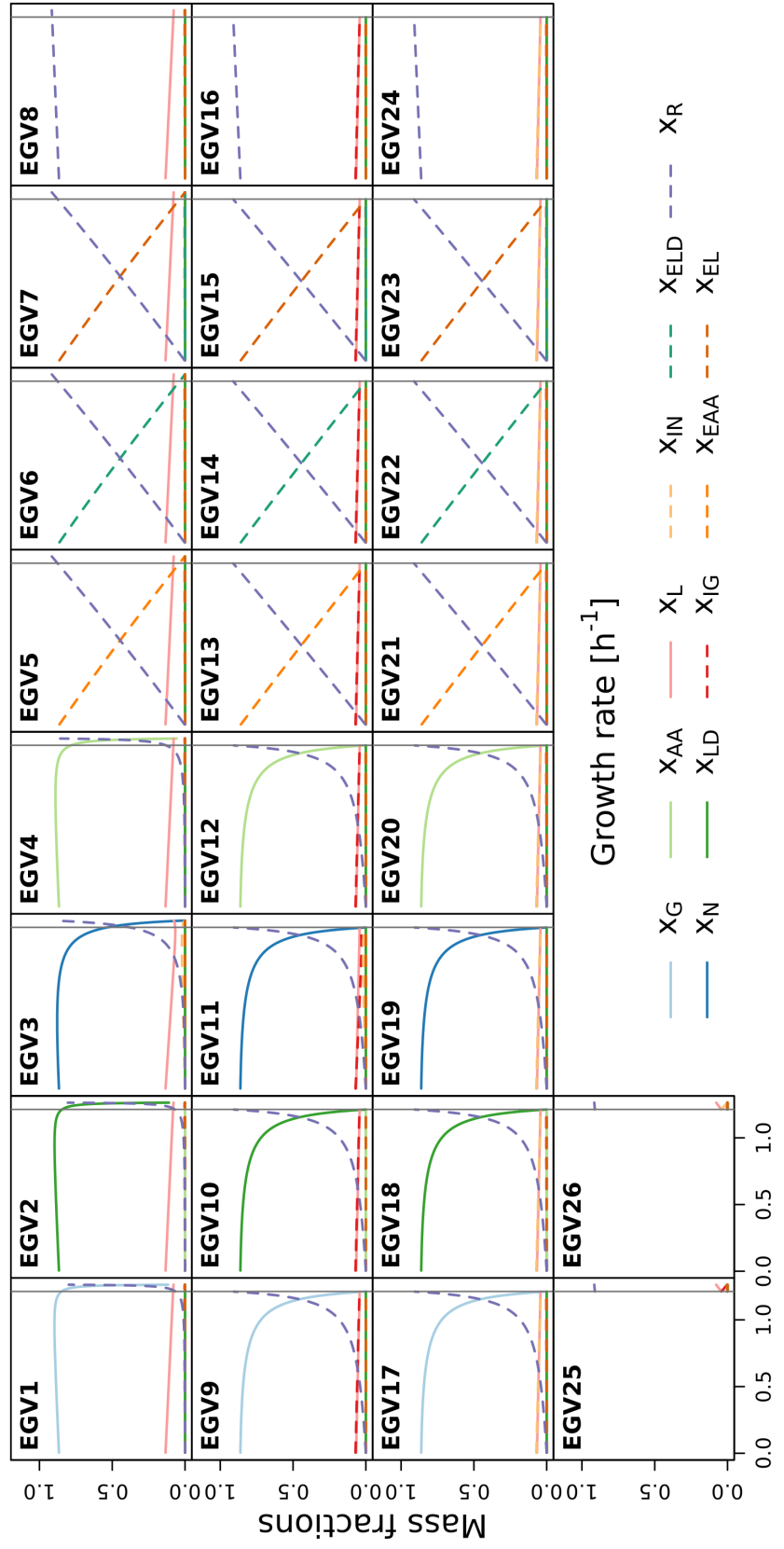


Fig B. Mass fractions for all 26 EGVs.

