

1 **Overview**

2 **Balance equations**

Introduction

The fundamental unit of biochemical engineering is the cell. Cells take in extracellular nutrients such as sugars and amino acids and use these for cell growth and the synthesis of useful products. These substrates are converted to biomass, products, and waste through a series of chemical reactions known as cellular metabolism.

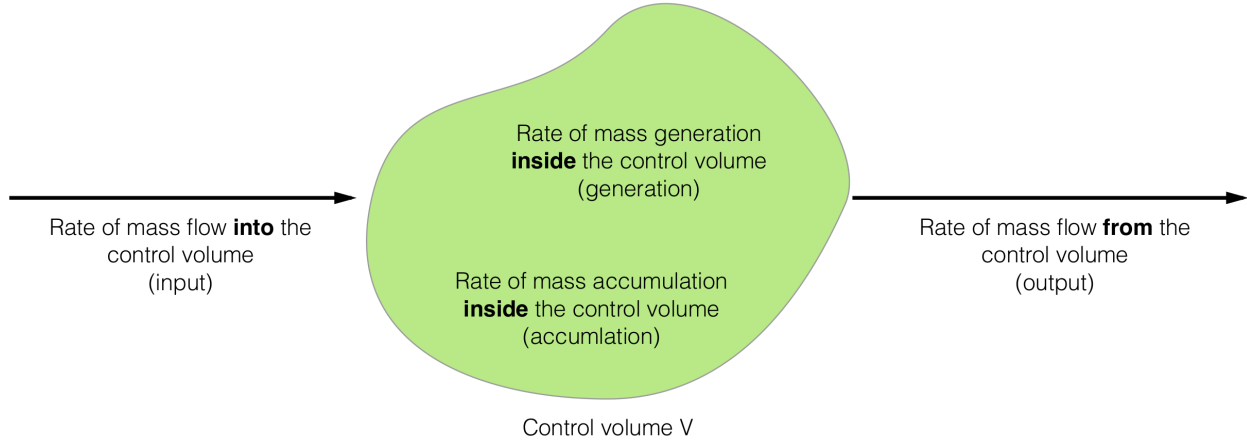


Fig. 1: Schematic of an idealized control volume V. Total macroscopic material balance equations consist of four types of terms. Material flow into and from the control volume, material generation inside the control volume and material accumulation inside the control volume. The abundance of material can be measured using both a mass or mole basis.

Given an idealized control volume, the concentration of any species (metabolite) within will change according to four processes: flow in, flow out, generation, and consumption. We can define a mass flow rate \dot{m}_s for each stream into or out of the control volume, and a direction parameter v_s such that $v_s = 1$ for streams entering the control volume and $v_s = -1$ for streams exiting the control volume. Furthermore, we can define generation minus consumption as net generation \dot{m}_{gen} , and the overall resulting change in concentration as accumulation \dot{m}_{acc} . This gives us:

$$\dot{m}_{acc} = \sum_{s=1}^S v_s \dot{m}_s + \dot{m}_{gen} \quad (1)$$

However, we can also write a mole balance in the same way:

$$\dot{n}_{acc} = \sum_{s=1}^S v_s \dot{n}_s + \dot{n}_{gen} \quad (2)$$

16 Conservation of mass dictates that net mass generation is zero $\dot{m}_{gen} = 0$, but no such law
 17 exists for a mole balance $\dot{n}_{gen} \neq 0$. We can write a mole balance around an individual
 18 species as well:

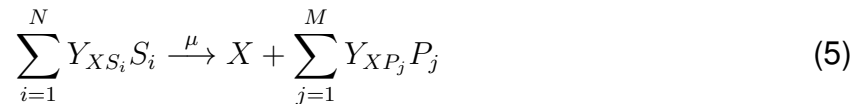
$$\dot{n}_{j,acc} = \sum_{s=1}^S v_s \dot{n}_s x_{j,s} + \dot{n}_{j,gen} \quad j = 1, 2, \dots, \mathcal{M} \quad (3)$$

19 where $x_{j,s}$ is the mole fraction of species j in stream s . We can also construct a balance
 20 using volumetric flow rate F_s as a basis for each stream, in which case mole fraction will
 21 be replaced by species concentration $C_{j,s}$:

$$\frac{d}{dt} (C_j V) = \sum_{s=1}^S v_s F_s C_{j,s} + \dot{n}_{j,gen} \quad j = 1, 2, \dots, \mathcal{M} \quad (4)$$

22 Here the mole rate of accumulation of species j is equal to the rate of change of the
 23 species concentration times the volume.

