Maximum entropy and population heterogeneity in continuous cell cultures, validation with experimental data

Jose A. Pereiro-Morejon, Jorge Fernandez-de-Cossio-Diaz, Roberto Mulet

March 16, 2020

1 Introduction

The study of cellular metabolism is a research area with direct potential impact in biological sciences, medicine and industry. An example is that cell culture-derived products are a major part of the multi-billion biotechnological industry portfolio. This products are obtained by exploding the capability of cellular metabolism to produce molecules with a wide range of chemical complexity. For this propose, cells are cultivated in three common modes: batch, fed-batch and continuous. In batch, cultivation starts with a medium rich in nutrients that will be consumed completely by the cells, often till starvation. Similarly, fed-batch cultures starts will a nutrient pool, but it is resupplied in discrete time intervals. On the other hand, in continuous mode, fresh medium constantly replaces culture fluid at a given rate.

The chemostat is an example of a continuous cultivation devise developed in the 50's (*cireRequired*6). Chemostat cultures are often run at constant working volume and in steady state condition, which is reach when the macroscopic variables of the culture stays constant with time (mainly cell and external metabolites concentrations). Although advantages of continuous cultivation have been commonly mentioned in the literature (*citeRequired*1), the preferred use of these technics over batch or fed-batch struggle with the complexity displayed by continuous systems, i.e., hysteresis, multi-stability or sharp transitions between metabolic states(*citeRequired*2). Mathematical modeling can guide our efforts to understand the cellular metabolism and then suggest strategies to improve production efficiency using continuous cultivation methods. This could lead to a decrease in the production costs and subsequently, to a reduction on medicine prices and a broader accessibility to treatments. Therefor, for the industry, it is fundamental the development of theoretical frameworks that allows, for example, predict how to reach an optimal stable cultivation state given a cell of interest and the medium to be used.

A major driver of biological discovery are currently found in increasingly accurate experimental techniques that generate unprecedented amount of thought-

ful data. Information about cellular metabolism, at individual reaction (gene) level, had lead to the development of genome-scale metabolic networks ($GEMs$) ($citeRequire3$). These networks have been used to build mathematical models of the cellular metabolism ($citeRequired4$). Constraint-based technics as Flux balance Analysis (FBA) had have good results predicting, for example, cellular metabolism in growth phase of batch cultures ($citeRequired5$). Meanwhile, a methodology for applying (FBA) in the context of a chemostat continuous cultivation have been proposed by (Fernandez-de Cossio-Diaz et al., 2017). It introduces a detailed characterization of the steady state of the chemostat, which is achieved by coupling cell metabolism with the dynamics of extra-cellular concentrations. Although this approach can explain different phenomena in the context of a detailed metabolic model, it have the drawback that it assume cell population to be homogeneous.

Cells in a culture are not identical, many different processes can contribute to heterogeneity ($citeRequired7$). In this context, with the goal of understanding which is the impact of considering cell population in a continuous culture to be homogeneous, a model that include cell heterogeneity is develop by (Fernandez-de Cossio-Diaz & Mulet, 2019). It is proposed to apply the maximum entropy principle ($MaxEnt$) ($citeRequired8$) to model the heterogeneity in a continuous bioreactor. The authors also shows that the methodology is applicable to GEMs of realistic sizes with commonly available computational resources.

The objective of this work is to search thorough literature-available data of chemostat cultures and test the impact of including heterogeneity. For that propose we will face the two model proposed in (Fernandez-de Cossio-Diaz et al., 2017) and (Fernandez-de Cossio-Diaz & Mulet, 2019) to the same data and using the same metabolic networks for different cell lines.

2 Materials and Method

2.1 Model framework

The model framework used in this work is detaily explained in (Fernandez-de Cossio-Diaz et al., 2017) and (Fernandez-de Cossio-Diaz & Mulet, 2019).

2.2 Chemostat experimental data

2.2.1 Escherichia coli KJ134

Continuous cultivation data for Escherichia coli ($E.coli$) were taken from (van Heerden & Nicol, 2013). The used Escherichia coli KJ134 strain was genetically modified for succinic acid fermentation. A dozen of steady states were recorded at different dilution rates and GLC availability in the feed medium. Data for the effluent concentration of GLC, succinic acid (SA), acetic acid (AcA), formic acid (FA) and malic acid (MA) for the different steady states were given. Also, the cell concentration was reported.

2.2.2 Human cell line AGE1.HN.AA1

Experimental data were taken from (Rath, 2017). In this work, the author performed 6 continuous cultures of the cell line AGE1.HN.AA1. Its parental line AGE1.HN was established by the company ProBioGen (ProBioGen AG, Berlin, Germany) from a tissue sample of a human brain. The experiments were run under various conditions, differing mainly in the dilution rate (D) and the feed medium composition of glucose (GLC), glutamine (GLN) and galactose (GAL) $Table$ 2.

