dingo: a Python library for metabolic networks sampling & analysis

Apostolos Chalkis[1,2,3], Vissarion Fisikopoulos[1,3], Elias Tsigaridas[3,4], Haris Zafeiropoulos[5,6]

[1] Department of Informatics & Telecommunications, National & Kapodistrian University of Athens | [2] Athena Research Innovation Center, Greece | [3] GeomScale org
[4] Inria Paris and IMJ-PRG, Sorbonne Université | [5] Department of Biology, University of Crete | [6] Institute of Marine Biology Biotechnology and Aquaculture - Hellenic Centre for Marine Research

Author names are in alphabetical order

dingo on **GitHub**:

dingo **tutorial** GColab notebook:

GeomScale Org. group website:

BOSC 2021

Introduction

Metabolic networks and their corresponding models give a thorough description of the ***gene - protein - reaction*** associations from a single cell to the organism and the ecosystem level. At the species level, a **genome - scale metabolic model (GEM)**, enables detailed and quantitative predictions of the organism's behavior.

Metabolic networks analysis relies heavily on computational geometry methods and interweaves with constraint - based approaches. Commonly known methods such as **Flux Balance Analysis (FBA)** and **Flux Variability Analysis (FVA)** have been widely used to identify an optimal flux distribution by optimizing a linear objective function; e.g the biomass value of the organism and the extreme values that a flux may get correspondigly (Figure 1).

On the contrary, **flux sampling** is an unbiased method, as an objective function is not mandatory in this case. Sampling allows to mask all the possible flux values by estimating a probability distribution for the flux values of a given reaction (Figure 1). Yet, sampling is possible at optimal steady states, optimizing an objective function first.

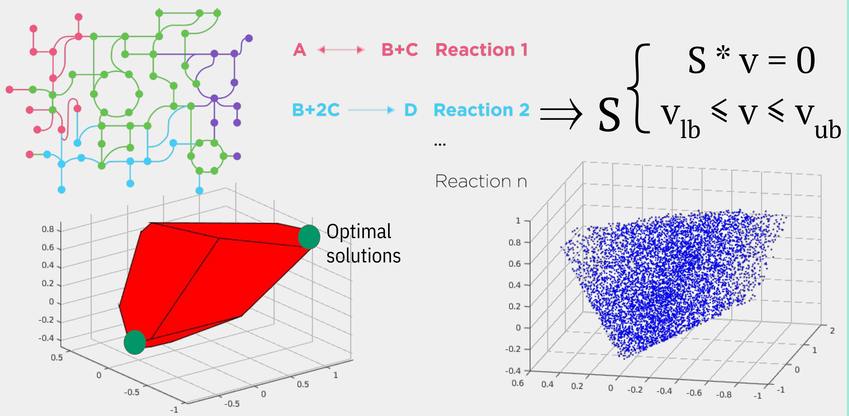

Figure 1: Given a genome of an organism, a genome - scale metabolic model (GEM) can be built from the chemical reactions found. The stoichiometric matrix (S) along with their corresponding constraints represent the flux space. Given the system is at steady state FBA returns optimal flux values while sampling all the possible flux values.

Sampling Methods

To sample metabolic polytopes we employ the **Multi-phase Monte Carlo Sampling (MMCS) algorithm from [1]** with the following key features:

- ➔ sampling procedure is mixed with rounding and split in phases for efficiency
- ➔ an optimized variant of Billiard Walk with faster arithmetic complexity per step is used
- ➔ the modular design of the method enables the use of any random walk

