CS524 Project: Flux Balance Analysis for Metabolic Networks

Chengwei Dong

A. What is the issue being addressed? Biological systems has numerous interconnected metabolic (chemical) reactions that convert nutrients into energy, biomass, and other products. Understanding how the reaction fluxes are distributed across a metabolic network is an important question in biology. **Flux Balance Analysis (FBA)** is a computational method to model the reaction fluxes in a metabolic network at steady state.[1] It uses a complete set of all reactions of a cell/organism, and formulate a linear program model to optimize a biological objective such as growth rate, and hence obtaining a metabolic network model with optimal flux distribution. In this project, I want to use the idea of FBA to construct metabolic models using public data, and then perform subsequent analyses.

B. Where does the data come from and how will it be obtained? The data come from metabolic networks from the **BiGG Models** database (https://bigg.ucsd.edu) which is freely accessible. I will first use *E. coli* (one of the simplest and best-studied organisms) and later explore larger-scale networks afterwards. This database contains everything I need to build the model and it allows comparison of my results with published models.

C. What is the optimization problem underlying this project? The underlying optimization problem is a linear program (LP) on the reaction network. The decision variables are the reaction fluxes (chemical reaction rates) v for all reactions. The constraints are (1) **Steady state**: all intermediate substances (metabolites) have constant concentrations, meaning that for each metabolite, the total flux producing it equals the total flux consuming it; (2) **Reaction bounds**: each flux is limited by some physical constraints. The objective is to **maximize biomass production**, represented as an artificial "biomass reaction", which corresponds to the *growth rate* of the cell. This problem is conceptually similar to a max network flow problem, but the difference is that, in the view of chemical reactions, fluxes can be negative (reversible reactions).

D. What are the deliverables? First, I will construct a metabolic network model for *E. coli* and estimate its growth rate, and apply the method to larger single-cell organisms (with thousands of reactions). Based on the FBA models, I will analyze the **critical reactions** by constraining each flux to zero and examining its effect on the optimal objective value. I will also perform so-called **Flux Variablility Analysis** to determine the available range of each flux. Essentially, this is fixing the optimal objective and solving two LPss (min and max) for each flux.

E. Other points for me to consider when evaluating. (1) Performance and scalability. Because the model may contain thunsands of reactions and substances, my program should efficiently handle large-scale LPs. I can compare different solvers and options to evaluate the performance. (2) Validation. I can compare my model with published FBA results from the BiGG database to check if it captures the realistic behaviors.

[1] Orth, J., Thiele, I. & Palsson, B. *Nat Biotechnol* **28**, 245–248 (2010). https://doi.org/10.1038/nbt.1614