For each experiment, a steady-state condition was reached, (steady states were labeled A, B, C, D, E, $F01$), and several observables were reported. Particularly relevant for this work was the growth rate (μ), D, the viable cell concentration (Xv) and the medium concentration (s) and derived uptake rate (q) for a set of metabolites (GLC, lactose (LAC), GLN, ammonium ($NH4$), GAL, pyruvate (PYR), glutamate (GLU), alanine (ALA), asparagine (ASP)). A unit conversion was required to make experimental data and our model compatible. For this propose the only external data needed was the cell mass density, 0.25 pgDW/ μm^3 (Niklas et al., 2011).

2.3 Preparing GEMs

2.3.1 Escherichia coli KJ134

For modeling the cellular metabolism, the $E.coli$ genome-scale metabolic model (GEM) iJR904 (Reed et al., 2003) (download link: $https://darwin.di.uminho.pt/models$) was modified to match with (van Heerden & Nicol, 2013) data. The reported genetic modifications for the Escherichia coli KJ13 strain were included. It was also added enzymatic cost constraints in concordance with (Beg et al., 2007).

2.3.2 Human cell line AGE1.HN.AA1

With the same propose of modeling, the metabolism of the human-derived cell AGE1.HN.AA1, it was used a GEM from (Shlomi et al., 2011)(download link: $https://www.ebi.ac.uk/biomodels/MODEL1105100000$). The GEM biomass equation was modified in agreement with the biomass composition for AGE1.HN.AA1 reported in (Niklas et al., 2013). The production of α1-antitrypsin ($A1AT$) was also considered, it was used the protein sequence, $https://www.drugbank.ca/polypeptides/P01009$, and the reported production rate for $A1AT$ (Rath, 2017) to set maintenance for the required amino acids. Additionally, an atp demand ($ATPM$) was included, taking the maintenance energy demand for mammalian cells mentioned by (Fernandez-de Cossio-Diaz & Vazquez, 2018).

3 Results and Discussion

The main objective of this work is to explore the capability of an application of the Maximum Entropy ($MaxEnt$) framework to the modeling of cellular metabolism in a continuous cultivations regime. $MaxEnt$ has been used for the analysis of other biology-related problems with success. It has become a useful tool when we are dealing with limited data, which actually is a common scenario in biology (De Martino & De Martino, 2018).

In particular, the framework proposed in (Fernandez-de Cossio-Diaz et al., 2017) and (Fernandez-de Cossio-Diaz & Mulet, 2019) allows to link effectively the macroscopic variables, commonly accessible, that defined the state of a chemostat steady-state with the underlying cellular metabolism of the cells. In the former article, a more traditional approach to model the metabolism was taken. It delegates in Flux Balance Analysis (FBA) (Orth et al., 2010) for choosing the vector of reactions fluxes assumed to be determining the current metabolic state of the cell under the cultivation conditions. This method has been heavily used in the past decades with good results and a variety of applications (O'Brien et al., 2015). It has the advantage of not requiring kinetic parameters from cellular metabolism. This is possible because FBA is based in a steady-state assumption, justified in the time scale difference between regulatory (slow) and metabolic (fast) processes (De Martino & De Martino, 2018). Another assumption FBA makes is that the cell population in the culture is homogeneous and is optimizing a given metabolic objective. FBA returns the vector of all reactions fluxes that optimize the objective function subject to the applied constraints (Orth et al., 2010). This vector will be taken as the definition of the metabolic state for all the cells in the culture, and with it, all the predictions or analyses will be made (Fernandez-de Cossio-Diaz et al., 2017).

To overcome the limitation of considering culture homogeneity, a probability distribution can be defined over the set of all the possible metabolic states. This distribution describes how probable is to find a cell in one metabolic state. To infer such probability distribution in agreement with available experimental data, the $MaxEnt$ principle can be used (Fernandez-de Cossio-Diaz & Mulet, 2019). In this work $MaxEnt$ was applied in such a way that it returns the probability distribution that maximized the entropy and ensures the expected value of the growth rate to match with the experimentally observed. Now, instead of a vector of reactions fluxes that optimize an objective function, the model will return a vector containing the probability distribution for each reaction flux due to the inferred $MaxEnt$ distribution. This is a major advantage with respect to the FBA framework, $MaxEnt$ not assume that the cells "*have a goal*", an objective function, it claims to compute the bias-less probability distribution in concordance with the imposed, data-driven, constraints (De Martino & De Martino, 2018). On the other hand, the main practical problem of using the metabolic $MaxEnt$ model is that it can represent a computationally expensive task, but this problem was also addressed by (Fernandez-de Cossio-Diaz & Mulet, 2019) using expectation propagation algorithms.

3.0.1 Escherichia coli KJ134

Because good predictive results had been made using *E coli* $GEMs$ (Vazquez & Oltvai, 2016), we include this model organism in our study. $Figures\ 1-2$ shows the results of both, FBA and $MaxEnt$ models, facing the experimental data reported in (van Heerden & Nicol, 2013) for *Escherichia coli* $KJ134$ continuous cultures. In the first figure, plots of several chemostat observables vs the cell-specific dilution rate (ξ) are shown. The parameter ξ is computed as Xv/D and can be interpreted as the number of cells sustained in the culture per unit of medium supplied per unit time (Fernandez-de Cossio-Diaz et al., 2017). In the second figure, the corresponding correlations are displayed.