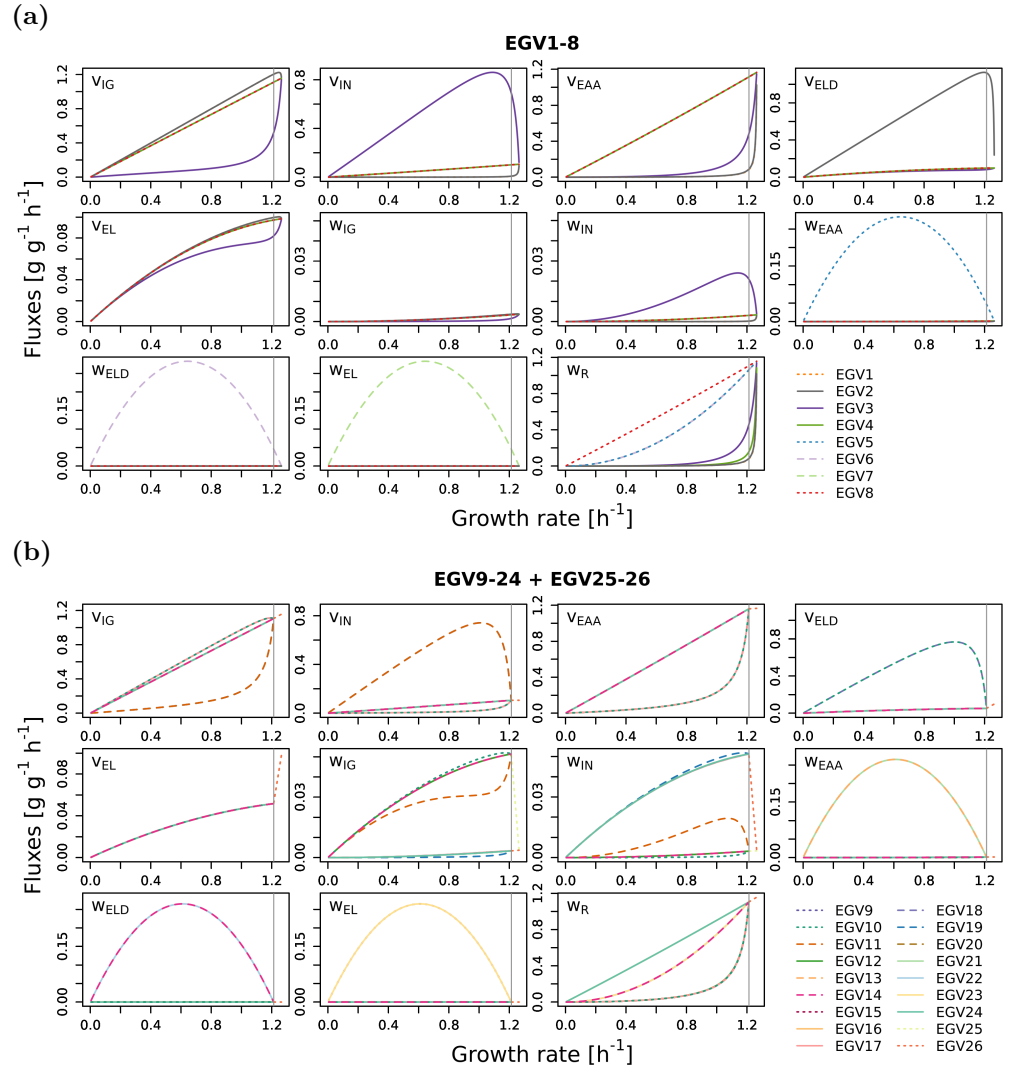
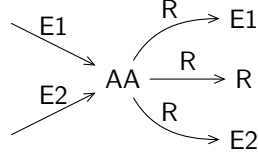


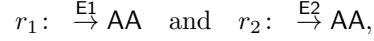
Fig C. (Scaled) fluxes for (a) the 8 EGVs that exist for all growth rates and (b) the 16 EGVs that exist in regime L plus the 2 EGVs that exist in regime H.

D Minimal growth model with alternative pathways

Consider the following minimal model of cellular growth with two alternative pathways:



The cell takes up external substrates and forms amino acids (AA) via two “reactions”,



catalyzed by the “enzymes” E1 and E2, respectively. Amino acids are then used by the ribosome (R) to synthesize the enzymes and the ribosome itself,



The set of molecular species is $\text{Mol} = \{AA, E1, E2, R\}$, and the set of reactions is $\text{Rxn} = \text{Rmet} \cup \text{Rsyn}$ with metabolic reactions $\text{Rmet} = \{r_1, r_2\}$ and synthesis reactions $\text{Rsyn} = \{s_1, s_2, s_R\}$.

