Benchmarking different types of Escherichia coli Scale Metabolic Models on oxic and anoxic Datasets

By
Vanessa Carta

Laboratory of Systems & Synthetic Biology

29-04-2024

Supervisor 1: Claudia de Buck
Supervisor 2: Maria Suarez Diez

Abstract

Genome-scale models are powerful computational tools widely employed to guide metabolic engineering approaches, especially in the design of optimised organism for industrial applications. However, the limited predictive accuracy represents the primary limitation of GEMs. In recent years, several studies were focused on the improvement of GEMs leading to the construction of extended-GEMs, improved metabolic models that accounts for additional factors, such as the enzyme availability and the protein cost allocation. The design of extended GEMs represents a remarkable improvement in the field of metabolic modelling. Unfortunately, extended GEMs are typically validated on different datasets, mainly containing oxic data, making challenging the comparison of their performance. This project is focus on the benchmarking of three metabolic models of *Escherichia coli.* The main interest was in the evaluation of predictive accuracy of the two extended models eciML1515 and pciML1515 compared to the conventional GEM iML1515, especially under anoxic conditions. The benchmarking did not provide exhaustive conclusions for all the considered conditions, however, pinpoint eciML1515 as the most accurate model amount the evaluated models.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's regulations and Code of Practice for Research Degree Programmes, and that it has not been submitted for any other academic awards. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in this dissertation are those of the author.

Signed:

Date:

List of Tables

List of Figures

1. Introduction

1.1 Constraint-based modelling

Constrained-based modelling is a computational approach widely used to investigate metabolic networks. It performs the simulation of metabolism based on the stoichiometry and reversibility of metabolic reactions. One of the most relevant applications of constraint-based modelling is represented by Genomic-scale metabolic models (GEMs), consisting of a genome wide reconstruction of a metabolic network (Sánchez et al., 2017).

The main advantage of GEMs is represented by the possibility to easily simulate the organism's metabolism under different conditions. In fact, in GEMs the genotype and metabolic phenotype are linked through the gene-protein-reaction (GPR) relationship. This allows to directly connect each metabolic reaction to its catalysing enzyme and corresponding gene. This feature is important to simulate metabolic phenotypes when gene deletions or heterologous pathways are introduced. Thus, the GPR makes GEMs a powerful computational tool to guide metabolic engineering approaches, especially driven to the optimization of microorganism strains used for industrial applications. Indeed, GEMs are widely used to identify optimal synthetic pathways to produce valuable molecules, predict growth phenotypes across different conditions and carbon sources, and locate potential targets for metabolic engineering (Moreno-Paz et al., 2022; Yang et al., 2021).

In GEMs, metabolism is described as a network of reactions stored in a stoichiometric matrix, with columns that define the stoichiometry of each reaction and rows that indicate the mass balance of each metabolite involved. Each reaction is associated with a metabolic flux that can assume a possible range of values according to the constrains applied to the model. GEMs are based on two main constraints, namely the stoichiometry and the boundaries associated with each reaction. The boundaries limit the values' range associated with a metabolic flux. Usually, bounds are set to define the reversibility of each reaction,

but they can be fixed to specific values to further constrain the model. For instance, GEMs are commonly constrained in the carbon source uptake. These metabolic models work under the steady state assumption for the concentration of intracellular metabolites, excluding intracellular accumulation (Bekiaris & Klamt, 2020; Grigaitis et al., 2021).

Flux Balance Analysis (Orth et al., 2010) andFlux Variability Analysis (Müller & Bockmayr, 2013) are the two most used approaches applied to GEMs. Both are based on linear programming, and they return a flux distribution as the solution of the optimization process of a specific objective function. Diverse reactions can be selected as the objective function such as biomass, carbon source, ATP or a specific

metabolite (Grigaitis et al., 2021). FBA and FVA present relevant drawbacks that reduce the information achievable from the model.

The main limitation of FBA consists in its inability to describe properly the metabolism's variability caused by the presence of redundant or alternative pathways in the cell. In fact, FBA returns only a single set of metabolic fluxes across multiple feasible flux distributions compatible with the objective optimisation, causing the loss of important information. Differently, FVA supplies an interval of possible flux values compatible with the optimised objective without specifying their feasibility or probability with which they could occur (Herrmann et al., 2019). Flux sampling methods can be combined to FBA to investigate the whole solution space. The sampling analyses are conducted by specific algorithms such as the Artificially centred Hit-and-Run (Saa & Nielsen, 2016), which generate a sequence of feasible solutions associated with their probability distribution. Furthermore, FBA and FVA are based on the steady state assumption for intracellular concentrations, excluding intracellular accumulation and variation of the extracellular conditions (Herrmann et al., 2019). Unfortunately, this assumption does not represent a realistic scenario for all growth conditions, such as batch growth, where the concentration of extracellular metabolites changes dynamically over the growth process. An alternative approach is represented by dynamic FBA (dFBA), which integrates the conventional FBA with differential equations that describe the dynamic variation of extracellular concentrations over time. This methodology allows to represent a more realistic scenario for batch systems, in which the conditions change over the growth process (Mahadevan et al., 2002).