As explained in the last section, the inferred $MaxEnt$ distribution ensures that the predicted growth rate (μ), more precisely its expected value, coincides with the observed value, as evidenced in the upper-right graphs $Figures\ 1-2$. This extra piece of information added to the model allow us to have much better correlations with respect to the FBA results, $Figure\ 2$. In the cases of GLC and AcA, two important metabolites related to the energetic metabolism, the improvements are notable. The FA correlations for the $MaxEnt$ model wasn't good, but in contrast, FBA is completely blind to it. Also, it is possible to observe in the graphs of μ and cell concentration (Xv) that FBA really over-determinate these quantities. This fact may suggest, among others, that the imposed constraints are too permissive. Also, although in the case of *E. coli*, the hypothesis of biomass maximization made by FBA is generally accepted, it could not be valid for all situations. As explained before, $MaxEnt$ do not make this assumption, it just ajust the model to available experimental data.

3.0.2 Human cell line AGE1.HN.AA1

A more difficult task is to model the human metabolism. Human related $GEMs$ are not as mature as bacteria networks (Gu et al., 2019). Anyway, we applied the methodology to experimental data extracted from (Rath, 2017). $Figures\ 3$ shows the plots of several exchange fluxes as a function of the parameter ξ. If one sees the data correlations, $Figure\ 4$, this time the results are not as polarized as in the *E. coli* section but, generally $MaxEnt$ model (right columns) predictions are closer to the experimental data than ones produced by FBA model (left columns). Remember again that, $MaxEnt$ distribution ensures that the model growth rate match with experimental results (second row, third pair column $Figure\ 3-4$). Remarkable are the cases of GLN and PYR, where $MaxEnt$ really make a good job improving the predictions of FBA. Less precise, but still better for $MaxEnt$, are the cases of ALA and LAC. A curious case is $NH4$, where FBA results were more consistent with respect to $MaxEnt$. Again, FBA ends up over-estimating the values of μ and Xv, which may suggest that the cell is not actually in a state that maximizes the biomass production rate. For the rest of the results, GLC, GLU, GAL and ASP, both model shows similar predictions, only satisfactory for GLC. Because exchange fluxes and the concentration of metabolites in the effluent are two related quantities, results

showed in *Figure* $5-6$ mirror the former ones.

It is notable that, although both models are not using a cell-line context-specific network, which is in our opinion the most significant source of error, they were capable of capturing many of the metabolic features encode in the experimental data.

