

## Lecture 4: Metabolism and Metabolic Networks

*Lecturer: Jeffrey Varner*

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

## 1 Introduction

Metabolism is the set of chemical reactions that occur in living organisms to maintain life. These chemical reactions are organized into metabolic pathways, which are sequences of chemical reactions that convert a substrate into a product. Metabolic pathways are interconnected to form metabolic networks, which are the focus of this week in lecture. Metabolic networks are complex systems that are responsible for the synthesis and degradation of metabolites, which are small molecules that are intermediates or products of metabolism. Metabolic networks are also responsible for the generation of energy, which is required for the maintenance of life. This week, we will discuss the structure and function of metabolic networks, and how they can be analyzed using mathematical models. Let's get started by discussing catabolism and anabolism, which are the two main types of metabolic pathways.

## 2 Catabolism and Anabolism

Metabolic pathways can be divided into two main types: catabolic pathways and anabolic pathways. Catabolic pathways are responsible for the breakdown of complex molecules into simpler molecules, which releases energy. On the other hand, anabolic pathways are responsible for the synthesis of complex molecules from simpler molecules, which requires energy.

The main purpose of catabolic pathways is to generate energy and small molecules that can be used as building blocks for anabolic pathways. Here we will discuss three important catabolic pathways (which we'll collectively call central carbon metabolism): glycolysis, the pentose phosphate pathway and the citric acid cycle. Glycolysis is a catabolic pathway that converts sugars, e.g., glucose into **pyruvate**, which can then be further metabolized to generate energy. The first step of glycolysis is the conversion of glucose into glucose-6-phosphate, which is then converted into fructose-6-phosphate, and so on. The pentose phosphate pathway is a catabolic pathway that converts glucose-6-phosphate into ribose-5-phosphate, which is a precursor for the synthesis of nucleotides, and some aromatic amino acids. It is also an important source of **NADPH**, which is a reducing agent that is required for biosynthetic reactions. Finally, the citric acid cycle is a catabolic pathway that converts acetyl-CoA (generated from pyruvate) into **CO<sub>2</sub>**, **NADH**, **FADH<sub>2</sub>** and **GTP**. The **NADH** and **FADH<sub>2</sub>** generated by the citric acid cycle are used to generate energy through another pathway called **oxidative phosphorylation**.

Anabolic pathways are responsible for the synthesis of complex molecules from simpler molecules, which requires energy. For example, the synthesis of macromolecules such as proteins, nucleic acids, lipids and polysaccharides are examples of anabolic pathways, each of which requires energy in the form of high-energy phosphate bonds and reducing power in the form of **NADH** (**NADPH**). Anabolic pathways are often regulated by the availability of substrates and energy, as well as the demand for the end products of the pathway.

The details of the central carbon metabolism and other metabolic pathways can be found in many textbooks, such as (1) or Chapter 3 of Wünschiers et al (2). Thus, we will not discuss them in detail here, but instead abstract the structure and function of metabolic networks into a mathematical model.

### 3 Metabolic Networks

#### 3.1 Stoichiometric matrix

Suppose we have a set of chemical (or biochemical) reactions  $\mathcal{R}$  involving the chemical species (metabolite) set  $\mathcal{M}$ . Then, we can represent the metabolic network using the stoichiometric matrix  $\mathbf{S}$  (Defn 1):

**Definition 1** (Stoichiometric Matrix). *The stoichiometric matrix is a  $\mathbf{S} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{R}|}$  matrix that holds the stoichiometric coefficients  $\sigma_{ij} \in \mathbf{S}$  such that:*

- $\sigma_{ij} > 0$ : Chemical species  $i$  is produced by reaction  $j$ . Species  $i$  is a product of reaction  $j$ .
- $\sigma_{ij} = 0$ : Chemical species  $i$  is not connected with reaction  $j$
- $\sigma_{ij} < 0$ : Chemical species  $i$  is consumed by reaction  $j$ . Species  $i$  is a reactant of reaction  $j$ .

