

Intracellular Metabolic Networks and Flux Balance Analysis

Jeffrey D. Varner*

School of Chemical Engineering*

Purdue University, West Lafayette IN 47907

Copyright © Jeffrey Varner 2016. All Rights Reserved.

Introduction

Cells operate vast interconnected biochemical *networks* to process starting materials such as sugars into valuable products of interest, waste products and eventually more cells. To start exploring metabolic networks, let's first write material balances around the reactants and products being consumed or produced by the enzymes (catalytic protein) in these networks.

General intracellular species balance equations. To derive the material balance governing the specific intracellular concentration of the j^{th} metabolite x_j ([mmol/gdw]) in a metabolic network, we start from the general material balance:

$$\dot{n}_{X,acc,j} = \dot{n}_{X,in,j} - \dot{n}_{X,out,j} + \dot{n}_{X,gen,j} \quad j = 1, 2, \dots, \mathcal{M} \quad (1)$$

where \mathcal{M} denotes the number of metabolites and X denotes a specific cell from our population of cells. However, we have millions of cells in a bioreactor. Thus, we can't track metabolite levels in each individual cell. Instead, we make the well mixed assumption (WMA) for the cells in the reactor. The WMA for cells implies that the intracellular level of the metabolite x_j does not vary between cells i.e., if we sample x_j in many cells, the levels of x_j would be the same to within measurement error. Second, WMA also implies that x_j does not vary with position *inside* the cells. In reality, the WMA assumption for cells is incorrect; there are many interesting phenomena (both technological and human health related) that occur because of cellular diversity. Moreover, it is well known (especially in eukaryotes) that the abundance of metabolites varies between the different cellular compartments. However, for now we'll make this assumption.

There are only two terms that we need to consider for intracellular balances, the accumulation and generation terms. This is because there is no *convective* transport into or from the cellular phase, i.e., cell walls block convective transport of x_j into or from the cell. Thus, the transport terms vanish $\dot{n}_{X,in,j} = \dot{n}_{X,out,j} = 0$ and the intracellular material balance around metabolite x_j becomes:

$$\dot{n}_{X,acc,j} = \dot{n}_{X,gen,j} \quad j = 1, 2, \dots, \mathcal{M} \quad (2)$$

The well-mixed assumption simplifies the accumulation term:

$$\dot{n}_{X,acc,j} = \frac{d}{dt} \int_X x_j dX \simeq X \frac{dx_j}{dt} + x_j \frac{dX}{dt} \quad (3)$$

while the generation term becomes:

$$\dot{n}_{X,gen,j} = \int_X \left(\sum_r^{\mathcal{R}} \sigma_{jr} v_r + \sum_k^{\mathcal{T}} \tau_{jk} q_k \right) dX \simeq \left(\sum_r^{\mathcal{R}} \sigma_{jr} v_r + \sum_k^{\mathcal{T}} \tau_{jk} q_k \right) X \quad (4)$$

We have split the reaction rates into those rates that occur completely within cells (v'_k s) and those that cross the cell boundary (q'_k s) where σ_{jr} and τ_{jk} denote the respective stoichiometric coefficients. These rates, called *metabolic fluxes*, have a special unit system called specific units. The most used example of specific units for fluxes is [mmol/gdw-hr], however other examples are possible. In general, metabolic fluxes are complex functions of the concentration of the enzymes that catalyze the reactions (intracellular protein catalysts manufactured by cells), as well as substrates (starting materials) and sometimes even the products of other reactions. Substituting the accumulation and generation terms into Eqn (2) gives:

$$X \frac{dx_j}{dt} + x_j \frac{dX}{dt} = \left(\sum_r^{\mathcal{R}} \sigma_{jr} v_r + \sum_k^{\mathcal{T}} \tau_{jk} q_k \right) X \quad j = 1, 2, \dots, \mathcal{M} \quad (5)$$

Rearranging Eqn (5) and dividing by the cellmass gives:

$$\frac{dx_j}{dt} = \left(\sum_r^{\mathcal{R}} \sigma_{jr} v_r + \sum_k^{\mathcal{T}} \tau_{jk} q_k \right) - \frac{x_j}{X} \frac{dX}{dt} \quad j = 1, 2, \dots, \mathcal{M} \quad (6)$$

The last term on the right-hand side of Eqn (6) is a dilution term. This term accounts for the loss of x_j because of the growth of the cellmass population. The dilution term couples the intracellular metabolite balances to the extracellular cellmass balance in a bioreactor, and thus to the operation of the bioreactor.

