

Lecture 6a: Incorporating Gene Expression into the Flux Bounds

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:

- Flux balance analysis (FBA) is a mathematical approach used to analyze the flow of metabolites through a metabolic network. It assumes a steady state where metabolite production, consumption, and transport rates are balanced. The FBA problem is formulated as a linear programming (LP) problem to maximize or minimize fluxes through the network, subject to constraints and bounds.
- **Material balance constraints** are used to ensure that the flow of metabolites into and out of a system is balanced (and physical). These constraints are typically represented as a set of linear equations, where the coefficients represent the stoichiometry of the reactions in the system in combination with transport and dilution terms.
- **Flux bounds constraints** limit the range of possible fluxes through a metabolic network. These bounds can incorporate additional information, such as experimental data or prior knowledge about the system, into the FBA problem.

1 Introduction

In this lecture, we dive deeper into the material balance constraints and the flux bounds in the flux balance analysis (FBA) framework. In particular, we will discuss incorporating gene expression data into the flux bounds.

2 Material Balance Constraints

Flux balance analysis (FBA) aims to estimate intracellular reaction rates (fluxes) using available experimental data. The more available data, the more constraints we can put on the system, the more accurate our flux estimates will be. However, one of the powerful aspects of FBA is that it can estimate metabolic fluxes even for systems with 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. The optimal reaction rates can be estimated using the following linear program (Defn. 1).

Definition 1 (Flux Balance Analysis Problem)). *The flux balance analysis problem, whose solution provides estimates for the values of the unknown fluxes, is the linear program:*

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

Here, σ_{ij} are elements of the stoichiometric matrix \mathbf{S} , c_i are objective coefficients (you choose), \hat{v}_i are unknown fluxes, and \mathcal{L}_j and \mathcal{U}_j are the lower and upper flux bounds, respectively. The d_s terms are the direction parameters for the streams, and the \dot{V}_s terms are the volumetric flow rates for the streams.

However, this formulation looks slightly different from the one in Orth et al. (1). Let's explore this further.

2.1 Simplified Species Constraints

In the Orth et al. primer (1), the material balance constraints were written as $\mathbf{S}\hat{\mathbf{v}} = 0$. This is a simplification; in reality, the material balance constraints are more complex. But let's see how we get there. When doing FBA, we often will make three assumptions:

- **Steady-state, constant volume:** The biological system (or at least part of it) is in a steady state, and in whole-cell models, the volume of the culture is constant. However, this is not the case in fed-batch cultures; many industrial biotechnology processes operate in fed-batch mode.
- **Specific units:** The volume basis for the intracellular species concentrations is in specific units, i.e., per unit cell mass measured in grams dry weight (units: gDW). For cell-free systems, the volume basis is the volume of the reaction.
- **No transport or dilution terms:** We will assume that there are no physical transport or dilution terms in the material balance equations. This is a simplification, but it is often used in FBA. For example, this assumption is not the case for continuous cell-free systems.

Derivation of the Palsson Constraints

Let the volume of our system be written in specific units, i.e., $V = B\bar{V}$, where B is the biomass concentration (units: gDW/L) and \bar{V} is the volume of the culture (units: L). The material balance constraints can be simplified by assuming the species in our system are in a steady state. But there is more to this story. Let's

expand the accumulation terms:

$$\begin{aligned}
\frac{d}{dt}(C_i V) &= \frac{d}{dt}(C_i B \bar{V}) \\
&= B \bar{V} \underbrace{\left(\frac{dC_i}{dt}\right)}_{\text{steady state} = 0} + C_i B \underbrace{\left(\frac{d\bar{V}}{dt}\right)}_{\text{steady state} = 0} + C_i \bar{V} \left(\frac{dB}{dt}\right) \\
C_i \bar{V} \left(\frac{dB}{dt}\right) &= \underbrace{\sum_{s \in \mathcal{S}} d_s C_{i,s} \dot{V}_s}_{\text{no transport} = 0} + \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j V \\
C_i \bar{V} \left(\frac{dB}{dt}\right) &= \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j B \bar{V} \\
C_i \underbrace{\left[\frac{1}{B} \left(\frac{dB}{dt}\right)\right]}_{\text{specific growth rate } \mu} &= \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j \\
C_i \mu &= \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j \\
\sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j - \underbrace{C_i \mu}_{\text{small} \ll 1} &= 0 \\
\sum_{j \in \mathcal{R}} \sigma_{ij} \hat{v}_j &= 0 \quad \forall i \in \mathcal{M} \quad \blacksquare
\end{aligned}$$

