

# Mathematical Modeling of Metabolic Flux: Integrating COBRApy and Escher for Enhanced Analysis

João S. Lopes, José A. Rodrigues and Jacinta Serpa

**Abstract**—Metabolic flux analysis (MFA) is a widely used computational approach to investigate cellular metabolism by predicting the rates of metabolic reactions in a given network. In this study, we applied Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA) using a genome-scale metabolic model of *Escherichia coli* to illustrate the theoretical application of these methods in microbial metabolism. Our goal was to analyze key metabolic pathways under different simulated conditions and evaluate cellular growth rates based on biomass production. FBA was employed to optimize biomass production as the objective function, while FVA was used to explore the flexibility of the metabolic network and the range of allowable flux distributions. The analyses provided insights into substrate utilization, metabolic efficiency, and the overall adaptability of the cell's metabolism under various simulated scenarios. This theoretical framework demonstrates the utility of FBA and FVA, as implemented through COBRApy and visualized using Escher, in understanding and optimizing microbial metabolic networks.

**Keywords**—Metabolic flux analysis (MFA), Flux Balance Analysis (FBA), Cancer metabolism, COBRApy, Escher Builder.

## I. INTRODUCTION

The study of metabolic networks is central to understanding cellular physiology, as it provides insight into how cells convert nutrients into energy and biomass. Microorganisms like *Escherichia coli* (*E. coli*) serve as excellent model organisms for exploring metabolism due to their well-characterized genetic and biochemical networks. With the advent of genome-scale metabolic models (GEMs), computational approaches such as Metabolic Flux Analysis (MFA) have become essential tools for studying metabolism on a system-wide scale. MFA allows researchers to quantitatively predict the rates of intracellular metabolic reactions by integrating constraints derived from experimental data, thermodynamic limits, and stoichiometry, thus offering a comprehensive understanding of the cell's metabolic behavior under different conditions.

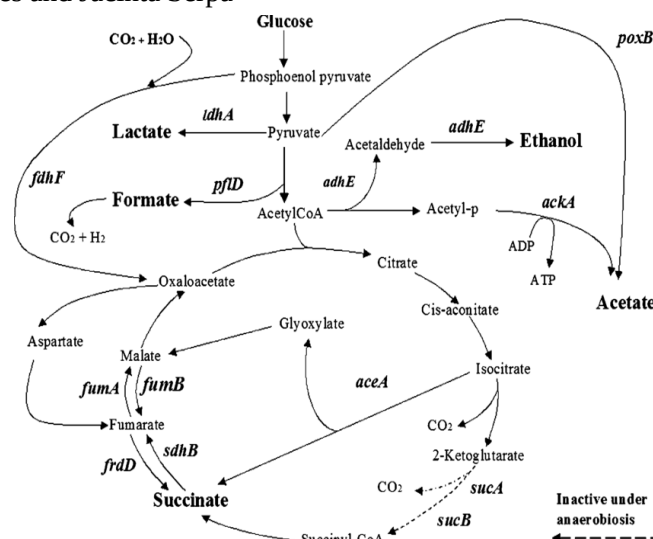


Fig. 1. Simplified *E. coli* Core Metabolism. Lara et al. (2006). *Biotechnology and Bioengineering*, 93(2), 372-385.

Flux Balance Analysis (FBA) is one of the most widely used methods within MFA. FBA is grounded in the assumption of a steady-state system, where the accumulation of metabolites is minimal, and it predicts the flow of metabolites through a network by optimizing a specific objective function, such as growth rate or ATP production. FBA has been applied extensively to microbial systems, particularly for *E. coli*, to study a wide range of cellular phenomena, from optimizing metabolic pathways for genetic perturbations to investigating the cell's responses to environmental stress. The accuracy and versatility of FBA have made it a fundamental tool in metabolic engineering, synthetic biology, and bioprocess optimization [1].