Let's assume we are growing cells in a batch process. This assumption implies the dilution term can be re-written as:

$$\frac{x_j}{X} \frac{dX}{dt} = \mu x_j \quad (7)$$

where

$$\mu = \frac{1}{X} \frac{dX}{dt} \quad (8)$$

Putting everything together, we arrive at the general intracellular material balance governing x_j in index-form:

$$\frac{dx_j}{dt} = \left(\sum_r^{\mathcal{R}} \sigma_{jr} v_r + \sum_k^{\mathcal{T}} \tau_{jk} q_k \right) - \mu x_j \quad j = 1, 2, \dots, \mathcal{M} \quad (9)$$

or in matrix-vector:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} + \mathbf{T}\mathbf{q} - \mu\mathbf{x} \quad (10)$$

where \mathbf{x} denotes the species abundance vector ($\mathcal{M} \times 1$), \mathbf{S} denotes the stoichiometric matrix ($\mathcal{M} \times \mathcal{R}$), \mathbf{T} denotes the transport (or exchange) matrix ($\mathcal{M} \times \mathcal{T}$) and the \mathbf{I} denotes the ($\mathcal{M} \times \mathcal{M}$) identity matrix.

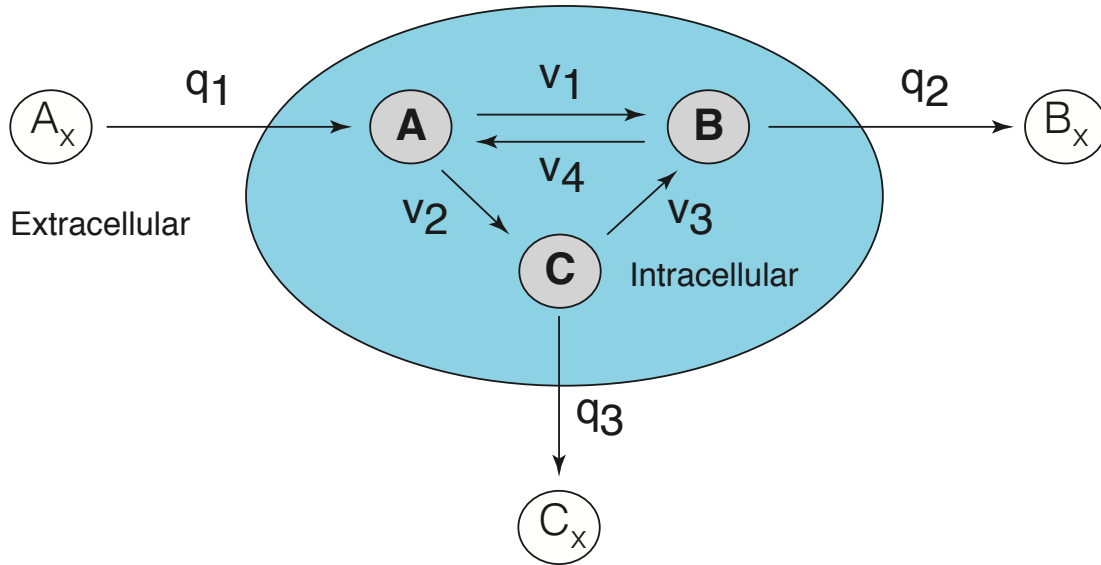


Fig. 1: Schematic of simple metabolic reaction network. This system has a total of six metabolites and seven reactions. The intracellular metabolites (metabolites A , B and C) are often treated as balanced (no accumulation) while the extracellular metabolites (A_x , B_x and C_x) are not conserved (can accumulate). The blue-oval denotes the cell boundaries, where q_j denotes flux across these boundaries ([mmol/gdw-hr]) and v_k denotes the intracellular fluxes.

