

Lecture 5b: Flux Balance Analysis (FBA)

*Lecturer: Jeffrey Varner***Disclaimer:** *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture, we will discuss the following topics:

- **Metabolism and metabolic networks:** A metabolic network is the complete set of metabolic (chemical) processes determining a cell's biochemical state. It encompasses all the chemical reactions associated with metabolism, i.e., the breakdown of raw materials such as sugars (catabolism) and the production of macromolecules, e.g., proteins, lipids, etc (anabolism).
- **A stoichiometric matrix** is a mathematical representation of a metabolic network that encodes the relationships between reactants and products in the network, where rows correspond to different metabolites, contained in the set \mathcal{M} , and columns correspond to reactions, contained in the set \mathcal{R} . Thus, the stoichiometric matrix is a $\mathbf{S} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{R}|}$ matrix holding the stoichiometric coefficients $\sigma_{ij} \in \mathbf{S}$ for $i = 1, 2, \dots, |\mathcal{M}|$ and $j = 1, 2, \dots, |\mathcal{R}|$.
- **Structural analysis of \mathbf{S} :** Structural analysis of the stoichiometric matrix involves examining its connectivity distribution and using tools such as eigendecomposition to explore the network's fundamental pathway structures and other topological properties. These types of analyses give us more insight into the structure of the network (and perhaps some indication of the importance of particular metabolites of reactions).

1 Introduction

In the previous lecture, we discussed the stoichiometric matrix \mathbf{S} , which is a mathematical representation of a metabolic network. The stoichiometric matrix encodes the relationships between reactants and products in the network, where rows correspond to different metabolites, contained in the set \mathcal{M} , and columns correspond to reactions, contained in the set \mathcal{R} . Thus, the stoichiometric matrix is a $\mathbf{S} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{R}|}$ matrix holding the stoichiometric coefficients $\sigma_{ij} \in \mathbf{S}$ for $i = 1, 2, \dots, |\mathcal{M}|$ and $j = 1, 2, \dots, |\mathcal{R}|$. In this lecture, we will discuss how we can use the stoichiometric matrix to analyze the metabolic network and predict the fluxes through the network using a technique called Flux Balance Analysis (FBA). In particular, we will explore the ideas in the paper by Orth et al (1).

2 Material Balance Equations

Before we dive into FBA, let's first consider the material balance equations for a metabolic network. These form the constraints that we will use in FBA. Each metabolite, enzyme, genem, mRNA species in the metabolic network is subject to a material balance equation, which describes the rate of change of the concentration of the metabolite (enzyme, etc) in the system. Material balance equations are a fundamental concept in chemical engineering that can bbe applied to many different types of systems, including metabolic networks. Material balances consist for four types of terms: accumulation, generation, transport in and out of the system.

Imagine we have an **abstract biophase**, which is a well-mixed compartment where our metabolic reactions are occurring. This biophase can be a single cell, a compartment within a single cell, such as the cytoplasm or mitochondria, a cell-free system or a even collection of cells in a bioreactor. Let the volume of the biophase be V (depending on the context, V could be the volume of a single cell, the volume of the cytoplasm, cell number, cell mass, thus it will units specific to the case). Then, we can write balances around each chemical species $i \in \mathcal{M}$ in the system as (Defn. 1):

Definition 1 (Material Balance Equation). *A material balance equation for species i in a biophase with volume V has four terms:*

$$\text{Accumulation} = \text{Generation} + \text{Transport In} - \text{Transport Out} \quad (1)$$

The **accumulation** term is the rate of change of species i in the system, **generation** is the rate of production (consumption) of species i by chemical reactions in the system, the **transport** terms describe the rate of physical transport (convection or passive diffusion) of species i into (from) the system.

Material balance equations can be written in terms of species concentration, the number of moles of the species, or the mass of the species. The choice of units depends on the context of the problem, and the type of data that is available. However, we'll consider species mole balances and species concentration balances (as these are the most common types of material balances used in flux balance analysis).