Flux Variability Analysis (FVA) complements FBA by analyzing the flexibility of the metabolic network. While FBA provides a single optimal solution, FVA explores the range of flux values that can achieve the same objective, allowing for a deeper understanding of the alternative metabolic pathways that a cell may employ to sustain growth. This is particularly useful for microbial systems like *E. coli*, where multiple metabolic routes can be activated depending on environmental conditions. FVA, therefore, provides a more comprehensive view of the network's capacity and robustness, highlighting metabolic plasticity that could be leveraged in biotechnological applications [2].

The implementation of these methods is made accessible

Manuscript received October 16, 2024

João S. Lopes (Author), Master in Biomedical Engineering, Instituto Superior de Engenharia de Lisboa, Portugal

José A. Rodrigues (Author), CIMA and Department of Mathematics of ISEL-Higher Institute of Engineering of Lisbon, Portugal

Jacinta Serpa (Author), NOVA Medical School—Faculdade de Ciências Médicas, Lisbon Portugal

through tools such as COBRApy, a Python-based framework designed for constraint-based modeling and analysis of metabolic networks [3]. COBRApy enables the simulation of metabolic fluxes, gene knockouts, and pathway optimizations, making it widely adopted for studying microbial metabolism. Additionally, visualization tools such as Escher provide intuitive and interactive ways to visualize metabolic pathways and flux distributions, enhancing the ability to interpret and communicate results [4]. Escher's interactive maps allow users to visualize reaction fluxes dynamically, offering clear insights into how changes in environmental conditions or genetic modifications affect metabolic pathways.

In this study, we demonstrate the theoretical application of FBA and FVA to *E. coli* metabolism, focusing on the prediction of growth and substrate utilization. The integration of COBRApy and Escher facilitates not only the computational analysis but also the intuitive visualization of these complex networks. By modeling the behavior of *E. coli* under various simulated conditions, we aim to illustrate how these methods can be applied to a wide range of microbial systems, providing valuable insights into cellular metabolism, metabolic engineering, and potential industrial applications such as biofuel production and bioremediation [5, 6]. The general applicability of these approaches makes them powerful tools in both fundamental research and applied biotechnology.

## II. METHODS

### A. Metabolic Model and Data Input

The metabolic flux analysis (MFA) was performed using the COBRApy library, which allows for the simulation of cellular metabolism based on a stoichiometric model of biochemical reactions. A pre-existing genome-scale metabolic model was used, containing reactions, metabolites, and stoichiometric coefficients that represent cellular metabolism in a generic organism, such as *Escherichia coli* or other microorganisms. This model can be adapted to study various biological systems by adjusting constraints based on specific environmental conditions or organism characteristics.

- 1) COBRApy: Used for metabolic model analysis and simulation.
- 2) cobraio load\_model: Loads predefined metabolic models available in public databases.
- 3) cobra.flux\_analysis.flux\_variability\_analysis: Performs flux variability analysis on the network.
- 4) escher.Builder: Visualizes flux distributions in metabolic networks.

For the analysis:

- 1) Reaction fluxes were constrained by predefined upper and lower limits to simulate nutrient availability and other environmental conditions.
- 2) The objective function of the model was set to maximize biomass production, a common proxy for growth rate in microbial systems.

This approach can be applied to various conditions, ranging

from optimal growth scenarios to stress conditions, by modifying the model's environmental inputs and nutrient constraints.

### B. Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) was used to estimate the steady-state flux distribution of metabolic reactions. The FBA problem is solved by optimizing a linear objective function, commonly the biomass production, under the assumption that the system operates in a steady state where metabolites do not accumulate. In this study, the cobra.flux\_analysis.pfba() function was employed, ensuring mass balance across all reactions in the network. The resulting flux vector provided the optimal rates for each reaction in the network, given the imposed constraints. FBA is a general approach that can be applied to various biological systems, ranging from microbial organisms to mammalian cells, by adjusting the model to represent the specific organism or metabolic pathway of interest.