The resulting stoichiometric matrix and the corresponding flux vector are given by

$$\begin{array}{c}
 \text{AA} \\
 \text{E1} \\
 \text{E2} \\
 \text{R}
 \end{array}
 \begin{pmatrix}
 r_1 & r_2 & s_1 & s_2 & s_R \\
 1 & 1 & -1 & -1 & -1 \\
 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1
 \end{pmatrix}
 = N$$

and

$$(v_1, v_2; w_1, w_2, w_R)^T = v.$$

By mass conservation (for the synthesis reactions), the molar masses obey $\omega = \bar{\omega} \cdot (1, 1, 1, 1)^T$. The growth cone is given by

$$C_g = \{(v_1, v_2; w_1, w_2, w_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid v_1 + v_2 - w_1 - w_2 - w_R \geq 0\}.$$

As it turns out, there are $4 \times 2 = 8$ EGMs (up to scaling), corresponding to the four species and the two alternative pathways, that is, $e^{s,i} \in \mathbb{R}_{\geq}^{\text{Rxn}}$ with $s \in \text{Mol}$ and $i \in \{1, 2\}$. Explicitly,

$$\begin{aligned}
 e^{\text{AA},1} &= \mu/\bar{\omega} \cdot (1, 0; 0, 0, 0)^T, \\
 e^{\text{AA},2} &= \mu/\bar{\omega} \cdot (0, 1; 0, 0, 0)^T, \\
 e^{\text{E1},1} &= \mu/\bar{\omega} \cdot (1, 0; 1, 0, 0)^T, \\
 e^{\text{E1},2} &= \mu/\bar{\omega} \cdot (0, 1; 1, 0, 0)^T, \\
 e^{\text{E2},1} &= \mu/\bar{\omega} \cdot (1, 0; 0, 1, 0)^T, \\
 e^{\text{E2},2} &= \mu/\bar{\omega} \cdot (0, 1; 0, 1, 0)^T, \\
 e^{\text{R},1} &= \mu/\bar{\omega} \cdot (1, 0; 0, 0, 1)^T, \\
 e^{\text{R},2} &= \mu/\bar{\omega} \cdot (0, 1; 0, 0, 1)^T.
 \end{aligned}$$

Every EGM “produces” exactly one molecular species, as indicated by its name, thereby using either pathway 1 or 2. (For every EGM, there is exactly one species with nonzero associated concentration.) Due to the factor $\mu/\bar{\omega}$, all EGMs have associated growth rate μ .

GMs need not be AC, in particular, no EGM is AC. In fact, every (nonzero) GM is BC, since all reactions are catalytic, however, a GM need not be CC. MAC subsets of reactions are the supports of AC GMs. There are two MAC sets, namely $M_1 = \{r_1, s_1, s_R\}$ and $M_2 = \{r_2, s_2, s_R\}$, corresponding to the two alternative pathways. AC GMs with support M_1 are generated by the EGMs $e^{AA,1}, e^{E1,1}, e^{R,1}$. (Analogously for the MAC set M_2 .)

For a GM $v \in C_g$, the associated growth rate amounts to

$$\mu = \omega^T N v = \bar{\omega} (v_1 + v_2),$$

determined by the “exchange” fluxes v_1 and v_2 . For fixed growth rate μ , the growth cone becomes a growth polytope,

$$P_g(\mu) = \{(v_1, v_2; w_1, w_2, w_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid w_1 + w_2 + w_R \leq v_1 + v_2, \\ v_1 + v_2 = \mu/\bar{\omega}\}.$$

Further, for scaled fluxes $\hat{v} = \bar{\omega}/\mu \cdot v$, the polytope becomes independent of μ ,

$$\hat{P}_g = \{(\hat{v}_1, \hat{v}_2; \hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid \hat{w}_1 + \hat{w}_2 + \hat{w}_R \leq 1, \\ \hat{v}_1 + \hat{v}_2 = 1\}.$$

In particular, its projection to the synthesis fluxes $\hat{w} \in \mathbb{R}_{\geq}^{\text{Rsyn}}$ is the “growth simplex”

$$\mathcal{P}_g = \{(\hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rsyn}} \mid \hat{w}_1 + \hat{w}_2 + \hat{w}_R \leq 1\},$$

spanned by the projections of the scaled EGMs,

$$\begin{aligned} e^{AA} &= (0, 0, 0)^T, \\ e^{E1} &= (1, 0, 0)^T, \\ e^{E2} &= (0, 1, 0)^T, \\ e^R &= (0, 0, 1)^T. \end{aligned}$$

The “projections” of the MAC sets M_1 and M_2 to the synthesis reactions are $m_1 = \{s_1, s_R\}$ and $m_2 = \{s_2, s_R\}$. Scaled projections of AC GMs with support m_1 lie in the (two-dimensional) simplex generated by e^{AA}, e^{E1}, e^R . (Analogously for m_2 .)