Despite the considerable role of GEMs in metabolism research, previous studies suggested a limited accuracy of GEMs' predictions, highlighting the inconsistency between predicted fluxes and experimental data. A clear example of this is represented by the inability of conventional GEMs to correctly simulate the overflow metabolism of *Escherichia coli* (Grigaitis et al., 2021) and the Crabtree effect in *Saccharomyces cerevisiae* (Sánchez et al., 2017). The low accuracy suggests an oversimplification of the metabolic network in GEMs, and lack of important additional factors that impact deeply the metabolism behaviour, such as physical limitations, regulation, reactions' kinetics, enzymatic availability and protein cost allocation (Grigaitis et al., 2021). To increase the accuracy, extended versions of GEMs have been developed, which include additional constraints as the cell volume limitation (Beg et al., 2007), enzyme availability (Mao et al., 2022; Sánchez et al., 2017), protein source allocation(Grigaitis et al., 2021) and thermodynamics (Yang et al., 2021).

In this project, the attention is directed mainly toward enzyme-constrained and protein sources allocation models. Enzyme-constrained models differ mainly in the formulation of the enzyme constraint and complexity. The Enzymatic Constraints using Kinetic and Omics data (GECKO) models are constructed to incorporate the enzyme availability as constraint to the metabolic fluxes, combining protein abundance and kinetics parameters (Sánchez et al., 2017). Subsequently, the protein-constrained models were developed combining the enzyme availability, based on the GECKO concept, and the distribution of sources required for the synthesis, folding and degradation of protein. This approach consents to explicitly

account for sources allocation between the different cellular processes, adding an extra constraint layer, which increases the prediction accuracy compared to GEKCO (Grigaitis et al., 2021). An alternative approach has been proposed by Mao et al., 2022, which has developed a workflow that converts GEMs into enzyme-constrained models by adding a single enzyme constraint defined from the experimental measure of the total protein abundance and kinetic parameters.

The introduction of extended-GEMs represents a remarkable advancement in the field of constrained-base modelling that has permitted to achieve a more accurate description of metabolic processes, as shown by several published studies that state how enzyme-constrained models significantly outperform their GEM counterparts (Grigaitis et al., 2021; Mao et al., 2022; Moreno-Paz et al., 2022; Sánchez et al., 2017; Yang et al., 2021). However, the assessment of the predictive performance across alternative extended-GEMs remains unexplored. Indeed, all the above-mentioned models have been extensively validated using different datasets mostly containing experimental data referred to aerobic growth, rendering ineffective the comparison of their predictive accuracy especially under anoxic conditions.

During this thesis project, three genome-scale models of *E. coli* are compared to evaluate their predictive accuracy. The selected models are the conventional GEM iML1515 (Monk et al., 2017), the EMPy enzyme-constrained eciML1515 model (Mao et al., 2022) and the protein allocation model pcIML1515 (Grigaitis et al., 2021).The Gecko model is excluded from the analysis in view of the results of Mao et al., 2022,attesting whose framework the protein constraint model was built (Grigaitis et al., 2021).For the benchmarking process three distinct datasets were selected, specifically chosen to include data collected from diverse experimental conditions, focusing on anoxic settings. The identification of a metabolic model that is highly accurate in the anaerobic simulations would represent a valid tool in the context of the ICEMAN project, in which this thesis is located. The ICEMAN project: Increasing Conservation of Energy by Microorganisms, conducted by the PhD candidate Claudia de Buck, strives to devise an innovative energy-conservation strategy in the model organism *E. coli*.

Thus, the final aim of this thesis consists in pinpointing the most accurate extended model, especially under anoxic conditions, between pciML1515 and eciML151 compared to the conventional GEM iML1515. The analyses will consider not only the predictive differences under oxic and anoxic conditions but will also evaluate their ability to properly simulate different growth processes such as chemostat and batch.