Example: Formulate model system. Formulate the intracellular metabolite balances for the network shown in Fig. (1). The example network has a total of six metabolites and seven reactions. The metabolites A , B and C are intracellular (inside the cell), while A_x , B_x and C_x are extracellular (outside the cell). Using the general intracellular species mass balance formulation (and neglecting the dilution due to growth terms) yields the material balances:

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} \quad (11)$$

The material balances can be rewritten in matrix vector form:

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} + \mathbf{T}\mathbf{q} \quad (12)$$

where $\mathbf{x} \equiv \{A, B, C\}^T$, the intracellular flux vector is defined as $\mathbf{v} \equiv \{v_1, v_2, v_3, v_4\}^T$ and the specific transport (exchange) vector is defined as $\mathbf{q} \equiv \{q_1, q_2, q_3\}^T$. The stoichiometric matrix \mathbf{S} is

then given by:

$$\mathbf{S} \equiv \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

while the transport or exchange matrix \mathbf{T} is given by:

$$\mathbf{T} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Why did we neglect the μx_j terms? Most researchers in the systems biology community neglect intracellular metabolite dilution terms because they are small relative to the magnitudes of the reaction and transport fluxes ($v_k \sim q_t \gg \mu x_j$). Additionally, they frustrate the mathematical analysis of intracellular networks using approaches such as Flux Balance Analysis (FBA) as described by Palsson and coworkers (1). However, the shrinking and often marginalized kinetic modeling community still includes these terms because they modify intracellular concentration, and thus the reaction kinetics encoded in the v_k 's and q_k 's. These terms follow from the classic work of Fredrickson (2) who showed they can be important in some cases.

What can we learn from the structure of a metabolic system? We can calculate interesting properties about a metabolic system by analyzing the structure of the stoichiometric and transport matrices. Starting from the binary form of the augmented matrix (only entries in the matrices or 0 and 1):

$$\mathcal{S} \equiv [\hat{\mathbf{S}} | \hat{\mathbf{T}}] \quad (13)$$

where $\hat{\mathbf{S}}$ and $\hat{\mathbf{T}}$ denote the stoichiometric and transport arrays with all non-zero terms replaced by 1. The participation number for reaction j , defined as the number of metabolites that participate in reaction j , is given the symbol π_j and defined as:

$$\pi_j \equiv \sum_{i=1}^{\mathcal{M}} \mathcal{S}_{ij} \quad j = 1, 2, \dots, \mathcal{R} + \mathcal{T} \quad (14)$$

while the connectivity number ρ_i , defined as the number of reactions that a metabolite participates in, is defined as:

$$\rho_i \equiv \sum_{j=1}^{\mathcal{R}+\mathcal{T}} \mathcal{S}_{ij} \quad i = 1, 2, \dots, \mathcal{M} \quad (15)$$

We can also learn about the extent of cross talk in a network by analyzing the compound adjacency ($\mathcal{A}_x = \mathcal{S}\mathcal{S}^T$) and the reaction adjacency ($\mathcal{A}_r = \mathcal{S}^T\mathcal{S}$) arrays. The diagonal elements of \mathcal{A}_x give the number of reactions in which compound x_i participates (the connectivity number), while the off diagonal elements give the number of reactions in which both compounds x_i and x_j participate. One the other hand, the diagonal elements of \mathcal{A}_r give the participation number and the off diagonal elements quantify how many compounds two reactions have in common.

Example: Analyze the structure of a metabolic network. Analyze the structure of the metabolic network shown in Fig. (1). The augmented binary array for this network is given by:

$$\mathcal{S} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

The metabolite and reaction adjacency arrays are given by:

$$\mathcal{A}_x = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 3 \end{bmatrix} \quad (17)$$

and

$$\mathcal{A}_r = \begin{bmatrix} 2 & 1 & 1 & 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (18)$$

respectively. From these arrays we can estimate connectivity and crosstalk information about the metabolic network.

Flux balance analysis (FBA). Flux balance analysis (FBA) is a strategy to estimate intracellular fluxes given a model of the metabolic network, and information about the rate of material exchange between cells and the extracellular environment. Flux balance analysis was developed by Palsson and coworkers to estimate fluxes of genome scale models of *E. coli* (3). The FBA problem recasts the estimation of intracellular fluxes as a Linear Programming (LP) problem. Linear programming is a type of convex optimization problem in which a linear objective function is maximized (or minimized) subject to linear algebraic constraints. LPs can be solved easily for genome scale problems with thousands of unknown fluxes on standard hardware using packages such as MATLAB in a few seconds. The general form for a metabolic FBA problem is given by:

$$\max_{\mathbf{v}} \mathbf{c}^T \mathbf{v} \quad (19)$$

subject to the algebraic constraints:

$$\mathbf{Sv} + \mathbf{Tq} \leq \mathbf{b} \quad (20)$$

$$\mathbf{v}, \mathbf{q} \geq \mathbf{0} \quad (21)$$

where \mathbf{c}^T is the *objective selection vector* and \mathbf{b} denotes the time-derivative vector for all system states (intracellular and extracellular), while \mathbf{S} and \mathbf{T} denote the stoichiometric and transport arrays

for both intracellular and extracellular metabolites (rows) for unmeasured and measured fluxes (columns).