Dynamic Species Mole Balances

Suppose we have a system with volume V in which we have a set of chemical species \mathcal{M} , a set of streams \mathcal{S} , and a set of chemical reactions \mathcal{R} . Using the four terms in Defn. 1, we can write the dynamic species mole balance for species i in the system. Let n_i denote the number of moles of chemical component $i \in \mathcal{M}$ (units: mmol or μmol , etc). Further, each stream $s \in \mathcal{S}$ (flowing into or from) our volume has a direction parameter $\nu_s \in [-1, 1]$. If stream s enters the system $\nu_s = +1$, however if stream s exits the system then $\nu_s = -1$.

Definition 2 (Dynamic species mole balance). *The number of moles of chemical component n_i (unit: mol) in the system as a function of time is described by the open species mass balance equation:*

$$\sum_{s \in \mathcal{S}} \nu_s \dot{n}_{s,i} + \dot{n}_{G,i} = \frac{dn_i}{dt} \quad \forall i \in \mathcal{M} \quad (2)$$

where $\dot{n}_{s,i}$ denotes the mole flow rate of component i in stream s (units: mol i/time), $\dot{n}_{gen,i}$ denote the rate of generation of component i in the system (units: mol-i/time), and dn_i/dt denotes the rate of accumulation of the number of moles of component i in the system (units: mol-i/time).

Let's discuss the generation terms in Defn 2. Generation terms describe the impact of chemical reactions; theoretically, we can describe reactions using a mass or mole basis. However, typically we operate on a mole (or concentration) basis when dealing with chemical reactions. Thus, the species generation rate $\dot{n}_{G,i}$ can be written in terms of the open extent of reaction:

$$\dot{n}_{G,i} = \sum_{r \in \mathcal{R}} \sigma_{ir} \dot{\epsilon}_r \quad (3)$$

where σ_{ir} denotes the stoichiometric coefficient of species i in reaction r , and $\dot{\epsilon}_r$ denotes the open extent of reaction r (units: mol/time). Putting these ideas together, we can rewrite the dynamic species mole balance as:

$$\sum_{s \in \mathcal{S}} \nu_s \dot{n}_{s,i} + \sum_{r \in \mathcal{R}} \sigma_{ir} \dot{\epsilon}_r = \frac{dn_i}{dt} \quad \forall i \in \mathcal{M} \quad (4)$$

Dynamic Species Concentration Balances

When describing systems with chemical reactions where we write reaction rate expressions in terms of concentration, e.g., mole per unit volume basis, it is inconvenient to use mole or mass based units. In these cases, we need a new type of balance equation, the concentration balance. Let n_i denote the number of moles of chemical component i in the system (units: mmol or μmol). Further, denote the set of chemical species as \mathcal{M} , the set of streams flowing into (or from) the system as \mathcal{S} , and the set of chemical reactions as \mathcal{R} . Using the four terms in Defn. 1, we can write the dynamic species concentration balance for species i in the system.

Definition 3 (Dynamic Species Concentration Balance). *The number of moles of chemical component n_i in the system is described by an open species mass balance equation:*

$$\sum_{s \in \mathcal{S}} \nu_s \dot{n}_{s,i} + \dot{n}_{G,i} = \frac{dn_i}{dt} \quad \forall i \in \mathcal{M} \quad (5)$$

However, we can re-write the number of moles of species i as $n_i = C_i V \quad \forall i \in \mathcal{M}$ where C_i denotes the concentration of species i (units: mole per volume), and V (units: volume) denotes the volume of the system. Thus, we can re-write the species mole balance in concentration units as:

$$\sum_{s \in \mathcal{S}} \nu_s C_{s,i} \dot{V}_s + \dot{C}_{G,i} V = \frac{d}{dt} (C_i V) \quad \forall i \in \mathcal{M} \quad (6)$$

where \dot{V}_s denotes the volumetric flow rate for stream s (units: volume/time), $C_{s,i}$ denotes the concentration of component i in stream s (units: concentration), and $\dot{C}_{G,i}$ denotes the rate of generation of component i by chemical reaction (units: concentration/time).