1.2 Metabolic Models

1.2.1 Standard GEM iML1515

The iML1515 model, published by Monk et al., 2017, represents the most up-to-date constrained-based model representation of the *E. coli* K-12 MG1655 metabolism. It consists of a stoichiometric matrix with *m* rows and *n* columns, where the rows represent the metabolites and the columns the reactions included in the model. Each reaction is constrained by an upper and lower bound, which define the range of possible flux's values. The upper and lower bounds of reversible reactions are set to 1000 mmol gDW^{-1} h^{-1}, while in case of irreversible reactions, the lower bound is fixed to zero (Monk et al., 2017). The first version, originally featuring 1515 coded enzymes, 2719 reactions and 1192 metabolites, underwent a continuous refinement up to the latest version available in BiGG Models database (King et al., 2016) containing a total of 1516 genes, 2712 reactions and 1877 metabolites. The exchange reactions, which describe the metabolites' exchange between the intracellular and extracellular compartments, are described as reversible reactions, as indicated in Table 1. Positive exchange reaction fluxes represent the metabolites' flow towards the extracellular compartment, indicating the synthesis of metabolism's products, whereas negative fluxes suggest the cellular metabolites' uptake, indicating the substrates' consumption.

1.2.2 Enzyme-constrained Model eciML1515

The eciML1515 is an enzyme-constrained metabolic model built using the *ECMpy* workflow, developed by Mao et al., 2022. The model is constructed from iML1515 and combines the GEM's stoichiometric matrix with a single enzyme constraint, which takes into account the total enzyme availability. The stoichiometric matrix, defined as described for iML1515, includes the whole set of reactions constrained by their bounds. In eciML1515, each reversible reaction is split into two irreversible ones in light of the different kinetics parameters (Kcat) in both directions of each reaction, as illustrated in Table 1. The irreversible reactions are identified by the original ID expanded adding the suffix 'reverse'. Also, the reactions involving the use of isozymes are split into several irreversible reactions. Similarly, their ID is modified adding the suffix 'numx', where x indicates the isoenzyme. Similarly to standard GEMs, irreversible reactions are defined setting the lower bound to zero and the upper one to 1000 mmol gDW^{-1} h^{-1}. The eciML1515 model includes an enzymatic constraint that limits the predicted metabolic fluxes to values compatible with the enzymes' availability experimental measured in *E. coli*.

This constraint is formulated as a single equation (*Eq.1*) that constrains the total sum of metabolic fluxes included in the model to the enzyme availability observed for *E. coli*. The total sum of fluxes is computed considering the flux associated with each reaction (*vi*) modulated by the molecular weight (*MWi*), saturation coefficient (σi) and turnover number *(Kcati*) of the enzyme that catalyses the reaction. The enzyme availability is determined by multiplying the total proteome *(ptot*), experimentally measured, and

the enzyme fraction (*f*), considering the ratio between the sum of abundances of all enzymes expressed in the model and the sum of abundances expressed in the experimental proteome (*Eq.2*).

$$\sum_{i=1}^{n} \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \leq ptot \cdot f \qquad \textit{(Eq. 1)}$$

V_i *:* flux reaction i
MW_i *:* Molecular Weight enzyme i
σ_i *:* saturation coefficient enzyme i
k_{cati} *:* turnover number enzyme i
ptot : measured total proteome
f = enzyme fraction

$$f = \sum_{i=1}^{p\_num} A_i MW_i / \sum_{j=1}^{g\_num} A_j MW_j \qquad \textit{(Eq. 2)}$$

A_i *:* enzyme abundance enzyme i (model)
MW_i *:* MW enzyme i (model)
A_j *:* abundance enzyme j (proteome
MW_j *:* MW enzyme j (proteome)
p_num : number of enzymes in the model
g_num : number of enzymes in the proteome

Unlike iML1515, all reactions in eciML1515, including the exchange ones, are defined as irreversible. For this reason, the metabolites' exchange between the intracellular and extracellular compartment is defined by a pair of irreversible reactions that return only positive fluxes.

1.2.3 Protein allocation Model pciML1515

The pciML1515 is a protein-constrained model developed by using the reference model ML1515. Besides the stoichiometric matrix, which describes the stoichiometric constraint, the model integrates the enzyme constraint built on the Gecko framework properly expanded to consider the protein allocation sources cost and a volume constraint (Grigaitis et al., 2021; Sánchez et al., 2017).