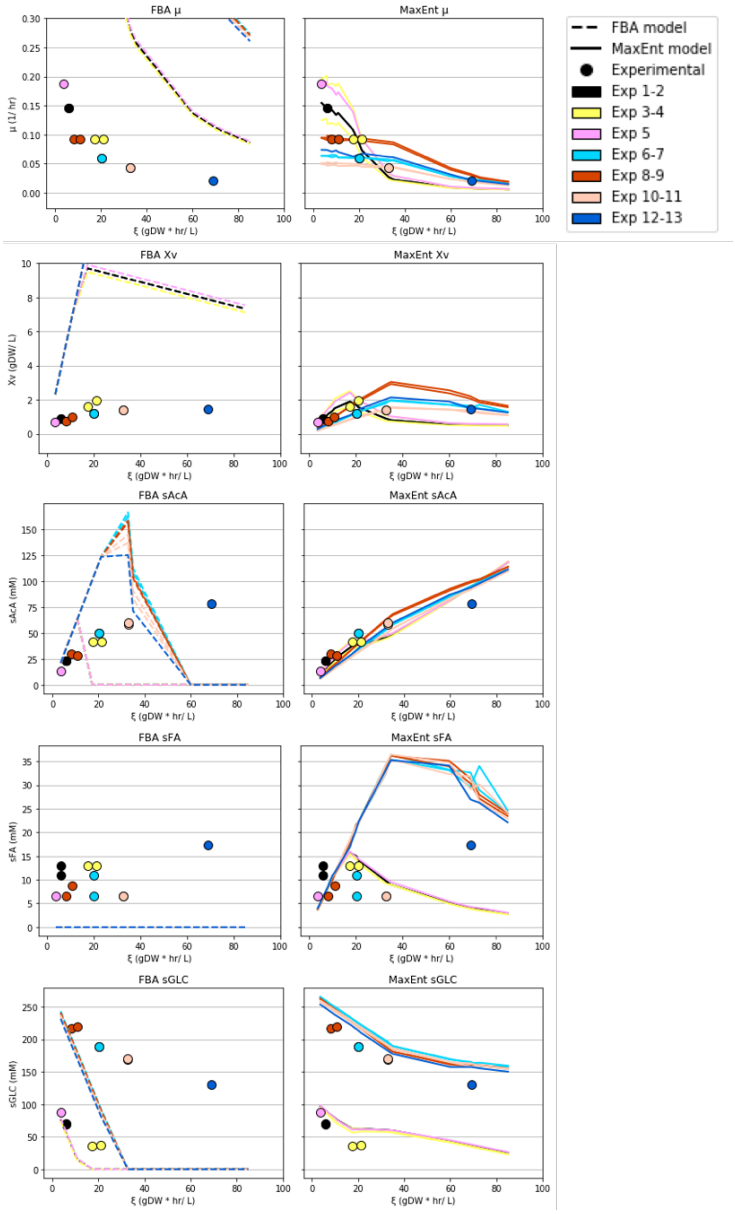

Figure 1: Predictions (lines) of both, FBA (left column) and $MaxEnt$ (right column) models, facing the experimental data (colored circles) reported in (van Heerden & Nicol, 2013) using *Escherichia coli* $KJ134$. Rows shows the growth rate (μ), the cell concentration (Xv), the effluent concentration (s) of acetic acid (AcA), formic acid (FA) and glucose (GLC) respectively as a function of the cell-specific dilution rate (ξ) . $MaxEnt$ distribution was chosen in a manner that the expected value of the growth rate coincide with the experimental data (see first row right column)

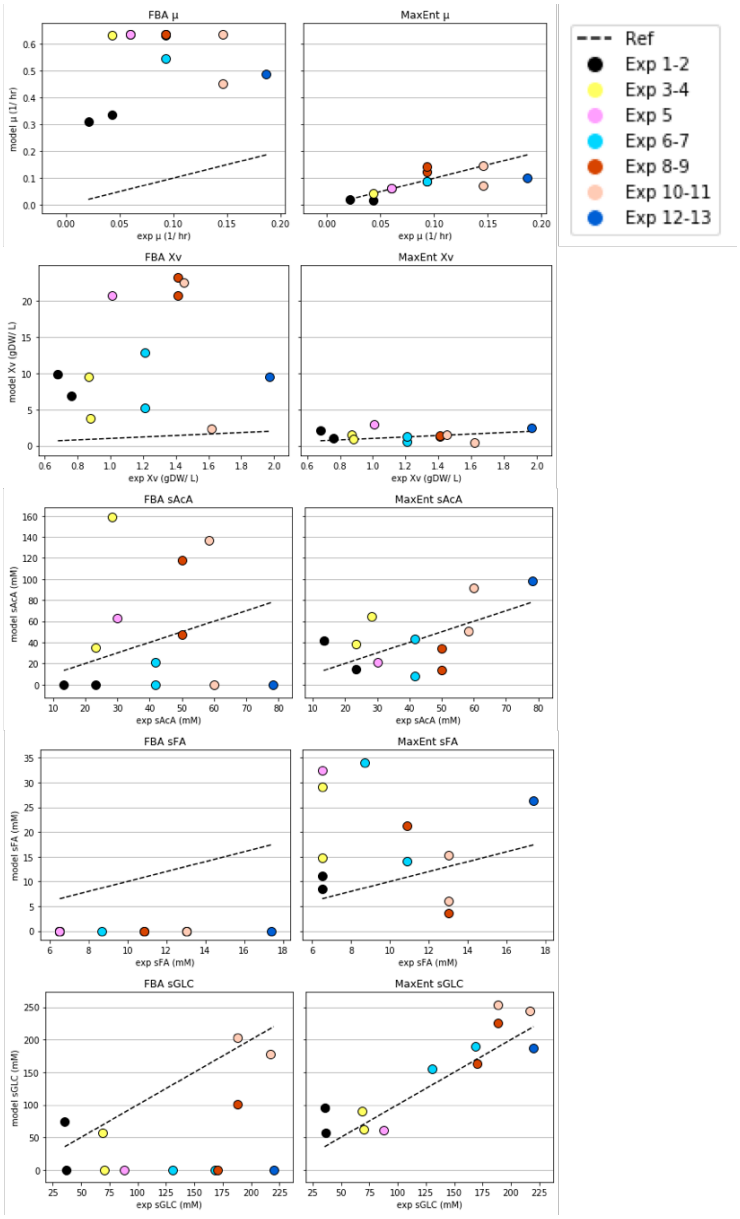

Figure 2: Correlations of the results of both models (y axis), FBA (left column) and $MaxEnt$ (right column), facing the experimental data (x axis) reported in (van Heerden & Nicol, 2013) using *Escherichia coli* $KJ134$. Rows shows the growth rate (μ), the cell concentration (Xv), the effluent concentration (s) of acetic acid (AcA), formic acid (FA) and glucose (GLC) correlations respectively. $MaxEnt$ distribution was chosen in a manner that the expected value of the growth rate coincide with the experimental data (see first row right column)