Exchange Reactions

Great! So then, does everything in our system have to be at a steady state? Not exactly. We can think of the system we are studying as an open system, i.e., it can exchange material with the surroundings. Thus, while the components of the system are in a steady state, the universe (system + surroundings) as a whole is not. This is a subtle point, but it is essential to understand. The exchange of material with the surroundings is captured in the context of our three assumptions by writing hypothetical reactions that exchange material with the surroundings. We call these reactions the **exchange reactions**. Consider the toy network in Fig. 1. The exchange reactions are written as those that cross the system's boundary.

3 Flux Bounds

The flux bounds are essential constraints in flux balance analysis calculations and the convex decomposition of the stoichiometric array. Beyond their role in the flux estimation problem, the flux bounds are integrative. These constraints integrate many types of genetic and biochemical information into the problem. A general model for these bounds is given by:

$$-\delta_j \underbrace{\left[V_{max,j}^{\circ} \left(\frac{e}{e^{\circ}} \right) \theta_j(\dots) f_j(\dots) \right]}_{\text{reverse: other functions or parameters?}} \leq \hat{v}_j \leq V_{max,j}^{\circ} \left(\frac{e}{e^{\circ}} \right) \theta_j(\dots) f_j(\dots)$$

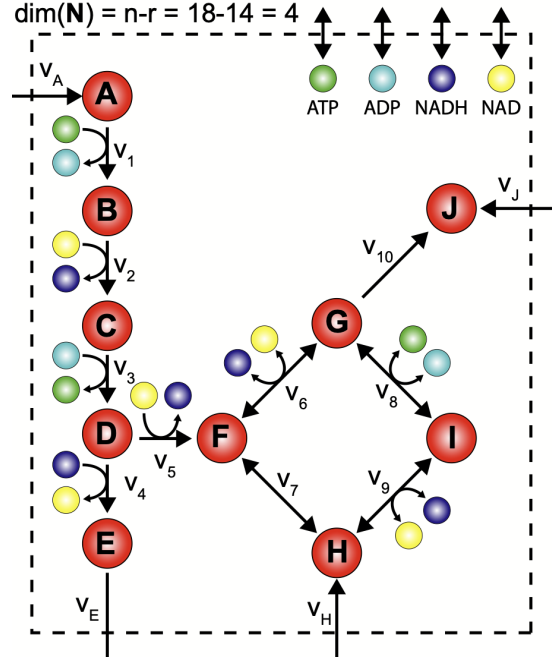


Figure 1: Example network reproduced from Bordbar et. al., (2). The exchange reactions cross the boundary of the system. We have (potentially) enzyme-catalyzed exchanges, e.g., $\{v_A, v_E, v_H, v_J\}$ and hypothetical exchanges, e.g., ATP or NADH.

where $V_{max,j}^\circ$ denotes the maximum reaction velocity (units: flux) computed at some characteristic enzyme abundance. Thus, the maximum reaction velocity is given by:

$$V_{max,j}^\circ = k_{cat,j}^\circ e^\circ$$

where $k_{cat,j}$ is the catalytic constant or turnover number for the enzyme (units: 1/time) and e° is a characteristic enzyme abundance (units: concentration). The term (e/e°) is a correction to account for the actual enzyme abundance catalyzing the reaction (units: dimensionless). The $\theta_j(\dots) \in [0, 1]$ is the current fraction of maximal enzyme activity of enzyme e in reaction j . The activity model $\theta_j(\dots)$ describes [allosteric effects](#) on the reaction rate, and is a function of the regulatory and chemical state of the system, the concentration of substrates, products, and cofactors (units: dimensionless). Finally, the $f_j(\dots)$ is a function describing the substrate (reactants) dependence of the reaction rate j (units: dimensionless).