In pciML1515, all the catalytic reactions are defined as irreversible reactions; the reactions originally reversible are split to define the forward and reverse direction. The exchange reactions, defined in iML1515 as reversible, remain unchanged, as shown in Table 1. The irreversible reactions are identified by the ID name followed by the suffix '*fwd*' or '*rev*'. The reactions are expanded to include an additional species, representing the corresponding enzyme. The enzyme enters the reaction with a stoichiometric coefficient specific for each protein, defined as $(k_{cati})^{-1}$. The product of the enzyme abundance and, [*Ei]* and of its catalytic constant defines the maximum flux capacity for the specific enzyme *i,* as shown in *Eq.3*. In this formulation, the enzyme is treated as additional substrate, of which availability limits the reaction, introducing a direct linkage between enzyme abundance and the metabolic flux.

$$\sum_{j \in R_i} v_j \leq k_{cat,i} \times [E_i] \qquad \textit{(Eq. 3)}$$

The pciML1515 model is integrated with additional reactions describing the synthesis, folding and degradation processes. The active enzyme is converted, during the reaction, to the 'used enzyme' E_i **,* as indicated in *Eq.4.*

$$A + B + \left(k_{cat,i}^{\square}\right)^{-1} \times E_i \rightarrow C + D + \left(k_{cat,i}^{\square}\right)^{-1} \times E_i^{*} \qquad \textit{(Eq. 4)}$$

E_i ** is* retrieved at the end of the reaction to insert it into further pathways involved in the protein degradation process. The detailed description of the protein life cycle is central to defining the constraint imposed by the enzyme availability. Indeed, the enzyme dispensability is determined by the activity of biosynthetic and degradative reactions constrained by experimental proteomic data. The reactions involved in the enzymes' biosynthesis play central role in the determination of enzymatic availability, they are constraining according to the experimental absolute abundance, referred to specific enzymes, and total protein abundance. For protein synthesis, the costs of carbon, nitrogen and ATP are considered while the degradation reactions permit to account for the turnover costs. This formulation allows to account for the sources allocation across the metabolic pathways, potentially providing a highly accurate description of the metabolism. Additionally, the model includes an extra constraint that limits the total enzymatic pool according to the available cellular volume.

Table 1 – Definition of reactions in across the models.

The table display a schematic representation of the definition of catalytic and exchange reactions present in the three models.

iML1515	eciML1515	pciML1515
Enzymatic reactions $A \leftrightarrow B$ $A \rightarrow B$ $B \rightarrow A$	*Enzymatic reactions* $A \rightarrow B$ $B \rightarrow A$	*Enzymatic reactions* $A \rightarrow B$ $B \rightarrow A$
Exchange reaction $A \Leftrightarrow B$	*Exchange reaction* $A \Rightarrow B$	*Exchange reaction* $A \Leftrightarrow B$

1.3 Experimental Datasets

In this project, the selected models were tested using three different datasets containing experimental data concerning the central carbon metabolism of *Escherichia coli*, collected under different growth conditions.

The first two, referred as the Exponential growth phase and Chemostat datasets, include published fluxomic data. For the purpose of analyses, a subset of fluxes was selected, which were related to the exchange and intracellular fluxes associated with the pentose phosphate pathway, the TCA cycle, the glycolytic pathway, the glyoxylate shunt and the ED pathway. The third one, indicated as Batch dataset, contains unpublished data obtained from a collaborator at the Bioprocess Engineering group. The dataset includes extracellular concentrations measured during anaerobic growth.

1.3.1 Exponential phase Dataset

The exponential growth phase dataset (Gonzalez et al., 2017) was produced from batch growth under oxic and anoxic conditions, using glucose as the sole carbon source. Despite the batch growth mode, the fluxomic data provided in the dataset do not correspond to the full timeline of the whole batch process; instead, each reaction in the dataset is reported as a single flux value representing only one time point with the corresponding confidence interval of 95%. The reported data are assumed to be referred to the exponential growth phase. The metabolic fluxes were obtained through the ^{13}C metabolic flux analysis (^{13}C-MFA) methodology(Crown & Antoniewicz, 2013).

1.3.2 Chemostat Dataset

The Chemostat dataset (Toya et al., 2012) contains fluxomic data measured from a chemostat culture of *E. coli* under oxic and anoxic conditions. The metabolic fluxes were measured from a chemostat growth process obtained setting the dilution rate to 0.2 h^{-1} and using glucose as the sole carbon source. The data were determined via the ^{13}C-MFA methodology. The computed fluxes are reported as optimal flux associated with the corresponding confidence interval of 95%.

1.3.3 Anaerobic batch dataset

The Batch dataset includes the extracellular concentrations of glucose and the main fermentative products of *E. coli* metabolism such as acetate, ethanol, succinate, formate and lactate. The data were collected, in duplicate, from an anaerobic batch growth process using High-performance liquid chromatography (HPLC) for the fermentative products and YSI analyser for the glucose. The concentrations are provided for each timepoint in the whole timeline of the batch process.

2. Methods

The analyses were performed in Python 3.11 (Van Rossum, 2009)using the COBRA Toolbox version (Ebrahim et al., 2013).
All the scripts are available at (git.wur.nl/ssb/studentprojects/vanessa_carta_thesis_msc).