*The stoichiometric matrix is a digital representation of the metabolic network, and it can be used to analyze the structure of the network, as well as to estimate operational values, such as the metabolic fluxes.*

The stoichiometric matrix  $\mathbf{S}$  is the digital representation of the metabolic network, and it can be used to analyze some of the structural properties of the network. There is a large body of literature on the analysis of the structure of metabolic networks using various decomposition approaches of stoichiometric matrices. Let's look at one particular approach to analyzing the structure of metabolic networks, using singular value decomposition (SVD) (3).

#### Singular Value Decomposition

Suppose we have a matrix  $A \in \mathbb{R}^{m \times n}$ . The SVD of  $\mathbf{A}$  is a factorization of the form:  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ , where  $\mathbf{U} \in \mathbb{R}^{n \times n}$  and  $\mathbf{V} \in \mathbb{R}^{m \times m}$  are orthogonal matrices, i.e.,  $\mathbf{U} \cdot \mathbf{U}^T = \mathbf{I}$  and  $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$  is a diagonal matrix containing the singular values  $\sigma_i = \Sigma_{ii}$ . The matrix  $\mathbf{A} \in \mathbb{R}^{n \times m}$  can be decomposed as:

$$\mathbf{A} = \sum_{i=1}^{r_{\mathbf{A}}} \sigma_i \cdot (\mathbf{u}_i \otimes \mathbf{v}_i)$$

where  $r_{\mathbf{A}}$  is the rank of matrix  $\mathbf{A}$  and  $\otimes$  denotes the outer product. The outer product  $\hat{\mathbf{A}}_i = \mathbf{u}_i \otimes \mathbf{v}_i$  is a rank-1 matrix with elements:

$$\hat{a}_{jk} = u_j v_k \quad j = 1, 2, \dots, n \text{ and } k = 1, 2, \dots, m$$

The vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$  are the left (right) singular vectors, and  $\sigma_i$  are the singular values (ordered). Singular value decomposition is a special sort of eigendecomposition, thus, we could use QR-iteration to compute the SVD. The columns of  $\mathbf{U}$  are eigenvectors of  $\mathbf{A}\mathbf{A}^T$ , the columns of  $\mathbf{V}$  are eigenvectors of  $\mathbf{A}^T\mathbf{A}$  and the singular values  $\sigma_i$  are the square roots of the eigenvalues of the matrix products  $\mathbf{A}\mathbf{A}^T$  or  $\mathbf{A}^T\mathbf{A}$ .

#### 3.2 Material Balance Equations

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 be applied to many different types of systems, including metabolic networks. Material balances consist of 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. 2):

**Definition 2** (Material Balance Equation). *The material balance equation for species  $i$  in the biophase is given by:*

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

where the accumulation term is the rate of change of species  $i$  in the system, generation is the rate of production (consumption) by chemical reactions of species  $i$  in the system, the transport in term is the rate of transport (convection or passive diffusion) of species  $i$  into the system, and the transport out term is the rate of transport of species  $i$  out of the system.

Material balance equations can be written in terms of the concentration of the species, 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 in this lecture (as these are the most common types of material balances used in metabolic network analysis).

### Dynamic Species Mole Balances

If we are interested in the number of moles of each chemical species in the system, we can write the open species mole balance. Let  $n_i$  denote the number of moles of chemical component  $i$  in the system (units:  $\star$ moles, e.g., mmol or  $\mu$ mol). Further, denote the number of chemical components in the system we want to track as  $\mathcal{M}$ , and the set of streams **flowing** into (or from) the system as  $\mathcal{S}$ , where each stream  $s \in \mathcal{S}$  has a direction parameter  $\nu_s \in [-1, 1]$ . If stream  $s$  enters the system  $\nu_s = +1$ , however is stream  $s$  exits the system then  $\nu_s = -1$ .