3.1 Simplified bounds model

Let's initially assume that $(e/e^\circ) \sim 1$, there are no allosteric inputs $\theta_j(\dots) \sim 1$, and the substrates are saturating $f_j(\dots) \sim 1$. Then, the flux bounds are given by:

$$-\delta_j V_{max,j}^\circ \leq \hat{v}_j \leq V_{max,j}^\circ$$

This is a simple model for the flux bounds. It is easy to see that the flux bounds are a function of the maximum reaction velocity, the catalytic constant or turnover number, and our assumed value of a characteristic enzyme abundance.

3.2 Turnover numbers and reversibility

The turnover number, k_{cat} , measures an enzyme's catalytic activity, defined as the number of substrate molecules converted to product per enzyme molecule per unit time. Units are typically 1/time (e.g., 1/s or 1/min). Values can be obtained from primary literature or databases like [BRENDA](#) (3). The second thing we need to estimate is the reversibility parameter. The reversibility parameter δ_j can be computed in several ways. For example, one method in the literature is to use the sign of Gibbs reaction energy:

$$\delta_i = \begin{cases} 0 & \text{if } \text{sign}(\Delta G^\circ - \Delta G^*) = -1 \quad \text{irreversible} \\ 1 & \text{if } \text{sign}(\Delta G^\circ - \Delta G^*) = +1 \quad \text{reversible} \end{cases}$$

where ΔG° is the standard Gibbs free energy change of the reaction, and ΔG^* is a threshold value (hyper-parameter). The threshold value can be set to zero or some other value. Alternatively, the value of δ_j can be assigned based upon a cutoff K^* on the equilibrium constant:

$$\delta_i = \begin{cases} 0 & \text{if } K_{eq} > K^* \quad \text{irreversible} \\ 1 & \text{if } K_{eq} \leq K^* \quad \text{reversible} \end{cases}$$

where you specify the value K^* based upon some intuition or other criteria. We can compute the ΔG° values using [eQuilibrator](#) (4).

4 Gene expression models

One of the shortcomings that we discussed about flux balance analysis was: FBA may conflict with experimental data, especially when regulatory loops are excluded. These discrepancies reveal the limitations of relying only on stoichiometric information without considering complex cellular regulation. This can be (partially) fixed by including gene expression in the bounds. Gene expression is easy(ish), but allosteric regulation (activity) is hard.

Let's investigate how we could describe gene regulation in flux balance analysis. Suppose the flux problem we were interested in was composed of enzymes encoded by the genes $\mathcal{G} = 1, 2, \dots, N$. The action of each gene is described by two differential equations, one for mRNA concentration (m_j , units: nmol/gDW) and a second for the corresponding protein concentration (p_j , units: nmol/gDW):

$$\begin{aligned} \dot{m}_j &= r_{X,j} u_j(\dots) - (\theta_{m,j} + \mu) \cdot m_j + \lambda_j \quad j = 1, 2, \dots, N \\ \dot{p}_j &= r_{L,j} w_j(\dots) - (\theta_{p,j} + \mu) \cdot p_j \end{aligned}$$

Terms in the balances:

- **Transcription:** The term $r_{X,j} u_j(\dots)$ in the mRNA balance denotes the regulated rate of transcription for gene j . This is the product of a kinetic limit $r_{X,j}$ (units: nmol/gDW-time) and a transcription control function $0 \leq u_j(\dots) \leq 1$ (dimensionless). The final term λ_j is the unregulated expression rate of mRNA j (units: nmol/gDW-time), i.e., the leak expression rate.
- **Translation:** The regulated rate of translation of mRNA j , denoted by $r_{L,j} w_j$, is also the product of the kinetic limit of translation (units: nmol/gDW-time) and a translational control term $0 \leq w_j(\dots) \leq 1$ (dimensionless).
- **Degradation:** Lastly, $\theta_{*,j}$ denotes the first-order rate constant (units: 1/time) governing degradation of protein and mRNA, and μ is the specific growth rate of the cell (units: 1/time). We get the latter term using cell-specific concentration units (e.g., nmol/gDW).