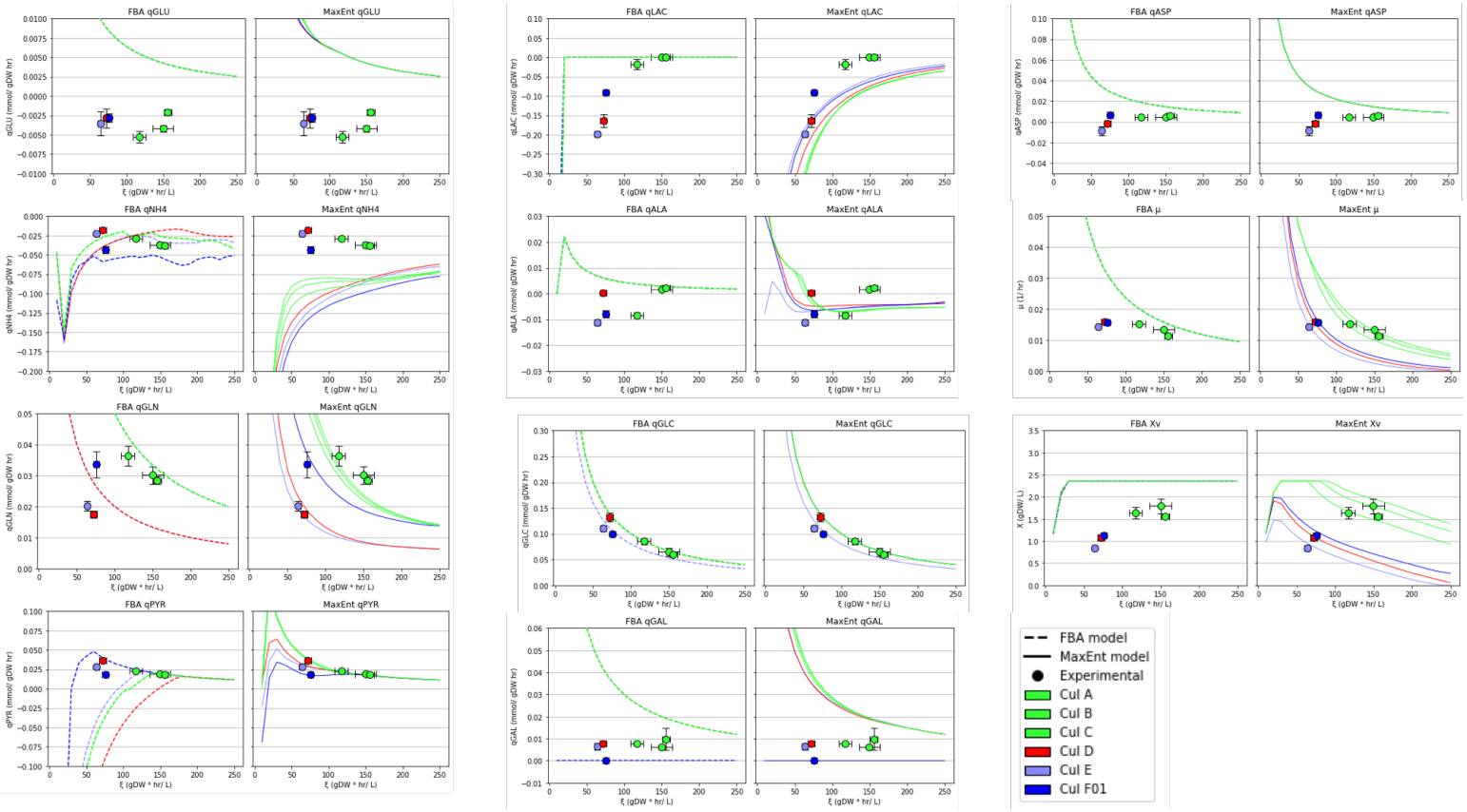

Figure 3: Exchange fluxes (q), growth rate (μ) and cell concentration (Xv) as a function of the cell-specific dilution rate (ξ). Graphs are arranged as three paired columns. Each pair have the models prediction (lines), FBA (left column) and $MaxEnt$ (right column), and experimental data (colored circles) from (Rath, 2017) for continuous cultures of the human derived cell line $AGE1.HN.AA1$. $MaxEnt$ distribution was chosen in a manner that the expected value of the growth rate coincide with the experimental data (see third paired colum, second row)