2.1 Simulation of the models

2.1.1 Models loading

The iML1515 was obtained from the BIGG webpage (http://bigg.ucsd.edu/models/iML1515) and loaded with Cobra. Then, the eciML1515 was retrieved from GitHub (https://github.com/tibbdc/ECMpy) where specific helper functions to replicate their result are also available. Loading the model with the information provided on their own documentation was not trivial. To facilitate the model's simulations, a brief description of the model loading steps is included here, as well as in the project code available at https://git.wur.nl/ssb/student-projects/vanessa_carta_thesis_msc Initially, the model was converted to a Cobra readable format, using the helper functions *'get_enzyme_constraint_model'* and '*json_load*', and then loading it with Cobra.

Finally, the pciML1515 model was collected from the GitHub pranasag/extendedEcoliGEM and loaded with Cobra. Before performing any analysis, it was necessary to add the enzyme constraint integrating the proteomic data contained in the '*protein_masses_file*' using the helper function '*add_proteome_constraint_cobra*' available in the above mentioned GitLab.

2.1.2 Oxic and Anoxic conditions settings

To simulate oxic conditions, the models have access to unlimited oxygen. For this purpose, in iML1515 and pciML1515 the lower and upper bounds of *EX_o2_e* were respectively set to -1000 mmol gDW^{-1} h^{-1} and +1000 mmol gDW^{-1} h^{-1}. In eciML1515, the lower and upper bounds of *EX_o2_e_reverse* and *EX_o2_e*' bounds were respectively fixed to 0 and 1000 mmol gDW^{-1} h^{-1}.

Conversely, in order to simulate anoxic conditions, the oxygen exchange was constrained to zero. For this purpose, the boundaries of *EX_o2_e* were set to zero in iML1515 and pciML1515. In eciML1515, the boundaries of both *EX_o2_e* and *EX_o2_e_reverse* were set to zero.

2.1.3 Glucose exchange reactions settings

In the iML1515 and pciML1515 models, the glucose exchange reactions are reversible and identified by the *EX_glc_D_e ID*. Its upper and lower bounds are set to -10 mmol gDW^{-1} h^{-1} and +1000 mmol gDW^{-1} h^{-1} by default. In the eciML1515 model, the glucose exchange is described by two irreversible reactions, *EX_glc__D_e* and *EX_glc__D_e_reverse.* The former indicates the glucose flow from the intracellular toward the extracellular compartment, while the latter represents the glucose uptake. By default, the reaction bounds are set to 0 and 10 mmol gDW^{-1} h^{-1} for *EX_glc__D_e_reverse* and to 0 and 1000 mmol gDW^{-1} h^{-1} for *EX_glc__D_e.*

To constrain the glucose uptake in eciML1515, the bounds of *EX_glc__D_e* are set to zero, while the bounds of *EX_glc__D_e_reverse* are fixed to the selected value for the glucose uptake rate. From now on, *EX_glc__D_e* in iML1515 and pciML1515, and *EX_glc__D_e_reverse* will be referred as glucose uptake reactions.

2.1.4 FVA under unlimited glucose conditions

In contrast to the standard approach that implies to constrain the carbon source, the models were initially analysed through FVA under unlimited glucose conditions. This procedure was specifically performed to identify the factor that constrains the models and evaluate the effect of the enzyme constraint on the predictions. The unlimited glucose conditions in iML1515 and pciML1515 were simulated setting the upper and lower bounds of the glucose uptake respectively to -1000 mmol gDW^{-1} h^{-1} and +1000 mmol gDW^{-1} h^{-1}. In eciML1515, the *EX_glc__D_e* and *EX_glc__D_e_reverse* were respectively set to 0 and +1000 mmol gDW^{-1} h^{-1}. The FVA was performed considering a flux range for reactions at 95% optimality. Subsequently, FVA was performed setting the glucose uptake to 10 gDW^{-1} h^{-1} in all models. The FVA was performed constraining the models according to the oxic and anoxic settings described in section 2.1.2.

2.1.5 Chemostat FBA

To remove default constraints, the glucose exchange reactions were initially set to allow unlimited glucose. The biomass flux was constraining to the dilution rate of 0.2 h^{-1}, according to the dilution rate used in the Chemostat dataset. The glucose exchange reaction was selected as the objective function to optimise. The optimisation direction was set to minimize the amount of consumed glucose. The FBA was performed constraining the models according to the oxic and anoxic settings described in section 2.1.2.

2.1.6. Exponential growth phase FBA

The models were constrained in the glucose uptake according to the experimental glucose included in the Exponential growth phase dataset. The glucose exchange reactions were set to mmol 8.8 gDW^{-1} h^{-1} and 13.1 mmol gDW^{-1} h^{-1} respectively to perform FBA under oxic and anoxic conditions. The biomass reaction was selected as the objective function to maximise. The FBA were performed constraining the models according to the oxic and anoxic settings described in section 2.1.2.

