Microbial synthesis of polyhydroxybutyrate in *Escherichia coli*: a systematic approach using thermodynamics and flux balance analysis studying different synthesis routes

Authors: Mariana I. Velasco Alvarez, Karel Olavarria Gamez, Mark van Loosdrecht and S. Aljoscha Wahl

Corresponding author: S. Aljoscha Wahl
Contact information: s.wahl@tudelft.nl

Delft University of Technology-Department of Biotechnology

Van der Maasweg 9

2629 HZ Delft

The Netherlands

General Introduction

This dataset contains data collected as part of the thesis project of Mariana-Itzel Velasco. It is open to the public as supplementary material for the thesis of Mariana-Itzel Velasco, in order for other researchers to follow up the results found.

The study and design of analysis was performed in the group of Cell Systems Engineering at the Faculty of Applied Sciences in TUDelft. This project was funded by NWO from The Netherlands and FAPESP from Brazil, which made possible the collaboration between Delft University of Technology and The University of São Paulo.

Purpose of the analysis

The aim of the study was to identify the most feasible metabolic pathway among three engineered strains through thermodynamics and Flux Balance Analysis. The theoretical results were confirmed by the experimental data performed in chemostats.

Test tool

The theoretical data was obtained through

1. Web-platform called Equilibrator (Noor E, Haraldsdóttir HS, Milo R, Fleming RMT. Consistent estimation of Gibbs energy using component contributions. PLoS Comput Biol. 2013;9: e1003098).
2. The COBRA toolbox was used for Flux Balance Analysis (Heirendt, L., Arreckx, S., Pfau, T. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc 14, 639–702 (2019). https://doi.org/10.1038/s41596-018-0098-2). The Ecolicore model was chosen to run the simulations, it can be downloaded from: http://systemsbiology.ucsd.edu/Downloads/EcoliCore (Orth J, Fleming R, Palsson B.

2010. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide, EcoSal Plus 2010; doi:10.1128/ecosalplus.10.2.1)

The experimental data was obtained in the fermentation laboratory through chemostat fermentations.

Description of the data in this dataset

A. Thermodynamic data that corresponds to Chapter 2 of the thesis.

A virtual environment was created in Python, the following steps have to be followed:

1. Several packages are needed to use equilibrator-API (`curl`, `pip`, Anaconda-jupyter notebook)
2. Open terminal and copy-paste the following link:

```
virtualenv -p python3 equilibrator
source equilibrator/bin/activate
pip install equilibrator-api jupyter
curl https://gitlab.com/elad.noor/equilibrator-
api/raw/develop/scripts/equilibrator_cmd.ipynb > equilibrator_cmd.ipynb
jupyter notebook
```

3. After the installation of `equilibrator-api` the python scripts can be ran.
   Three analyses were performed and are compiled in different files, the corresponding script for each analysis is as follows:
   a) Acetyl-CoA: `AcCoA_A1.ipynb`
   b) NADPH: `NADH_B1.ipynb`
   c) NADH: `NADPH_B2.ipynb`

B. Flux Balance Analysis predictions

The COBRA toolbox was ran using MATLAB 2017b to model the strain K12GAPNΔzwf+pPHB. The steps to follow to run the model are:

1. Install COBRA toolbox (`git` and `curl` are required solvers): https://opencobra.github.io/cobratoolbox/stable/installation.html
2. Open MATLAB and run function `initCobraToolbox`
3. Two simulations were performed:
   a. Prediction of different fluxes of ICL (isocitrate lyase), run the script:
      `Predictions_ICLfluxes.m`
      `Plot_ICLfluxes.m`
   b. Prediction of oxygen limited conditions:
      `Predictions_OxygenLimited.m`

C. Experimental Chemostat results

The experimental data collected corresponds to online data that is the monitoring of the fermentation parameters and offline data that corresponds to the samples taken during the experiment.
The raw data can be found in the .xlsx files, the data was further processed to estimate rates and balances. The files are divided in two different parameters that is aerobic and oxygen limited in the two strains studied.
The files are as follows:

a. For the strain K12+pPHB

   Rates calculations with the MATLAB script:

   `Chemostat_Ecoli_K12pPHB_qrates.m`

   Data reconciliation:

   `K12pPHB_dataReconc_AER01.m`
   `K12pPHB_dataReconc_AER005.m`
   `K12pPHB_dataReconc_OxyLim01.m`
   `K12pPHB_dataReconc_OxyLim005_1.m`
   `K12pPHB_dataReconc_OxyLim005_2.m`

b. For the strain K12GAPNΔzwf+pPHB

   Rates calculations with the MATLAB script:

   `K12GAPNzwfpPHB_AEROBIC_D01.m`
   `K12GAPNzwfpPHB_AEROBIC_D02.m`
   `K12GAPNzwfpPHB_AEROBIC_D005.m`
   `K12GAPNzwfpPHB_OXYGENLIMITED_D01_05.m`
   `K12GAPNzwfpPHB_OXYGENLIMITED_D01_10.m`
   `K12GAPNzwfpPHB_OXYGENLIMITED_D01_16.m`
   `K12GAPNzwfpPHB_OXYGENLIMITED_D01_26.m`

   Data reconciliation:

   `K12GAPNzwfpPHB_dataReconc_AER_D0.m`
   `K12GAPNzwfpPHB_dataReconc_AER_D02`
   `K12GAPNzwfpPHB_dataReconc_AER_D05.m`
   `K12GAPNzwfpPHB_dataReconc_OxyLim_D01_1.m`
   `K12GAPNzwfpPHB_dataReconc_OxyLim_D01_2.m`
   `K12GAPNzwfpPHB_dataReconc_OxyLim_D01_3.m`
   `K12GAPNzwfpPHB_dataReconc_OxyLim_D01_4.m`

D. Metabolite Analysis:

Heatmap figure displayed with the following MATLAB code:

`Heatmap_data.m`

The raw data can be found in the file: `MIC_GANP.xlsx`