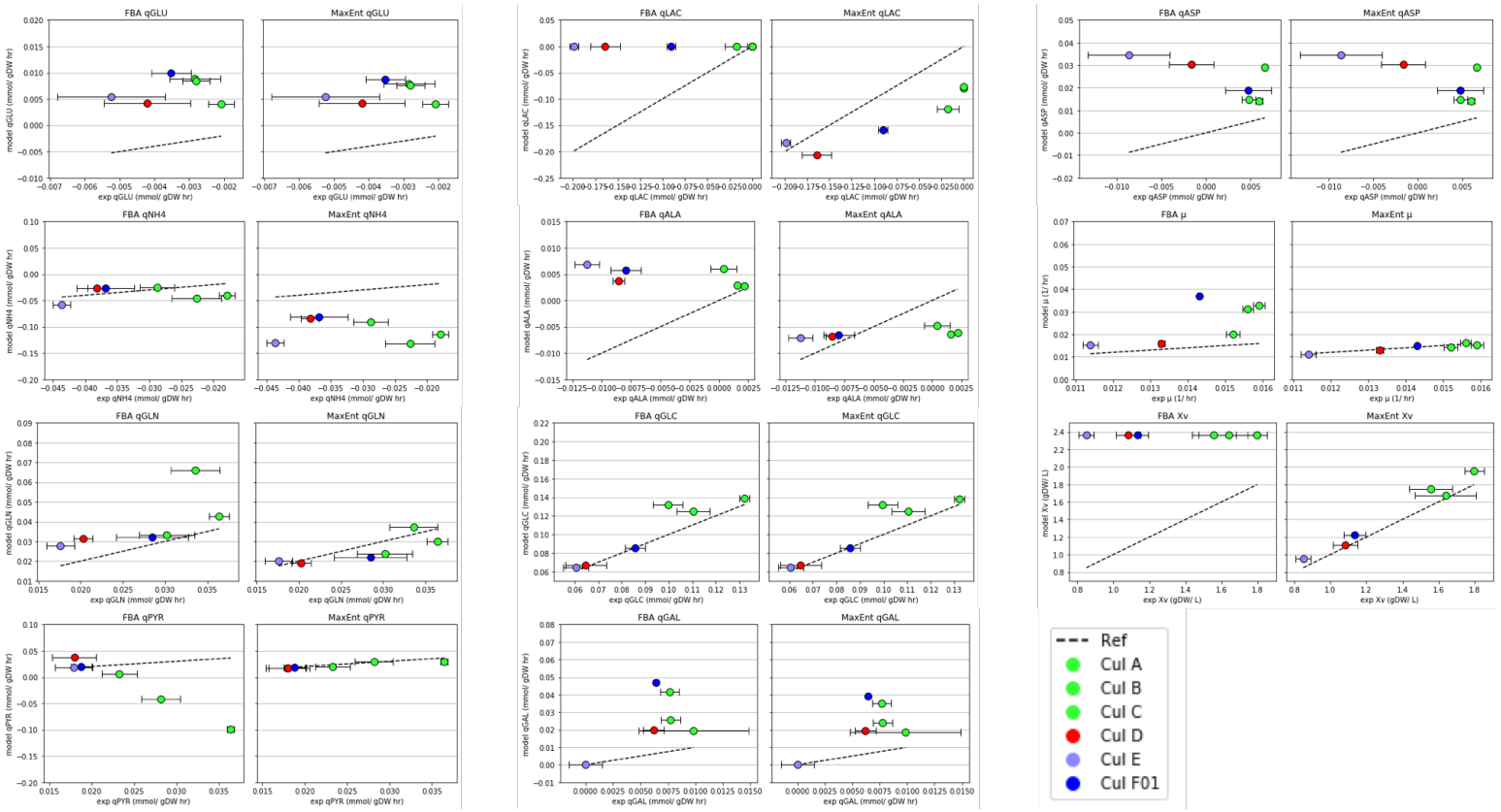

Figure 4: Correlations of the predicted (y axis) exchange fluxes (q), growth rate (μ) and cell concentration (Xv) against experimental data (x axis) from (Rath, 2017) for continuous cultures of the human derived cell line $AGE1.HN.AA1$. Graphs are arranged as three paired columns. Each pair have the correlations using FBA (left column) and $MaxEnt$ (right column) models. $MaxEnt$ distribution was chosen in a manner that the expected value of the growth rate coincide with the experimental data (see third paired colum, second row)