24 To better understand how moles of a species can be generated in a cell, we must
 25 consider the stoichiometry of cellular metabolism. Generally, cells consume substrates
 26 S_1, S_2, \dots, S_N to produce products P_1, P_2, \dots, P_M , as well as more cells X :



27 Y_{Xj} are yield coefficients, which relate the consumption or production of a species to μ ,
 28 the specific growth rate of cells. If a species is a substrate, Y_{Xj} represents how much of
 29 that substrate is consumed in the production of one unit of cells. $Y_{Xj} < 0$ for all substrates,
 30 and $Y_{Xj} > 0$ for all products. Thus, $Y_{XS_i} < 0$ and $Y_{XP_j} > 0$ in the equation above. This is
 31 the overall stoichiometry from substrates to products, a lumped version of all the chemical

32 reactions that occur in the cell.

$$\dot{n}_{j,gen} = \left(\sum_{r=1}^{\mathcal{R}} \sigma_{j,r} \hat{r}_r \right) V + \left(\sum_{k=1}^{\mathcal{T}} \tau_{j,k} q_k \right) XV \quad (6)$$

$$\frac{d}{dt} (C_j V) = \sum_{s=1}^{\mathcal{S}} v_s F_s C_{j,s} + \left(\sum_{r=1}^{\mathcal{R}} \sigma_{j,r} \hat{r}_r \right) V + \left(\sum_{k=1}^{\mathcal{T}} \tau_{j,k} q_k \right) XV \quad j = 1, 2, \dots, \mathcal{M} \quad (7)$$

Extracellular balances

$$\frac{dC_j}{dt} = \sum_{s=1}^S v_s D_s C_{j,s} + \left(\sum_{r=1}^{\mathcal{R}} \sigma_{j,r} \hat{r}_r \right) + \left(\sum_{k=1}^{\mathcal{T}} \tau_{j,k} q_k \right) X - \frac{C_j}{V} \frac{dV}{dt} \quad j = 1, 2, \dots, \mathcal{M} \quad (8)$$

$$\frac{dX}{dt} = \sum_{s=1}^S v_s D_s X_s + (\mu - k_d) X - \frac{X}{V} \frac{dV}{dt} \quad (9)$$

$$\frac{dV}{dt} = \sum_{s=1}^S v_s \frac{\rho_s}{\rho} F_s - \frac{V}{\rho} \frac{d\rho}{dt} \quad (10)$$

$$D_s \equiv \frac{F_s}{V} \quad (11)$$

34 **Batch cultures**

35 A batch process is defined by the absence of flow in or out of the reactor:

$$F_{in} = F_{out} = 0 \quad \therefore \quad D_{in} = D_{out} = 0 \quad (12)$$

36 This results in a simplification of our extracellular balances:

$$\frac{dS}{dt} = - \left(\frac{\mu}{Y_{XS}} + m_S \right) X \quad (13)$$

$$\frac{dX}{dt} = (\mu - k_d) X \quad (14)$$

$$\frac{dV}{dt} = 0 \quad (15)$$

Solving the system of ODEs

We used ordinary differential equations (ODEs) to model the time evolution of metabolite (x_i) and scaled enzyme abundance (ϵ_i) in hypothetical cell-free metabolic networks:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \quad i = 1, 2, \dots, \mathcal{M} \quad (16)$$

$$\frac{d\epsilon_i}{dt} = -\lambda_i \epsilon_i \quad i = 1, 2, \dots, \mathcal{E} \quad (17)$$

where \mathcal{R} denotes the number of reactions, \mathcal{M} denotes the number of metabolites and \mathcal{E} denotes the number of enzymes in the model. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j . Typically, reaction j is a non-linear function of metabolite and enzyme abundance, as well as unknown kinetic parameters \mathbf{k} ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j . If $\sigma_{ij} > 0$, metabolite i is produced by reaction j . Conversely, if $\sigma_{ij} < 0$, metabolite i is consumed by reaction j , while $\sigma_{ij} = 0$ indicates metabolite i is not connected with reaction j . Lastly, λ_i denotes the scaled enzyme activity decay constant. The system material balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$ and $\epsilon(t_o) = 1$ (initially we have 100% cell-free enzyme abundance).

Heuristic optimization

After solving the equations, we can calculate a variety of metrics to evaluate the model solution. One of these is the cost function, equal to the sum-squared-error between experimental data and model predictions:

$$\text{cost} = \sum_{i=1}^{\mathcal{D}} \left[\frac{w_i}{\mathcal{Y}_i^2} \sum_{j=1}^{\mathcal{T}_i} \left(y_{ij} - x_i|_{t(j)} \right)^2 \right] \quad (18)$$

Here, x_i are the simulated species values (the solution to the model ODEs), and y_{ij} are the experimental data against which we will compare these simulations. i is an index for the number of species for which data exist (usually less than the total number of species). j refers to the time index for the experimental data, not the simulations. Thus, in order to actually subtract our simulations from the data, we must interpolate them to fit the experimental time index: $x_i|_{t(j)}$. The number of experimental timepoints (\mathcal{T}_i) can depend on which species is being considered, hence the subscript i . We then calculate a sum-squared-error for each species, and sum across all species in the dataset ($i = 1$ to \mathcal{D}). However, each species' sum-squared-error is first normalized by the square of the maximal experimental value $\mathcal{Y}_i = \max_j (y_{ij})$. In addition, we can weight the various species' contributions to the overall cost: w_i are arbitrary, and should be used to prioritize fitting certain species over others.