**Definition 3** (Dynamic Species Mole Balance). *The number of moles of chemical component  $i$  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:  $\star$ mol i/time),  $\dot{n}_{gen,i}$  denote the rate of generation of component  $i$  in the system (units:  $\star$ mol i/time), and  $dn_i/dt$  denotes the rate of accumulation of the number of moles of component  $i$  in the system (units:  $\star$ mol i/time). Finally, let's discuss the generation terms. 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. Then, 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  $\sigma_{ir}$  denotes the stoichiometric coefficient of species  $i$  in reaction  $r$ , and  $\dot{\epsilon}_r$  denotes the open extent of reaction  $r$  (units:  $\star$ 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:  $\star$ moles, e.g., mmol or  $\mu$ mol). 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  $\mathcal{R}$ .

**Definition 4** (Dynamic Species Concentration Balance). *The number of moles of chemical component  $i$  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:

$$n_i = C_i V \quad \forall i \in \mathcal{M} \quad (6)$$

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:

$$\sum_{s \in \mathcal{S}} \nu_s C_{s,i} \dot{V}_s + \dot{C}_{G,i} V = \frac{d}{dt} (C_i V) \quad \forall i \in \mathcal{M} \quad (7)$$

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). The generation terms for species  $i$  in the concentration balance equation can be written as:

$$\dot{C}_{G,i} V = \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{r}_j V \quad \forall i \in \mathcal{M} \quad (8)$$

where  $\sigma_{ij}$  denotes the stoichiometric coefficient of species  $i$  in reaction  $j$ , and  $\hat{r}_j$  denotes the rate of reaction  $j$  (units: concentration/time). Putting these ideas together, we can rewrite the dynamic species concentration balances (in the system) as:

$$\sum_{s \in \mathcal{S}} \nu_s C_{s,i} \dot{V}_s + \sum_{j \in \mathcal{R}} \sigma_{ij} \hat{r}_j V = \frac{d}{dt} (C_i V) \quad \forall i \in \mathcal{M} \quad (9)$$

## 4 Timescales in Metabolic Networks

Metabolic systems operate on a wide range of timescales, which can be important for understanding the structure and function of the system. Metabolic reactions are fast, but transcription and translation are slow, which can lead to a mismatch between the timescales of the reactions and the timescales of the gene

expression. However, that mismatch has a practical consequence: the timescales of the reactions and the timescales of the gene expression can be decoupled, which allows us to think about these two processes separately. First, let's check the assertion that metabolic reactions are fast, while gene expression is slow.

## 4.1 Timescales of Metabolic Reactions

If we think about some reaction  $r \in \mathcal{R}$  in a metabolic network that depends upon concentration vector  $\mathbf{x}$ , the specific rate of the reaction is governed by a rate law  $\hat{r}_r$  (units:  $\star\text{mol}/\text{volume}\cdot\text{time}$ ) of the form:

$$\hat{r}_r = k_r \cdot E_r \cdot v_r(\mathbf{x}) \cdot g(\mathbf{x}, \theta) \quad (10)$$

where  $k_r$  denotes the rate constant of reaction  $r$  (units:  $1/\text{time}$ ),  $E_r$  denotes the enzyme concentration (units:  $\star\text{mol}/\text{volume}$ ), and  $g(\mathbf{x}, \theta)$  denotes the rate law function that depends on the concentration vector  $\mathbf{x}$  and some (unknown) parameters  $\theta$ . The rate law function  $0 \leq g(\mathbf{x}, \theta) \leq 1$  can be a simple function of the concentration of the reactants, or it can be a complex function that depends on the concentration of the reactants and other factors, such as temperature, pH, or the presence of other molecules. However, the maximum value of the rate law function is 1, which means that the rate of the reaction is limited by the rate constant and the enzyme concentration.

## 5 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. Let's start by discussing metabolic flux analysis, and conclude with a discussion of flux balance analysis.