2.1.7 Sampling

The flux solution space was sampled with the Artificial Cantering Hit-and-Run (ACHR) sampling algorithm, available in Cobra. The sampler was set to collect 1000 samples using a thinning factor equals to 10. The feasibility of each sample was verified through the sampler '*validate function' (achr.validate)*. Only the samples checked to respect the model's constraints were allowed in further analyses. The sampling process was operated constraining the biomass and glucose exchange reactions instead of optimizing a specific function.

To sample the models simulating the chemostat mode, the biomass flux was constrained to the dilution rate of 0.2 h^{-1}, including a variability of the 2% both to the upper and lower bound (resulting in a variation between the 98% and 102% of the dilution rate), and the glucose uptake was set according to the results of FBA described in section 2.1.5.
Similarly, to sample the models simulating the fluxes at the exponential growth phase, the glucose uptake was settled to match the experimental data of the exponential growth phase dataset and the biomass was constrained according to the results of FBA described in section 2.1.6, including the variability of 2%.

The median and median absolute deviation of samples were computed for each reaction present in model, using *statistics (3.4) and* scypy.*stats(1.13.0)* modules. For reaction with several isoenzymes active imultaneously, the final flux was computed as the sum of the median of fluxes when the difference between isoenzymes' fluxes was smaller than two order of magnitude. The total median absolute deviation was obtained by summing the squared median absolute deviation of each considered flux and calculating the square root of it.

2.1.8 Batch simulations

The batch process was simulated using the 'batch function' developed byMoreno-Paz et al., 2022, and properly adapted to the metabolic models benchmarked in this project. The batch simulations were performed to replicate the data included in the Batch dataset. Table 2 reports the parameters used in the simulations.

Table 2 – Parameters of Batch simulations.

Parameters		
c_s0	*47.2*	*mmol/L*
c_x0	*0.0384*	*gDW/L*
V	*0.05*	*L*
qs_max	*6.6*	*mmol_gDW/h*
Kms	*0.53*	*mmol/L*
t_end	*25*	*h*

C_{s0}: Initial glucose concentration in the bioreactor
C_{x0}: Initial biomass concentration
V : volume of bioreactor
q_{Smax}: Maximum glucose uptake rate
K_{MS}: Michaelis-Menten constant for glucose uptake

The initial biomass concentration was computed from the OD_{600nm} values reported in the Batch dataset and the relation 1 OD_{600nm} = 0.32 gDW/L described by Gonzalez et al., 2017 . The maximum glucose uptake rate and the Michaelis-Menten costant were obtained obtained by Sara Moreno Paz, 2019.

2.2 Evaluation of models' Accuracy

The quality of models' predictions was evaluated computing the coefficient of determination (R^2) between the predicted and experimental fluxes. The R^2 scores were computed using the *sklearn.metrics* 0.11

Additionally, the estimate error was calculated between each measure flux available in the datasets and the corresponding flux predicted by the three models.

3. Results

3.1 FVA under unlimited glucose conditions

The FVA analyses under unlimited glucose availability are not comparable to the data contained in the experimental datasets since they do not represent a realistic scenario. However, these simulations could highlight the differences concerning the constraining factors across the three models, when the carbon source is not limited.

The FVA results indicate that in eciML1515 and pciML1515 additional constraints are present, as suggested by the relatively narrow flux intervals that were obtained when unlimited availability of glucose is considered. As expected, iML1515 shows wide flux ranges indicating the absence of further limiting factors. In Table 3, the results of FVA are displayed, including the computed fluxes' intervals of biomass, exchange reactions of fermentative products and glucose, and representative intracellular fluxes.

Table 3 – FVA under unlimited glucose availability.
The table reports the FVA results of iML1515, eciML1515 and pciML1515 under oxic and unoxic conditions. The fluxes are expressed in mmol gDW-1 h-1.