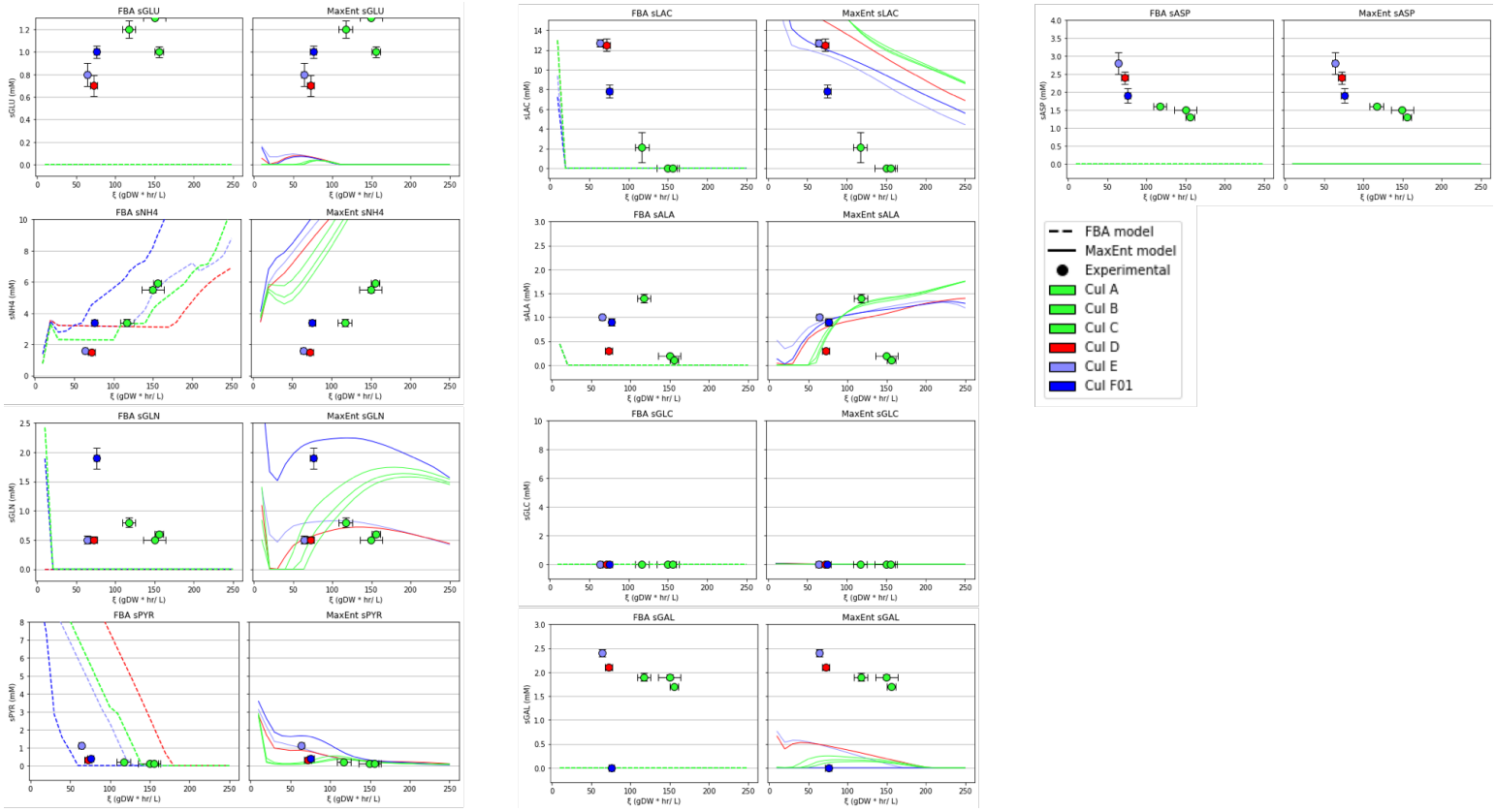

Figure 5: Metabolites effluent concentration (s) as a function of the cell-specific dilution rate (ξ). Graphs are arranged as three paired columns. Each pair have the models prediction (lines), FBA (left column) and $MaxEnt$ (right column), and experimental data (colored circles) from (Rath, 2017) for continuous cultures of the human derived cell line $AGE1.HN.AA1$

