EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATHEMATISCH-
NATURWISSENSCHAFTLICHE
FAKULTÄT

Master Thesis in Bioinformatics

Visual analysis of flux sampling simulations based on extracellular metabolomic measurments

Mentors:
Jun.-Prof. Dr. Michael Krone, Prof. Dr. Cora Weigert,
and Jun.-Prof. Dr. Andreas Dräger

February 2020

1 Background and Motivation

Background and Motivation Nowadays, there is a lot more biological data generated through high-throughput methods. Especially exciting in this context are the biochemical changes that lead to specific biological phenotypes. These can be influenced by disease states and environmental states, such as diet, exercise, and lifestyle. One can, therefore, study the metabolome under changing environments in order to associate them with each other [7]. The vast amount of knowledge led to the generation of genome-scale metabolic models (GEM) [3]. With these, one can simulate *in silico* flux states of reactions between metabolites in biological systems. Interaction with an environment can be reflected by constraining exchange fluxes on the basis of extracellular metabolomic measurements [10]. This approach is particularly useful since it is a way to analyze biological systems under the environmental influence in a non-invasive manner. There are currently different kinds of metabolic *in silico* simulation methods. The most popular are flux balance analysis (FBA) [11] and flux variability analysis (FVA) [8]. FBA computes steady-state solutions that optimize an objective function, as for maximum biomass production [4]. Nevertheless, often not only one solution occurs, but a set of steady-state solutions. This variability is being reflected in an FVA analysis that outputs the range of feasible flux states in a biological system optimized for an objective function. The use of an objective function nevertheless introduces bias into the analysis [6] [5]. Maximum biomass, for example, might be the ideal objective function in ideal laboratory conditions. However, wildlife organisms might be more generally adapted to the environment. Thus, balancing a multitude of different objective functions. Additionally, it is questionable if used objective functions reflect the true objective functions in the organism. A more unbiased approach is therefore

flux sampling [13] [12]. Flux sampling creates a distribution of feasible flux states by fully sampling the solution space under network constraints. This approach, besides being unbiased, not only provides a range of flux solutions as FVA but also introduces a way to analyze their probability of appearance in the system. For simple organisms, their metabolic pathways can easily be visually analyzed. However, especially for *Homo sapiens*, the number of reactions, metabolites, and genes are in the thousands [1]. For these large scale models, there is a need to be able to analyze this vast set of fluxes in an exploratory manner.

Aim

The aim of this master thesis project is to develop a visual analytics web application. This web application should load extracellular metabolomic measurement data and genome-scaled metabolomic models, create flux sampling simulations from the data and be used for visual exploration of the simulation results.

Approach

In their work, Mo et al. [10] describe a way to connect extracellular metabolite measurements with intracellular fluxes by constraining exchange fluxes. Additionally, they describe using derived z-scores from the distribution of randomly chosen flux differences between two conditions in order to evaluate reactions, metabolites, and subsystems. These approaches will be used for simulation and visualization. The flux sampling itself will be conducted using COBRApy [2] and the optimized general parallel sampler (OPTGP) [9] or other sampling algorithms. On the visualization part, the individual metabolisms of the GEM should be visualized by z-scores in an overview fashion. Pathway visualizations utilizing KEGG creates a way to zoom into the metabolisms, and details like the individual flux distributions could be shown on demand.

References

[1] Elizabeth Brunk, Swagatika Sahoo, Daniel C Zielinski, Ali Altunkaya, Andreas Dräger, Nathan Mih, Francesco Gatto, Avlant Nilsson, German Andres Preciat Gonzalez, Maike Kathrin Aurich, et al. Recon3d enables a three-dimensional view of gene variation in human metabolism. *Nature biotechnology*, 36(3):272, 2018.

[2] Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. Cobrapy: Constraints-based reconstruction and analysis for python. *BMC systems biology*, 7(1):74, 2013.

[3] Jeremy S Edwards and Bernhard O Palsson. Systems properties of the haemophilus influenzaerd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–17416, 1999.

[4] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3): 344–349, 2010.

[5] Adam M Feist and Bernhard O Palsson. What do cells actually want? *Genome biology*, 17(1):110, 2016.

[6] Carlos Eduardo García Sánchez and Rodrigo Gonzalo Torres Sáez. Comparison and analysis of objective functions in flux balance analysis. *Biotechnology progress*, 30(5):985–991, 2014.

[7] Helena A Herrmann, Beth C Dyson, Lucy Vass, Giles N Johnson, and Jean-Marc Schwartz. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ systems biology and applications*, 5(1):1–8, 2019.

[8] R Mahadevan and CH Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264–276, 2003.

[9] Wout Megchelenbrink, Martijn Huynen, and Elena Marchiori. optgpsampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PloS one*, 9(2), 2014.

[10] Monica L Mo, Bernhard Ø Palsson, and Markus J Herrgård. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology*, 3(1):37, 2009.

[11] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3): 245–248, 2010.

[12] Jan Schellenberger and Bernhard Ø Palsson. Use of randomized sampling for analysis of metabolic networks. *Journal of biological chemistry*, 284(9):5457–5461, 2009.

[13] Sharon J Wiback, Iman Famili, Harvey J Greenberg, and Bernhard Ø Palsson. Monte carlo sampling can be used to determine the size and shape of the steady-state flux space. *Journal of theoretical biology*, 228(4):437–447, 2004.