	iML1515 oxic		iML1515 anoxic		eciM1515 oxic		eciML1515 anoxic		pciML1515 oxic		pciML1515 anoxic	
	min	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
Biomass	*24.81*	*26.12*	*10.48*	*11.03*	*0.68*	*0.71*	*0.31*	*0.33*	*0.64*	*0.67*	*0.62*	*0.65*
Glucose uptake	*473.13*	*1000*	*518.71*	*1000*	*9.93*	*1000*	*17.31*	*1000*	*83.12*	*230.62*	*103.45*	*144.70*
Ethanol	*0*	*1000*	*473.92*	*1000*	*0*	*3.04*	*11.12*	*26.74*	*0*	*19.31*	*0*	*6.90*
Pyruvate	*0*	*120.16*	*0*	*99.41*	*0*	*30.88*	*0*	*24.29*	*0*	*232.48*	*0*	*30.22*
Succinate	*0*	*64.24*	*3.50*	*90.48*	*0*	*1.68*	*0.10*	*0.91*	*0*	*18.77*	*0.21*	*13.90*
Formate	*216.33*	*1000*	*378.35*	*657.17*	*0*	*7.74*	*16.56*	*36.62*	*0*	*9.48*	*0*	*9.92*
Acetate	*166.16*	*1000*	*170.35*	*518.30*	*0*	*11.99*	*0*	*17.44*	*0*	*61.18*	*0*	*6.67*
Lactate	*0*	*259.55*	*0*	*347.94*	*0*	*0.83*	*0*	*0.96*	*0*	*248.67*	*192.42*	*246.93*
PDH	*0*	*1000*	*245.13*	*1000*	*0*	*14.86*	*0*	*12.65*	*0*	*2.08*	*0*	*2.37*
PGL	*0*	*1000*	*12.20*	*654.26*	*0*	*10.38*	*0*	*4.27*	*0*	*13.05*	*0*	*13.15*
PYK	*0*	*1000*	*0*	*594.64*	*0*	*14.50*	*0*	*22.19*	*0*	*250.95*	*0*	*248.29*
ACONTa	*26.68*	*319.07*	*11.27*	*35.26*	*0*	*2.28*	*0*	*1.63*	*0*	*15.61*	*0*	*15.10*
CS	*26.68*	*319.07*	*11.27*	*69.26*	*0.73*	*3.32*	*0.33*	*1.35*	*0.68*	*12.73*	*0.67*	*7.07*

The iML1515 model returned large fluxes' intervals, indicating that the model is deeply influenced by the glucose availability, which was set unlimited for this analysis. Most of the fluxes are limited to the upper bound of reactions (1000 gDW^{-1} h^{-1}) while others display unrealistic values such as the glucose uptake, the formate and acetate fluxes. It is noteworthy that the values associated with the acetate flux are unexpected, considering that the model inability to describe the overflow metabolism is a known limitation of iML1515. Differently, eciML1515 clearly demonstrates the enzyme constraint impact,

returning realistic fluxes' intervals comparable to the resulta computed under physiological glucose uptake (Table 4).

Similarly, pciML1515 appeared constrained by the enzyme availability, especially considering the biomass fluxes under both conditions. However, several fluxes are characterised by unrealistic intervals such as the lactate and glucose fluxes. The comparison between the eciML1515 and pciML1515 displays relevant discrepancies between the predictions, suggesting a more impacting constriction on eciML1515. Remarkably, both models return anomalous values due to the unrealistic simulation settings. For instance, all the three models predict the PDH reaction as potentially inactive under oxic conditions.

Table 4 – FVA under glucose uptake 10 gDW^{-1} h^{-1}

The table reports the FVA results of iML1515, eciML1515 and pciML1515 under oxic and unoxic conditions. The fluxes are expressed in mmol gDW-1 h-1.

	iML1515 oxic		iML1515 anoxic		eciM1515 oxic		eciML1515 anoxic		pciML1515 oxic		pciML1515 anoxic	
	min	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
Biomass	*0.83*	*0.88*	*0.15*	*0.16*	*0.65*	*0.68*	*0.15*	*0.15*	*0.28*	*0.29*	*0.03*	*0.03*
Glucose uptake	*9.51*	*10*	*9.64*	*10*	*9.14*	*10*	*9.64*	*10*	*9.38*	*10*	*9.71*	*10*
Ethanol	*0*	*1.06*	*6.99*	*10.62*	*0*	*1.60*	*7.02*	*10.65*	*0*	*1.25*	*0*	*0.61*
Pyruvate	*0*	*1.26*	*0.00*	*1.21*	*0*	*3.56*	*0*	*1.21*	*0*	*2.11*	*0*	*0.48*
Succinate	*0*	*0.86*	*0.05*	*1.86*	*0*	*1.58*	*0.05*	*1.84*	*0*	*1.01*	*0.01*	*0.44*
Formate	*0*	*5.83*	*14.69*	*21.94*	*0*	*3.76*	*14.73*	*21.98*	*0*	*4.43*	*0*	*1.22*
Acetate	*0*	*1.79*	*6.77*	*10.70*	*4.03*	*9.12*	*6.75*	*10.73*	*0*	*4.02*	*0*	*0.55*
Lactate	*0*	*1.06*	*0*	*3.63*	*0*	*0.93*	*0*	*3.63*	*0*	*1.59*	*18.41*	*19.52*
PDH	*0*	*42.02*	*0*	*3.63*	*7.78*	*14.10*	*0*	*3.63*	*0*	*1.72*	*0*	*0.18*
PGL	*0*	*23.97*	*0*	*1.82*	*0.28*	*10.15*	*0*	*1.86*	*0*	*8.53*	*0*	*0.52*
PYK	*0*	*23.39*	*0*	*12.40*	*0*	*13.60*	*0*	*12.43*	*0*	*25.23*	*0*	*19.93*
ACONTa	*0.90*	*13.22*	*0.16*	*0.51*	*0*	*2.76*	*0*	*14.15*	*0*	*26.39*	*0*	*1.41*
CS	*0.90*	*19.44*	*0.16*	*1.07*	*0.69*	*2.95*	*0.16*	*1.06*	*9.67*	*18.59*	*0.03*	*0.31*