### C. Flux Variability Analysis (FVA)

Flux Variability Analysis (FVA) was conducted to explore the flexibility of the metabolic network by determining the range of possible flux values for each reaction while maintaining the optimal objective value. FVA was performed using the cobra.flux\_analysis.flux\_variability\_analysis() function. The analysis can reveal the robustness of a network by highlighting reactions with significant flexibility, as well as those that are tightly constrained. For this study, FVA was executed with a constraint that preserved at least 90

FVA is particularly useful for understanding the alternative pathways that a system can utilize under different environmental or genetic perturbations, making it a versatile tool for both theoretical and practical applications in biotechnology and metabolic engineering.

### D. Growth Rate and Substrate Consumption

The growth rate of the organism was estimated by calculating the flux through the biomass production reaction. The model also examined the fluxes through exchange reactions, which represent nutrient uptake and by-product secretion. The total consumption of key substrates, such as glucose and oxygen, was analyzed, and the ratio of biomass production to substrate consumption was calculated to assess the metabolic efficiency of the system under different conditions.

This analysis can be generalized to various types of organisms by altering the exchange reactions to reflect the substrates relevant to the specific system being studied.

### E. Visualization of Metabolic Flux Distribution

The Escher visualization tool was used to display the metabolic flux distribution across key pathways. For example, in microbial systems, maps of central carbon metabolism, such as glycolysis, the TCA cycle, and the pentose phosphate pathway, can be visualized. The tool enables users to see how fluxes are distributed under different conditions, providing a

clear visual representation of changes in metabolic activity. Escher's capabilities are generalizable to a wide range of models and can be applied to any organism for which a metabolic model exists.

#### F. Visualization of Metabolic Flux Distribution

The Escher Builder tool from COBRApy was used to visualize the distribution of metabolic fluxes across the network. The "RECON1.Glycolysis TCA PPP" map was employed to display how the fluxes changed between the control and therapeutic conditions. This visualization helped highlight key differences in pathway activity, particularly within glycolysis and the TCA cycle, under both conditions.

#### G. Optimization Methods

Linear programming techniques such as the Simplex Method and the Interior-Point Method were employed to optimize key metabolic fluxes. These optimization techniques are commonly used in FBA to solve the linear programming problem that maximizes the objective function, such as growth rate or ATP production, subject to metabolic constraints. The Simplex Method iteratively selects the optimal solution by adjusting fluxes within the feasible space, while the Interior-Point Method explores the solution space from within the feasible region.

These methods are widely applicable to any system where optimization of metabolic fluxes is required, providing a robust framework for studying metabolic networks.

Structure of the Linear Programming Problem A typical linear programming problem can be described as follows:

Objective Function: Maximize or minimize:  $Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$  Where  $Z$  is the objective function and  $c_1, c_2, \dots, c_n$  are the coefficients of the function.

##### Constraints:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m$$

$$x_1, x_2, \dots, x_n \geq 0$$

Where  $a_{ij}$  are the coefficients of the variables  $x_i$  in the constraints, and  $b_1, b_2, \dots, b_m$  are the limits of the constraints.

This linear programming problem can be described in the following vectorial form:

Maximize or minimize:

$$Z = \mathbf{c}^T \mathbf{x}$$

with  $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$  and  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ .

The constraints can be imposed, in matrix-vector form, by:

$$\mathbf{A} \mathbf{x} \leq \mathbf{b}$$

where

- $\mathbf{A}$  is an  $m \times n$  matrix of coefficients, with elements  $a_{ij}$  representing the coefficients of  $x_j$  in constraint  $i$ ,

- $\mathbf{b} = [b_1, b_2, \dots, b_m]^T$  is the vector of constraint limits. Additionally, the non-negativity constraint:

$$\mathbf{x} \geq 0$$

indicates that  $x_1, x_2, \dots, x_n \geq 0$ .

#### H. Background on COBRApy and Escher

COBRApy and Escher are open-source tools designed for metabolic network analysis and visualization. COBRApy provides a framework for performing FBA and FVA on genome-scale metabolic models, while Escher allows for the interactive visualization of flux distributions. These tools are highly adaptable, allowing researchers to study a wide variety of metabolic systems, from simple bacterial models to complex eukaryotic networks.