Catalytic closure can be ensured by additional constraints.

- In a constraint-based model, one considers *inequality* enzyme capacity constraints,

$$v_i \leq k_i^{\text{cat}} x_{Ei}$$

for $i \in \{1, 2\}$, whereas

- in a (semi-)kinetic model, one considers *equality* constraints arising from enzyme and ribosome kinetics,

$$v_i = \kappa_i(x_{AA}) x_{Ei} \quad \text{and} \quad w_j = \alpha_j \tau_j(x_{AA}) x_R$$

for $i \in \{1, 2\}$ and $j \in \{1, 2, R\}$. Thereby, κ_i, τ_j are functions of the amino acid concentration x_{AA} , and α_j are control parameters (ribosome fractions) for studying growth rate maximization, cf. [1, 3].

Moreover, one often considers ribosome capacity constraints: $w_1 + w_2 + w_R \leq k_{t1} x_R$ in constraint-based models and $\sum_{j \in \{1,2,R\}} \alpha_j \leq 1$ in (semi-)kinetic models. However, the (inequality) ribosome capacity constraint is treated separately in the (semi-)kinetic model, and for reasons of comparison, we just require $x_R > 0$ in the constraint-based model.

Additional constraints involve concentrations. For a GM $v \in C_g$, the associated concentrations x are given by

$$Nv = (*, w_1, w_2, w_R)^T = \mu x = \mu (x_{AA}, x_{E1}, x_{E2}, x_R)^T.$$

In particular, $x_{E1} = w_1/\mu$ and $x_{E2} = w_2/\mu$.

In the following, we consider the growth polyhedra and EGVs arising from the additional constraints.

Constraint-based model. For fixed growth rate $\mu > 0$, the growth polytope $P_g(\mu)$ above is further restricted by *inequality* constraints,

$$P_{g,\geq}(\mu) = \{(v_1, v_2; w_1, w_2, w_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid w_1 + w_2 + w_R \leq v_1 + v_2, \\ v_1 + v_2 = \mu/\bar{\omega}, \\ v_1 \leq k_1^{\text{cat}} w_1/\mu, v_2 \leq k_2^{\text{cat}} w_2/\mu \}.$$

For scaled fluxes $\hat{v} = \bar{\omega}/\mu \cdot v$, the polytope becomes

$$\hat{P}_{g,\geq}(\mu) = \{(\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid \hat{w}_1 + \hat{w}_2 + \hat{w}_R \leq 1, \\ \hat{v}_1 + \hat{v}_2 = 1, \\ \mu \hat{v}_1 \leq k_1^{\text{cat}} \hat{w}_1, \mu \hat{v}_2 \leq k_2^{\text{cat}} \hat{w}_2 \}.$$

Its projection to the synthesis fluxes $\hat{w} \in \mathbb{R}_{\geq}^{\text{Rsyn}}$ is given by

$$\mathcal{P}_{g,\geq}(\mu) = \{(\hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rsyn}} \mid \hat{w}_1 + \hat{w}_2 + \hat{w}_R \leq 1, \\ \mu \leq k_1^{\text{cat}} \hat{w}_1 + k_2^{\text{cat}} \hat{w}_2 \}.$$

(Scaled) EGVs are the ccND vectors of a (scaled) growth polytope. (Here, since $\hat{P}_{g,\geq}(\mu)$ is contained in the nonnegative orthant, EGVs are the vertices.) The number of EGVs depends on μ . For small μ , there are 8 EGVs. As it turns out, there are at most two EGVs with nonzero ribosome concentration/flux, namely

$$(1, 0; \mu/k_1^{\text{cat}}, 0, 1 - \mu/k_1^{\text{cat}})^T, \\ (0, 1; 0, \mu/k_2^{\text{cat}}, 1 - \mu/k_2^{\text{cat}})^T.$$

Only these EGVs can fulfill the additional ribosome capacity constraint. They are AC, and their supports are the MAC sets M_1 and M_2 . Their projections are

$$e^1 = (\mu/k_1^{\text{cat}}, 0, 1 - \mu/k_1^{\text{cat}})^T, \\ e^2 = (0, \mu/k_2^{\text{cat}}, 1 - \mu/k_2^{\text{cat}})^T,$$

which have supports m_1 and m_2 , respectively.