4.1 Steady-state gene expression

We have publically said (without proof yet) that gene expression is slow and metabolism is fast. Suppose this is true. This means that the mRNA and protein concentrations are at an approximate steady state, i.e., $\dot{m}_j = \dot{p}_j = 0$ from the perspective of the metabolic network. This allows us to solve the gene expression equations for the steady-state mRNA and protein concentrations. Let's show the steps to compute the steady-state mRNA concentration m_j^* :

$$\begin{aligned} r_{X,j}u_j(\dots) - (\theta_{m,j} + \mu) \cdot m_j + \lambda_j &= \dot{m}_j \\ r_{X,j}u_j(\dots) - (\theta_{m,j} + \mu) \cdot m_j^* + \lambda_j &= 0 \\ r_{X,j}u_j(\dots) + \lambda_j &= (\theta_{m,j} + \mu) \cdot m_j^* \\ \frac{r_{X,j}u_j(\dots) + \lambda_j}{\theta_{m,j} + \mu} &= m_j^* \quad \text{for } j = 1, 2, \dots, N \quad \blacksquare \end{aligned}$$

Following the same steps, we can compute the steady-state protein concentration p_j^* :

$$p_j^* = \frac{r_{L,j}w_j(\dots)}{\theta_{p,j} + \mu} \quad \text{for } j = 1, 2, \dots, N \quad \blacksquare$$

Some things to think about:

- **Nonlinearity:** The expressions for m_j^* and p_j^* are trickier than they may seem at first blush. The steady-state mRNA and protein concentrations are a function of the kinetic limits of transcription and translation, the control functions, the degradation rates, and the specific growth rate of the cell. At the mRNA level, the $u(\dots)$ model could be a function of metabolite and protein concentrations. While the kinetic limit of translation $r_{L,j}$ and the $w_j(\dots)$ terms will be functions of the mRNA concentrations and other factors, such as ribosome availability, etc.
- **Models:** We need to formulate the control functions $u_j(\dots)$ and $w_j(\dots)$, and the degradation rates $\theta_{m,j}$ and $\theta_{p,j}$, and the kinetic limits $r_{X,j}$ and $r_{L,j}$ for each gene in the system. These will be functions of parameters that must be estimated from literature studies or experimental data.
- **Complication:** Finally, the enzymes catalyzing the reactions in the metabolic network are often complexes of different protein subunits, where a different gene encodes each subunit. We need to formulate the gene-protein-reaction (GPR) rules that link the genes to the proteins and the proteins to the reactions in the metabolic network. This will allow us to integrate the gene expression logic into the flux balance analysis problem.

5 Summary

This lecture discussed the material balance constraints and the flux bounds in the flux balance analysis (FBA) framework. The material balance constraints ensure that the flow of metabolites into and out of a system is balanced, while the flux bounds limit the range of possible fluxes through a metabolic network. These constraints and bounds are essential for estimating the intracellular reaction rates using FBA and integrating genetic and biochemical information into the problem. We developed some simple models for modeling gene expression and incorporating gene expression data into the flux bounds. This will allow us to incorporate gene expression control into the FBA problem and improve the accuracy of our flux estimates.

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. Bordbar A, Nagarajan H, Lewis NE, Latif H, Ebrahim A, Federowicz S, et al. Minimal metabolic pathway structure is consistent with associated biomolecular interactions. *Mol Syst Biol.* 2014;10(7):737. doi:10.15252/msb.20145243.
3. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research.* 2020;49(D1):D498–D508. doi:10.1093/nar/gkaa1025.
4. Beber ME, Gollub MG, Mozaffari D, Shebek KM, Flamholz AI, Milo R, et al. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* 2022;50(D1):D603–D609. doi:10.1093/nar/gkab1106.