Example: Simple flux balance analysis problem. Estimate the intracellular fluxes for the simple reaction network shown in Fig. (1) using flux balance analysis (FBA). FBA requires that we choose an objective vector, so let's assume we want to minimize C_x production. This objective implies $\mathbf{c}^T = (0, 0, 0, 0, -1, 0, 0)$. Assuming we measured constrained $0 \leq \dot{A}_x \leq 1$ and $0 \leq \dot{B}_x \leq 0.5$, we can redefine the stoichiometric and transport arrays as:

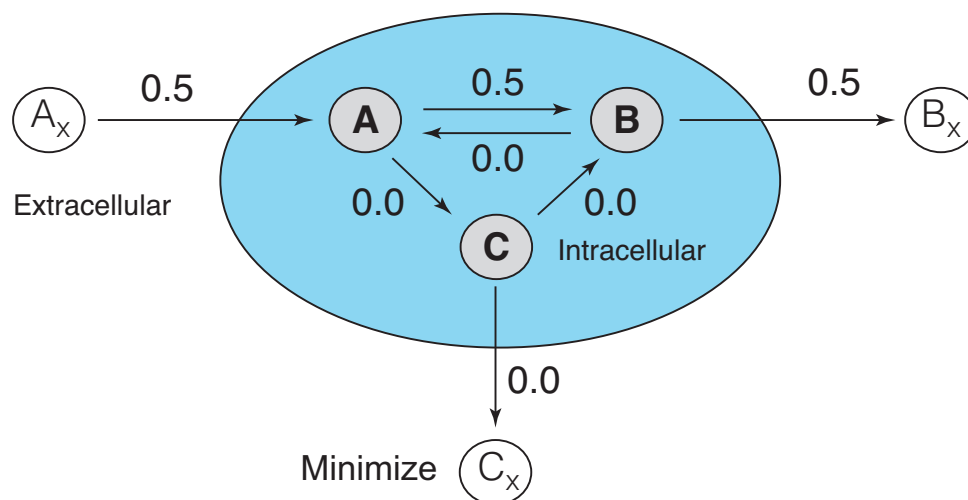
$$\mathbf{S} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (22)$$

and

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (23)$$

where the three extra rows are associated with the A_x , B_x and C_x balances. The minimum and maximum flux solutions for C_x are given in Fig. (2).

A



B

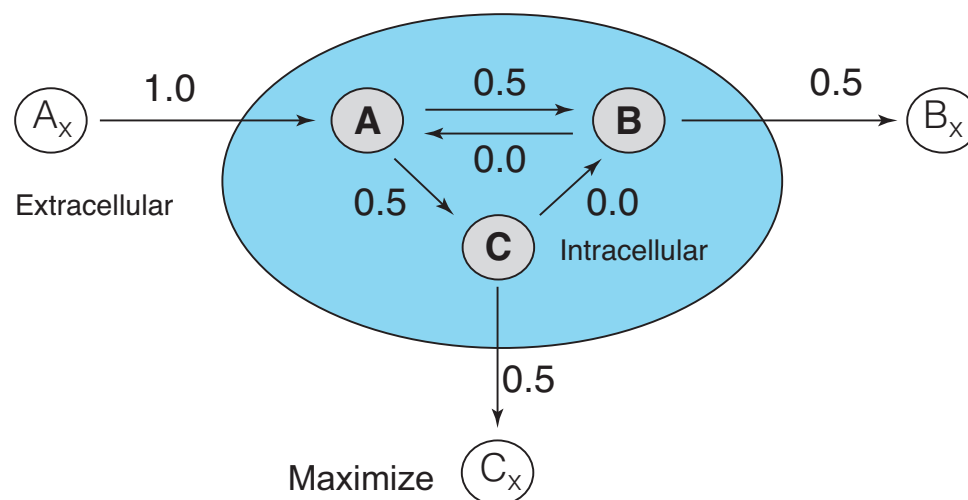


Fig. 2: Example flux solution $0 \leq \dot{A}_x \leq 1$ and $0 \leq \dot{B}_x \leq 0.5$ for minimizing or maximizing C_x .

References

1. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28: 245-8.
2. Fredrickson AG (1976) Formulation of structured growth models. *Biotechnol Bioeng* 18: 1481-6.
3. Edwards JS, Palsson BO (2000) The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* 97: 5528-33.