### III. THEORETICAL TEST RESULTS INTERPRETATION

We explore the application of two different optimization techniques to a metabolic network model: classical Linear Programming (LP) optimization and Flux Balance Analysis (FBA) using COBRApy. The optimization is performed on an *E. coli* core model, with the objective of maximizing biomass production. We will focus on the mathematical formulation of both methods, with special emphasis on the extraction of the stoichiometric matrix and objective coefficients from the COBRApy model.

#### A. Stoichiometric Matrix

In metabolic network modeling, the stoichiometric matrix,  $S$ , defines the relationship between metabolites and reactions (which corresponds to matrix  $A$  in the theoretical example of subsection II.G.). The matrix has dimensions  $m \times n$ , where  $m$  is the number of metabolites and  $n$  is the number of reactions. Each element  $S_{ij}$  represents the stoichiometric coefficient of metabolite  $i$  in reaction  $j$ .

In the context of COBRApy, the stoichiometric matrix can be extracted using the following command:

`S=cobra.util.create_stoichiometric_matrix(model)`

where `model` is the COBRApy representation of the *E. coli* metabolic network.

#### B. Objective Function

The goal of optimization is to maximize the flux through a particular reaction (in this case, biomass production). The objective function can be expressed on the vectorial form as:

$$Z = \mathbf{c}^T \mathbf{v}$$

where:

- $\mathbf{c}$  is a vector of objective coefficients (typically 0 or 1), indicating the reactions that contribute to the objective (e.g., biomass production).
- $\mathbf{v}$  is the vector of reaction fluxes.

In COBRApy, the objective coefficients can be extracted with the following command:

```
objective\_coeffs = np.array([rxn.objective\_coefficient for rxn
in model.reactions])
```

Here, the objective coefficients define the weight of each reaction in the objective function, and for maximizing biomass production, the biomass reaction will have a non-zero coefficient.

### C. Linear Programming Optimization

The classical Linear Programming problem can be formulated as:

$$\text{maximize } Z = \mathbf{c}^T \mathbf{v}$$

subject to:

$$\mathbf{S}\mathbf{v} = \mathbf{0}$$

and bounds on the fluxes:

$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u}$$

where:

- $\mathbf{v}$  is the vector of fluxes to be optimized.
- $\mathbf{S}$  is the stoichiometric matrix.
- $\mathbf{l}$  and  $\mathbf{u}$  represent the lower and upper bounds for each flux.

Most solvers, including `scipy.optimize.linprog`, are designed to minimize an objective function. To convert our maximization problem into a minimization problem, we minimize the negative of the objective function:

$$\text{minimize } -\mathbf{c}^T \mathbf{v}$$

This is implemented in the following Python command:

```
result = linprog(-objective\_coeffs, A\_eq=S,
b\_eq=np.zeros(S.shape[0]), bounds=bounds, method='highs')
```

### D. Flux Balance Analysis

Flux Balance Analysis is a constraint-based optimization technique tailored for metabolic network models. FBA solves the following optimization problem:

$$\text{maximize } Z = \mathbf{c}^T \mathbf{v}$$

subject to the steady-state constraint:

$$\mathbf{S}\mathbf{v} = \mathbf{0}$$

and the same flux bounds:

$$\mathbf{l} \leq \mathbf{v} \leq \mathbf{u}$$

COBRApy simplifies the process of running FBA with the following command:

```
solution = model.optimize()
```

This command automatically extracts the stoichiometric matrix, applies the constraints, and optimizes the objective function (biomass in this case).

## IV. COMPARISON OF RESULTS

We applied both optimization techniques (classical LP and COBRApy FBA) to the *E. coli* core model. The results of the objective values from both methods are as follows:

- **Objective value from LP optimization:** 0.8739215069684304
- **Objective value from COBRApy FBA:** 0.8739215069684295

The objective values obtained from both methods are essentially identical, with a negligible numerical difference (on the order of  $10^{-13}$ ). This small difference is likely due to solver-specific numerical tolerances rather than any significant difference in the underlying mathematical approach.