(Semi-)kinetic model. For fixed growth rate $\mu > 0$ and fixed amino acid concentration $x_{AA} =: \bar{x}$ (and hence fixed functions $\kappa_i(\bar{x}) =: \bar{\kappa}_i$), the growth polytope $P_g(\mu)$ is further restricted by *equality* constraints,

$$P_{g,=}(\mu, \bar{x}) = \left\{ (v_1, v_2; w_1, w_2, w_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid \begin{aligned} v_1 + v_2 - w_1 - w_2 - w_R &= \mu \bar{x}, \\ v_1 + v_2 &= \mu/\bar{\omega}, \\ v_1 &= \bar{\kappa}_1 w_1/\mu, v_2 = \bar{\kappa}_2 w_2/\mu \end{aligned} \right\}.$$

For scaled fluxes $\hat{v} = \bar{\omega}/\mu \cdot v$, the polytope becomes

$$\hat{P}_{g,=}(\mu, \bar{x}) = \left\{ (\hat{v}_1, \hat{v}_2, \hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rxn}} \mid \begin{aligned} \hat{v}_1 + \hat{v}_2 - \hat{w}_1 - \hat{w}_2 - \hat{w}_R &= \bar{\omega} \bar{x}, \\ \hat{v}_1 + \hat{v}_2 &= 1, \\ \hat{v}_1 &= \bar{\kappa}_1 \hat{w}_1/\mu, \hat{v}_2 = \bar{\kappa}_2 \hat{w}_2/\mu \end{aligned} \right\}.$$

Since \hat{v}_1, \hat{v}_2 depend on \hat{w}_1, \hat{w}_2 , the polytope is in one-to-one correspondence with its projection to the synthesis fluxes $\hat{w} \in \mathbb{R}_{\geq}^{\text{Rsyn}}$,

$$\mathcal{P}_{g,=}(\mu, \bar{x}) = \left\{ (\hat{w}_1, \hat{w}_2, \hat{w}_R)^T \in \mathbb{R}_{\geq}^{\text{Rsyn}} \mid \begin{aligned} 1 - \hat{w}_1 - \hat{w}_2 - \hat{w}_R &= \bar{\omega} \bar{x}, \\ \bar{\kappa}_1 \hat{w}_1 + \bar{\kappa}_2 \hat{w}_2 &= \mu \end{aligned} \right\}.$$

Again, scaled EGVs are vertices of the scaled polytope $\hat{P}_{g,=}(\mu, \bar{x})$. As it turns out, there are at most two EGVs, in particular, at most two EGVs with nonzero ribosome concentration/flux, namely

$$\begin{aligned} (1, 0; \mu/\bar{\kappa}_1, 0, 1 - \bar{\omega} \bar{x} - \mu/\bar{\kappa}_1)^T, \\ (0, 1; 0, \mu/\bar{\kappa}_2, 1 - \bar{\omega} \bar{x} - \mu/\bar{\kappa}_2)^T. \end{aligned}$$

These EGVs are AC, and their supports are the MAC sets M_1 and M_2 . They are in one-to-one correspondence with their projections,

$$\begin{aligned} \varepsilon^1 &= (\mu/\bar{\kappa}_1, 0, 1 - \bar{\omega} \bar{x} - \mu/\bar{\kappa}_1)^T, \\ \varepsilon^2 &= (0, \mu/\bar{\kappa}_2, 1 - \bar{\omega} \bar{x} - \mu/\bar{\kappa}_2)^T, \end{aligned}$$

the vertices of the polytope $\mathcal{P}_{g,=}(\mu, \bar{x})$, which are called *elementary growth states* (EGSs) in [1]. (In fact, EGSs are defined via the control parameters α , which are, however, in one-to-one correspondence with the synthesis fluxes w .)

The supports of the EGSs ε_1 and ε_2 , that is, m_1 and m_2 , the projections of the MAC sets, are called *elementary growth modes* in [1].

References

1. 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.
2. de Jong H, Casagrandi S, Giordano N, Cinquemani E, Ropers D, Geiselman J, et al. Mathematical modelling of microbes: metabolism, gene expression and growth. *Journal of The Royal Society Interface.* 2017;14(136):20170502.
3. Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol.* 2009;5(1):323.