Let's discuss the generation terms in Defn 3. The generation terms for species i in the concentration balance equation can be written as:

$$\dot{C}_{G,i} V = \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j V \quad \forall i \in \mathcal{M} \quad (7)$$

where σ_{ij} denotes the stoichiometric coefficient of species i in reaction j (units: dimensionless), and \hat{v}_j denotes the rate of the j th chemical reaction per unit volume (units: concentration/volume-time), and V denotes the volume of the system (units: volume). Putting these ideas together, we can rewrite the dynamic species concentration balances (in the system) as:

$$\sum_{s \in \mathcal{S}} \nu_s C_{s,i} \dot{V}_s + \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j V = \frac{d}{dt} (C_i V) \quad \forall i \in \mathcal{M} \quad (8)$$

3 Estimating Metabolic Fluxes

Metabolic fluxes are the rates of the reactions in a metabolic network, which are the unknowns that we want to estimate from various types of experimental data. You can think of metabolic fluxes as the flow of

metabolites through the metabolic network, which is analogous to the flow of water through a pipe. Thus, estimating the metabolic fluxes is important for understanding the structure and operation of a metabolic network as a function of the environment. Metabolic fluxes can be estimated using a variety of mathematical methods, such as metabolic flux analysis (MFA) and flux balance analysis (FBA). These two methods are based on different assumptions about the network, and they have different strengths and weaknesses. We are going to focus on FBA, but for those interested in MFA, we recommend the review of Antoniewicz (2).

3.1 Flux Balance Analysis

Flux balance analysis (FBA) is another mathematical method for analyzing the structure and function of metabolic networks (1). Flux balance analysis is arguably the most widely used computational tool to estimate the intracellular or cell-free reaction fluxes throughout a steady state reaction network (units: mol/volume-time). However, while we are introducing flux balance analysis in the context of metabolic network analysis, FBA is a general tool that can be used to estimate flows through many different types of networks and graphs, e.g., social graphs, communication networks, or other types of problems that can be represented as a network of graphs.

The goal of FBA is to estimate the intracellular reaction rates using whatever experimental data is available. The more data that is available, the more constraints we can put on the system, and the more accurate our estimates will be. However, one of the powerful aspects of FBA is that it can be used to estimate the metabolic fluxes in a system even when we have very little data. Suppose we have a system, with a species set \mathcal{M} , stream set \mathcal{S} and a reaction set \mathcal{R} . Further, suppose that the system (or at least part of it) is at or near a steady state. Then, the objective of the flux balance analysis is to estimate the value of the reaction rates operating in the system using linear programming. The flux balance analysis problem, whose solution provides estimates for the values of the unknown fluxes is the linear program:

$$\begin{aligned} & \text{minimize/maximize} && \sum_{i \in \mathcal{R}} c_i \hat{v}_i \\ & \text{subject to} && \sum_{s \in \mathcal{S}} \nu_s C_{s,i} \dot{V}_s + \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j V = \frac{d}{dt} (C_i V) \quad \forall i \in \mathcal{M} \\ & \text{subject to} && \mathcal{L}_j \leq \hat{v}_j \leq \mathcal{U}_j \quad \forall j \in \mathcal{R} \end{aligned}$$

The σ_{ij} are the elements of the stoichiometric matrix $\mathbf{S} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{R}|}$, the terms c_i denote the objective coefficients, \mathbf{v} denotes the unknown flux vector (the unknown that we are trying to estimate), and \mathcal{L} (\mathcal{U}) denote the permissible lower (upper) bounds on the unknown fluxes. The first set of constraints are material constraints, while the second imparts thermodynamic and kinetic information into the calculation. Finally, there are potentially other types of constraints (both linear and nonlinear) not shown that may be necessary for specific applications encountered in practice.

References

1. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245–8. doi:10.1038/nbt.1614.
2. Antoniewicz MR. A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. Metabolic Engineering. 2021;63:2–12. doi:https://doi.org/10.1016/j.ymben.2020.11.002.