The results from the *E. coli* metabolic model analysis provide key insights into the behavior of the network under optimized conditions. The objective function, which maximizes biomass production, resulted in a value of 0.2718, indicating the model's predicted growth rate under the given conditions. This value serves as a baseline for evaluating the system's efficiency.

### A. Key Reaction Fluxes

Several central metabolic reactions showed significant fluxes:

- **PFK (Phosphofructokinase):** A flux of 9.68 reflects active glycolysis, indicating that the model prioritizes glucose metabolism to fuel growth.
- **PGI (Phosphoglucose isomerase):** With a flux of 9.79, PGI further emphasizes glycolytic activity, showing high glucose conversion into downstream intermediates.
- **PGK (Phosphoglycerate kinase):** A flux of -19.22 suggests robust ATP generation, essential for biomass formation, through substrate-level phosphorylation.
- **PYK (Pyruvate kinase):** The flux of 3.87 highlights the crucial role of pyruvate generation, leading to both energy production and precursor supply for the TCA cycle.

The flux in ATP Synthase (ATPS4r) at 24.47 suggests a high rate of oxidative phosphorylation, indicating that the electron transport chain is heavily utilized for ATP production, further driving cell growth.

### B. Flux Variability Analysis (FVA)

FVA results revealed notable flexibility in several key reactions:

- **PFK:** The flux range of [2.92, 41.43] suggests that glycolysis can operate over a wide range of activities without significantly affecting growth. This indicates metabolic plasticity, allowing the cell to adjust its flux depending on environmental conditions while maintaining near-optimal growth.
- **PGI:** With a narrower range of [3.09, 9.94], the reaction shows less flexibility, implying a more tightly regulated role in glucose metabolism.

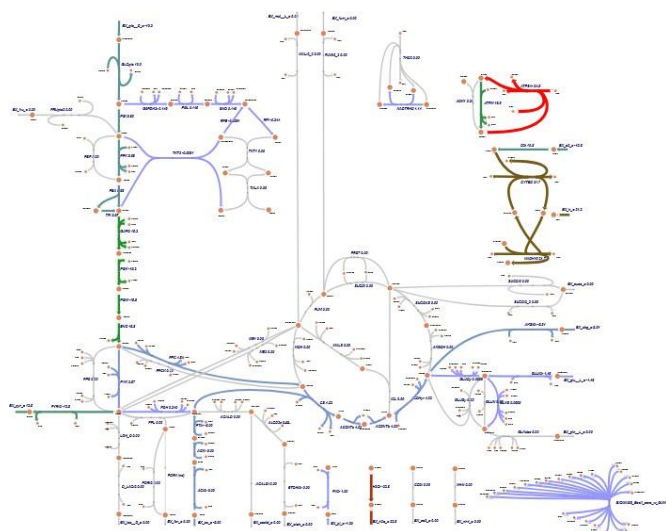


Fig. 2. Escher Simulation: E.Coli Core Metabolism

- ATP Synthase (ATPS4r): A wide range of [12.06, 40.00] highlights the cell's ability to vary its energy production through oxidative phosphorylation while maintaining robust growth.

### C. Growth Rate and Metabolic Efficiency

The calculated growth rate of 0.0024 units based on biomass production and substrate consumption is consistent with a model operating under efficient metabolic conditions. This suggests that *E. coli* can sustain growth with the available nutrients, particularly glucose and oxygen, without excessive waste production. The model prioritizes the generation of ATP and biomass precursors, indicating that the system is well-optimized for aerobic growth conditions.