### 5.1 Metabolic Flux Analysis

Metabolic flux analysis is a powerful tool for analyzing the structure and function of metabolic networks. The goal of metabolic flux analysis is to estimate the rates of the reactions in the network, which are called metabolic fluxes, from experimental data, in particular, isotopic labeling data or exchange flux data (rates of uptake or secretion of metabolites). The key assumption of metabolic flux analysis is that the network is at (or near) a steady state, which means that the concentrations of the metabolites and enzymes in the network are approximately constant over time. Under this assumption, the rate of accumulation of each metabolite (enzyme) in the network is zero, which allows us to write a set of linear algebraic equations that describe the network flux. Traditional metabolic flux analysis uses a stoichiometric model of the network, which is a set of linear equations that describe the relationships between the metabolites and reactions in the network. Alternatively, labeled isotopes can be used to estimate the metabolic fluxes in the network, which is called isotopic flux analysis.

#### Traditional Metabolic Flux Analysis

One of the first examples of what we might recognize as traditional metabolic flux analysis was published by Vallino and Stephanopoulos (4) (although there were several earlier studies from Papoutsakis et al (5, 6)

that used similar ideas). Vallino and Stephanopoulos used a stoichiometric model of the metabolism of *Corynebacterium glutamicum* in combination with exchange rate measurements to estimate the intracellular fluxes occurring during growth and lysine production.

The basic idea behind traditional metabolic flux analysis is to use the stoichiometric model of the network to write a set of linear equations that describe the metabolic flux through the network.

### Isotopic Metabolic Flux Analysis

Isotopic flux analysis uses the labeling patterns of metabolites to estimate the metabolic fluxes in the network. This is a very complex subject, and we will not discuss it in detail here, but the basic idea is that the labeling patterns of metabolites can be used to estimate the metabolic fluxes in the network. A recent review of isotopic flux analysis can be found was recently published de Falco et al (7) (although this review does not cover many of the important early studies in the field, not does it cover the mathematical basis of the method).

## 5.2 Flux Balance Analysis

Flux balance analysis (FBA) is another mathematical method for analyzing the structure and function of metabolic networks (8). Flux balance analysis is arguably the most widely used computational tool to estimate the intracellular or cell-free reaction rates (which are called fluxes) throughout a steady state reaction network (units:  $\star\text{mol}/\text{volume}\cdot\text{time}$ ). 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.

Like metabolic flux analysis, the goal of FBA is to estimate the rates of the reactions in the network from experimental data. However, FBA makes some additional simplifying assumptions about the network, which allows for the use of linear programming to solve for the metabolic fluxes. 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:

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

The matrix  $\mathbf{S}$  is the  $\mathcal{M} \times \mathcal{R}$  stoichiometric matrix,  $c_i$  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.

## 6 Summary

In this lecture, we discussed the structure and function of metabolic networks, and how they can be analyzed using mathematical models.

## References

1. Palsson BØ. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge University Press; 2015.
2. Wünschiers R, Jahn M, Jahn D, Schomburg I, Peifer S, Heinzle E, et al. 3. In: Metabolism. John Wiley & Sons, Ltd; 2012. p. 37–209. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118657072.ch3>.
3. Famili I, Palsson BO. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices. *J Theor Biol.* 2003;224(1):87–96. doi:10.1016/s0022-5193(03)00146-2.
4. Vallino JJ, Stephanopoulos G. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol Bioeng.* 1993;41(6):633–646. doi:10.1002/bit.260410606.
5. Papoutsakis ET, Meyer CL. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotechnol Bioeng.* 1985;27(1):50–66. doi:10.1002/bit.260270108.
6. Papoutsakis ET, Meyer CL. Fermentation equations for propionic-acid bacteria and production of assorted oxychemicals from various sugars. *Biotechnol Bioeng.* 1985;27(1):67–80. doi:10.1002/bit.260270109.
7. de Falco B, Giannino F, Carteni F, Mazzoleni S, Kim DH. Metabolic flux analysis: a comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas. *RSC Adv.* 2022;12:25528–25548. doi:10.1039/D2RA03326G.
8. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol.* 2010;28(3):245–8. doi:10.1038/nbt.1614.