The model solution, and thus the value of the cost function, depend on the model parameters (rate maxima, saturation constants, and control parameters).

Materials and Methods

Estimation of kinetic model parameters. We estimated an ensemble of diverse parameter sets using a constrained Markov Chain Monte Carlo (MCMC) random walk strategy. Starting from a single best-fit parameter set estimated by inspection and literature, We then perturbed each model parameter between an upper and lower bound that varied by parameter type:

$$k_i^{new} = \min(\max(k_i \cdot \exp(a \cdot r_i), l_i), u_i) \quad i = 1, 2, \dots, \mathcal{P} \quad (19)$$

where \mathcal{P} denotes the number of parameters ($\mathcal{P} = 815$), which includes 163 maximum reaction rates (V^{max}), 163 enzyme activity decay constants, 455 saturation constants (K_{js}), and 34 control parameters, k_i^{new} denotes the new value of the i^{th} parameter, k_i denotes the current value of the i^{th} parameter, a denotes a distribution variance, r_i denotes a random sample from the normal distribution, l_i denotes the lower bound for that parameter type, and u_i denotes the upper bound for that parameter type.

Model parameters were constrained by literature *Milo:2009aa*. The rate maximum for CAT transcription was calculated as 123 h^{-1} :

$$k_{cat}^{TX} = \left(\frac{k_{TX}}{l_{mRNA}} \right) \cdot P \quad (20)$$

where k_{TX} denotes the average mRNA elongation rate (25 nt/s, *Garamella:2016aa*), l_{mRNA} denotes the CAT mRNA length (660 nt), and P denotes the level of promoter activity (estimated at 0.9).

84 **Calculation of carbon yield.** The CAT carbon yield (Y_C^{CAT}) was calculated as the ratio
 85 of carbon produced as CAT divided by the carbon consumed as reactants:

$$Y_C^{CAT} = \frac{\Delta CAT \cdot C_{CAT}}{\sum_{i=1}^{\mathcal{R}} \Delta m_i \cdot C_i} \quad (21)$$

86 where ΔCAT denotes the abundance of CAT produced, C_{CAT} denotes carbon number of
 87 CAT, \mathcal{R} denotes the number of reactants, Δm_i denotes the amount of the i^{th} reactant
 88 consumed, and C_i denotes the carbon number of the i^{th} reactant.

89 **Calculation of energy efficiency.** Energy efficiency was calculated as the ratio of CAT
 90 production to substrate consumption, both in terms of equivalent ATP molecules:

$$\text{Efficiency} = \frac{\Delta CAT \cdot (2 \cdot (\text{ATP}_{\text{TX}} + \text{CTP}_{\text{TX}} + \text{GTP}_{\text{TX}} + \text{UTP}_{\text{TX}}) + 2 \cdot \text{ATP}_{\text{TL}} + \text{GTP}_{\text{TL}})}{\sum_{i=1}^{\mathcal{R}} \Delta m_i \cdot \text{ATP}_i} \quad (22)$$

91 where ATP_{TX} , CTP_{TX} , GTP_{TX} , UTP_{TX} denote the stoichiometric coefficients of each en-
 92 ergy species for CAT transcription, ATP_{TL} , GTP_{TL} denote the stoichiometric coefficients
 93 of ATP and GTP for CAT translation, Δm_i denotes the amount of the i^{th} substrate con-
 94 sumed, and ATP_i denotes the equivalent ATP number of the i^{th} substrate. $\text{ATP}_{\text{TX}} = 176$,
 95 $\text{CTP}_{\text{TX}} = 144$, $\text{GTP}_{\text{TX}} = 151$, $\text{UTP}_{\text{TX}} = 189$, $\text{ATP}_{\text{TL}} = 219$, $\text{GTP}_{\text{TL}} = 438$, $\text{ATP}_{\text{GLC}} = 21$,
 96 $\text{ATP}_{\text{PYR}} = 8$, $\text{ATP}_{\text{LAC}} = 9.5$, $\text{ATP}_{\text{SUCC}} = 11.5$, $\text{ATP}_{\text{MAL}} = 10.5$.