### D. Key Pathways in *E. coli* Core Metabolism using Escher:

- Glycolysis: Converts glucose to pyruvate, generating ATP and NADH. Key enzymes include hexokinase, phosphofructokinase, and pyruvate kinase.
- Pentose Phosphate Pathway (PPP): Produces NADPH and ribose-5-phosphate for biosynthesis. Key enzyme: glucose-6-phosphate dehydrogenase.
- TCA Cycle: Generates NADH, FADH<sub>2</sub>, and GTP from acetyl-CoA. Important enzymes: citrate synthase, isocitrate dehydrogenase, and alpha-ketoglutarate dehydrogenase.
- Oxidative Phosphorylation: Synthesizes ATP via NADH and FADH<sub>2</sub> from glycolysis and the TCA cycle. Key components: electron transport chain complexes and ATP synthase.
- Amino Acid Biosynthesis: Branches from TCA intermediates to produce amino acids like glutamate and aspartate.

### E. Visual Observations from the Map:

- Blue/Red Arrows: Indicate reaction direction (reversible/irreversible).
- Nodes: Represent intermediate metabolites, with branching points leading to different pathways (e.g., TCA, fermentation).
- Cycle Clusters: Likely represent cyclic pathways like the TCA cycle.

The map matches *E. coli* metabolism, showing conserved pathways like glycolysis, PPP, and the TCA cycle. In aerobic conditions, the TCA cycle and oxidative phosphorylation dominate, while in anaerobic conditions, *E. coli* shifts to fermentation, producing lactate, ethanol, or acetate [9][10].

## V. CONCLUSION

This study effectively demonstrates the application of Metabolic Flux Analysis (MFA), specifically through Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA), to explore and optimize the metabolic network of *Escherichia coli*. By utilizing genome-scale metabolic models and computational tools such as COBRApy and Escher, we elucidated key metabolic pathways and their flexibility under varying conditions, thereby enhancing our understanding of microbial metabolism. The findings indicate that *E. coli* exhibits significant metabolic plasticity, allowing it to adapt its flux distributions while maintaining optimal growth rates.

The results highlight the critical role of central metabolic reactions, such as glycolysis and oxidative phosphorylation, in sustaining cellular growth and energy production. Moreover, the ability to assess the range of flux values through FVA provides invaluable insights into metabolic robustness and the potential for biotechnological applications. Overall, the theoretical framework established in this study not only reinforces the utility of FBA and FVA in metabolic engineering but also paves the way for future investigations aimed at optimizing microbial and eukaryotic cells on medical settings.

## REFERENCES

- [1] Orth, J.D., Thiele, I., & Palsson, B.O. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3), 245-248. <https://doi.org/10.1038/nbt.1614>
- [2] Mahadevan, R., & Schilling, C.H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4), 264-276. <https://doi.org/10.1016/j.ymben.2003.09.002>
- [3] Ebrahim, A., Lerman, J.A., Palsson, B.O., & Hyduke, D.R. (2013). COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7:74. <https://doi.org/10.1186/1752-0509-7-74>
- [4] King, Z.A., Draeger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., & Palsson, B.O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Computational Biology*, 11(8), e1004321. <https://doi.org/10.1371/journal.pcbi.1004321>
- [5] Feist, A.M., & Palsson, B.O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3), 344-349. <https://doi.org/10.1016/j.mib.2010.03.003>
- [6] Zhuang, K., Vemuri, G.N., & Mahadevan, R. (2011). Economics of membrane occupancy and respiro-fermentation. *Current Opinion in Microbiology*, 14(5), 619-624. <https://doi.org/10.1016/j.mib.2011.09.009>

- [7] COBRApy Documentation. (n.d.). COBRApy: A Python package for Constraint-Based Reconstruction and Analysis. [Online]. Available: <https://opencobra.github.io/cobrapy/>
- [8] Escher Documentation. (n.d.). Escher: A web-based tool for creating and sharing interactive maps of metabolic pathways. [Online]. Available: <https://escher.github.io/>
- [9] Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., & Palsson, B.O. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Molecular Systems Biology*, 7(1), 535. <https://doi.org/10.1038/msb.2011.73>
- [10] Palsson, B.O. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139854610>