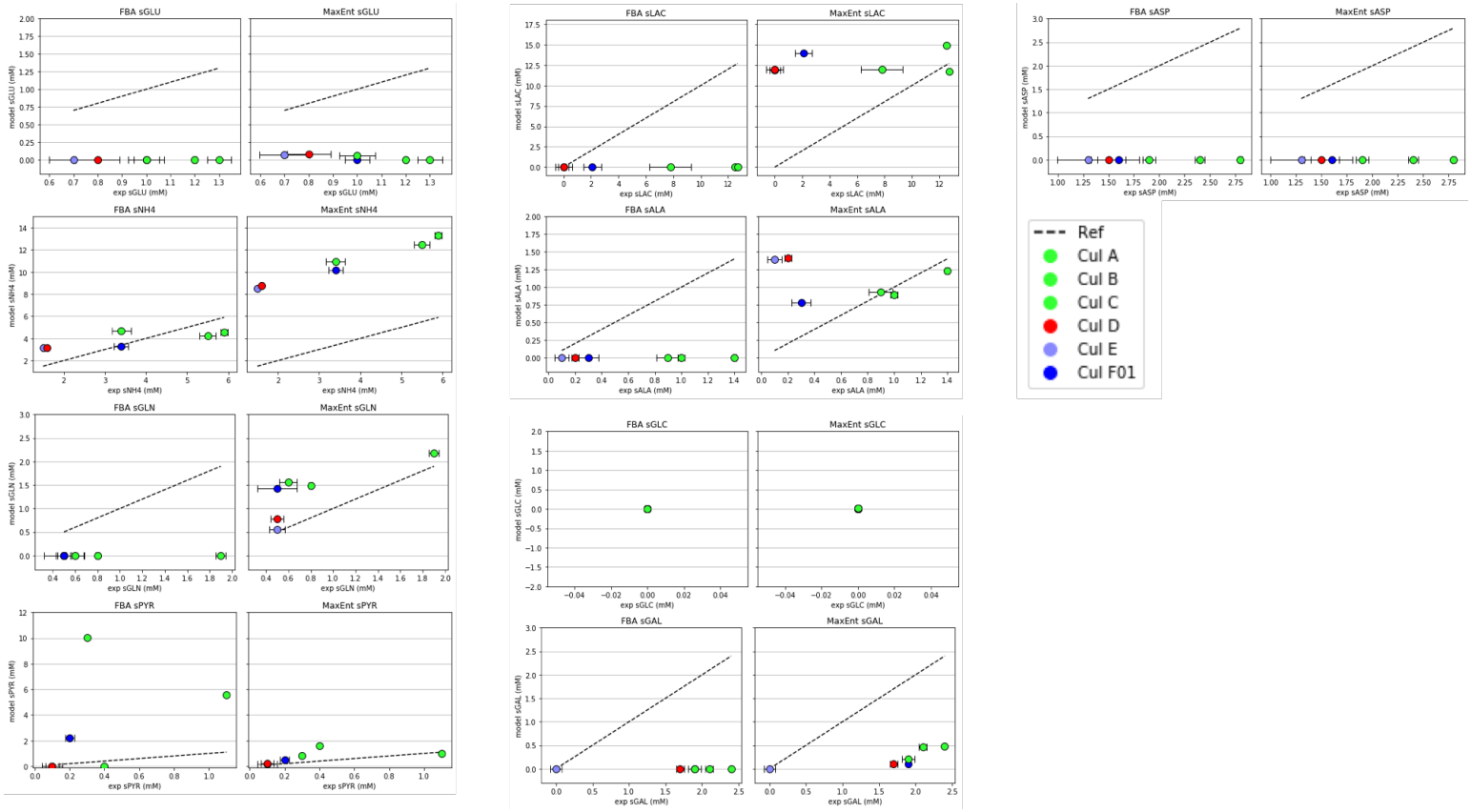

Figure 6: Correlations of the predicted (y axis) metabolites effluent concentration (s) against experimental data (x axis) from (Rath, 2017) for continuous cultures of the human derived cell line $AGE1.HN.AA1$. Graphs are arranged as three paired columns. Each pair have the correlations using FBA (left column) and $MaxEnt$ (right column) models.

4 References

Beg, Q. K., Vazquez, A., Ernst, J., Menezes, M. A. d., Bar-Joseph, Z., Barabási, A.-L., & Oltvai, Z. N. (2007, July). Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, *104*(31), 12663–12668. Retrieved 2019-11-06, from `https://www.pnas.org/content/104/31/12663` doi: 10.1073/pnas.0609845104

De Martino, A., & De Martino, D. (2018, April). An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, *4*(4), e00596. Retrieved 2019-10-25, from `https://linkinghub.elsevier.com/retrieve/pii/S2405844018301695` doi: 10.1016/j.heliyon.2018.e00596

Fernandez-de Cossio-Diaz, J., Leon, K., & Mulet, R. (2017). Characterizing steady states of genome-scale metabolic networks in continuous cell cultures. *PLoS Computational Biology*, *13*(11), 1–22. doi: 10.1371/journal.pcbi.1005835

Fernandez-de Cossio-Diaz, J., & Mulet, R. (2019, February). Maximum entropy and population heterogeneity in continuous cell cultures. *PLOS Computational Biology*, *15*(2), e1006823. Retrieved 2019-10-31, from `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006823` doi: 10.1371/journal.pcbi.1006823

Fernandez-de Cossio-Diaz, J., & Vazquez, A. (2018, May). A physical model of cell metabolism. *Scientific Reports*, *8*(1), 1–13. Retrieved 2019-12-09, from `https://www.nature.com/articles/s41598-018-26724-7` doi: 10.1038/s41598-018-26724-7

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology*, *20*(1), 1–18. doi: 10.1186/s13059-019-1730-3

Niklas, J., Priesnitz, C., Rose, T., Sandig, V., & Heinzle, E. (2013). Metabolism and metabolic burden by a 1 -antitrypsin production. *Metabolic Engineering*, *16*, 103–114. Retrieved from `http://dx.doi.org/10.1016/j.ymben.2013.01.002` doi: 10.1016/j.ymben.2013.01.002

Niklas, J., Schräder, E., Sandig, V., Noll, T., & Heinzle, E. (2011). Quantitative characterization of metabolism and metabolic shifts during growth of the new human cell line AGE1.HN using time resolved metabolic flux analysis. *Bioprocess and Biosystems Engineering*, *34*(5), 533–545. doi: 10.1007/s00449-010-0502-y

O'Brien, E. J., Monk, J. M., & Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell*, *161*(5), 971–987. Retrieved from http://dx.doi.org/10.1016/j.cell.2015.05.019 doi: 10.1016/j.cell.2015.05.019

Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. Retrieved from http://dx.doi.org/10.1038/nbt.1614 doi: 10.1038/nbt.1614

Rath, A. (2017). *Characterisation of cell growth, metabolism and recombinant protein production during transient and steady state conditions for the human cell line AGE1.HN-AAT* (PhD Thesis).

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003, August). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*(9), R54. Retrieved 2019-12-13, from https://doi.org/10.1186/gb-2003-4-9-r54 doi: 10.1186/gb-2003-4-9-r54

Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., & Ruppin, E. (2011, March). Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect. *PLOS Computational Biology*, *7*(3), e1002018. Retrieved 2019-11-05, from https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002018 doi: 10.1371/journal.pcbi.1002018

van Heerden, C. D., & Nicol, W. (2013, September). Continuous and batch cultures of Escherichia coli KJ134 for succinic acid fermentation: metabolic flux distributions and production characteristics. *Microbial Cell Factories*, *12*(1), 80. Retrieved from https://doi.org/10.1186/1475-2859-12-80 doi: 10.1186/1475-2859-12-80

Vazquez, A., & Oltvai, Z. N. (2016, August). Macromolecular crowding explains overflow metabolism in cells. *Scientific Reports*, *6*, 31007. Retrieved 2019-11-04, from https://www.nature.com/articles/srep31007 doi: 10.1038/srep31007