The dingo library

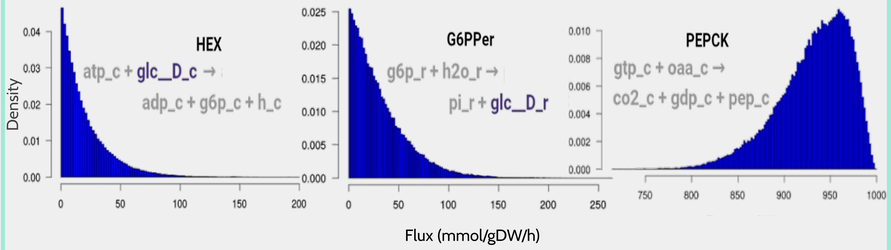

Figure 2: Flux distributions of reactions from the most recent human metabolic network Recon3D. We estimate the flux distributions of the reactions catalyzed by the enzymes Hexokinase (D-Glucose:ATP) (HEX), Glucose-6-Phosphate Phosphatase, Edoplasmic Reticular (G6PPer) and Phosphoenolpyruvate carboxykinase (GTP) (PEPCK). As we sample steady states, the production rate of glc__D_c should be equal to its consumption rate.

We introduce **dingo**, an efficient Python toolkit to perform some of the most fundamental analyses in metabolic networks. **dingo** supports both FBA and FVA and the random sampling on the flux space of metabolic networks of very high dimension. As shown in Figure 2, **dingo** supports sampling on the flux space of the latest human metabolic network reconstruction (Recon3D), including **13,543** metabolic reactions; **dingo** required <30 hours on a personal computer to do so. FBA and FVA implementations are also supported.

Based on the iAB-AMØ-1410 GEM of human alveolar macrophages and the SARS-CoV-2 virus biomass objective function (VBOF) they built, Renz *et al.* (2020) [2] investigated for alterations in the metabolism to come up with potential antiviral targets. **dingo** was used to analyse the metabolic model built from the macrophage model and the VBOF.

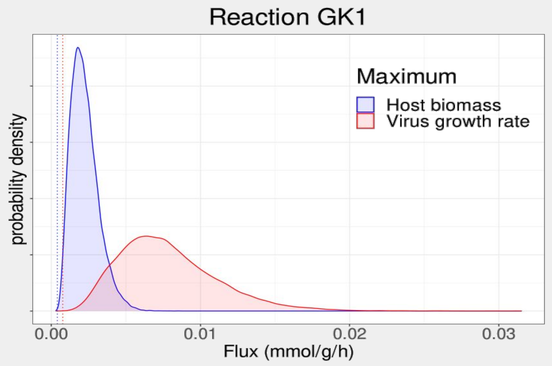

Figure 3: Flux sampling on the metabolic model built by the iAB-AMØ-1410 model and the VBOF by Renz et al. The flux distribution of GK1 reaction changes radically between the wild case (blue) and the infected by Sars-Cov-2 case (red).

Our findings where in line with those of Renz et al. (2020), indicating the **guanylate kinase (GK1)** as a potential antiviral target in all cases; both FBA and flux sampling (after maximizing the VBOF) (Figure 3) returned the same outcome. As shown in Figure 3, flux sampling provides insight on the distribution that a certain flux may take at the various steady states possible. Pairs of reactions can also be studied together to investigate the correlation between various reactions.

Open source science & future work

dingo is a part of Geomscale Org. a research and development project that delivers open source code for state-of-the-art algorithms at the intersection of data science, optimization, geometric, and statistical computing. **dingo** is based on the **volesti** package; a C++ library for volume approximation and sampling of convex bodies.

Parralelization of the flux sampling method is ongoing in an attempt to reduce even furtner the computational time needed. Methods for sampling the flux space of pairs of organisms, mostly between Bacteria, to infer metabolic interactions, is also under construction.

References

1. Chalkis, Apostolos, Vissarion Fisikopoulos, Elias Tsigaridas, and Haris Zafeiropoulos. "Geometric algorithms for sampling the flux space of metabolic networks," LIPIcs, Volume 189, SoCG 2021.
2. Renz, Alina, Lina Widerspick, and Andreas Dräger. "FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2." Bioinformatics 36.Supplement_2 (2020): i813-i821.
3. Chalkis, Apostolos, and Vissarion Fisikopoulos. "volesti: Volume approximation and sampling for convex polytopes in R." arXiv preprint arXiv:2007.01578 (2020).