Table 4 reports the FVA results when the glucose uptake is set to 10 gDW^{-1} h^{-1}. Under these conditions the constraint on the carbon source has a relevant effect on iML1515. The comparison of the predicted intervals indicates that iML1515 and eciML1515 returned comparable values concerning the production of fermentative products, intracellular fluxes, and biomass under anoxic conditions. Differently, pciML1515 did not predict the production of ethanol, formate and acetate providing fluxes' intervals close to zero, while the model suggests the production of lactate, which is missing in the other models.

As expected, iML1515 did not predict acetate production under oxic conditions, returning a range comprised between 0 and 1.79 mmol gDW^{-1} h^{-1}. Conversely, eciML1515 describes the overflow metabolism with an acetate flux that spans from 4.03 to 9.12 mmol gDW^{-1} h^{-1} and predicts the PDH reaction to be active with possible fluxes from 7.78 to 14.1 mmol gDW^{-1} h^{-1}. Then, the pciML1515 model returned results similar to iML1515 by not predicting either the acetate production or the PDH flux. Furthermore, the biomass fluxes appear remarkably underestimated compared the eciML1515 and iML1515 both under oxic and anoxic conditions.

3.2 Chemostat FBA

The FBA was used to predict the glucose uptake simulating the chemostat mode. The glucose uptake fluxes were computed using FBA by constraining the experimental dilution rate reported in the chemostat dataset (D = 0,2 h^{-1}) and minimising the glucose uptake flux. Table 5 summarizes the FBA results.

Table 5 – Chemostat FBA results

	Glucose		D
	Oxic	Anoxic	
Data	*3.2*	*10.9*	*0.2*
iML1515	*2.5*	*11.96*	*0.2*
eciML1515	*2.52*	*12.18*	*0.2*
pciML1515	*6.71*	*36.38*	*0.2*

The fluxes reported in table 5 show that iML1515 and eciML1515 predicted comparable glucose uptake fluxes, with values close to 2.5 mmol gDW^{-1} h^{-1} and 12 mmol gDW^{-1} h^{-1} under oxic and anoxic conditions respectively. Differently, the pciML1515 returned discordant glucose fluxes, notably higher in comparison with the previous results, with fluxes of 6.71mmol gDW^{-1} h^{-1} under oxic conditions and 36.38 mmol gDW^{-1} h^{-1} under anoxic ones.

3.3 FBA to simulate the exponential phase of growth

The FBA analysis was used to predict the biomass fluxes associated with the measured glucose uptake reported in the Exponential growth phase. The analyses were performed setting the glucose uptake of to 8.8 mmol gDW^{-1} h^{-1} and 13.1 mmol gDW^{-1} h^{-1} to simulate the exchange fluxes respectively under oxic and anoxic conditions with the three models. Table 6 reports the biomass fluxes that resulted from the FBA.

Table 6 – Exponential growth phase FBA results

	Oxic		Anoxic	
	Glucose	Biomass	Glucose	Biomass
Data	*8.8*	*0.824*	*13.1*	*0.487*
iML1515	*8.8*	*0.768*	*13.1*	*0.131*
eciML1515	*8.8*	*0.632*	*13.1*	*0.128*
pciML1515	*8.8*	*0.262*	*13.1*	*0.053*

The biomass fluxes predicted by iML1515 and eciML1515 appeared in an overall agreement to the experimental data both under oxic and anoxic conditions, although the values computed by iML1515 were slightly higher especially under oxic conditions. Conversely, the results by pciML1515 were notably lower, with values of 0.26 mmol gDW^{-1} h^{-1} and 0.05 mmol gDW^{-1} h^{-1} respectively under oxic and anoxic conditions.

3.4 Sampling

The sampling of pciML1515 was not successfully completed. The sampling process was correctly kickstarted several times, but unfortunately the program got stuck during each attempt, regardless of the variation of settings. On the other hand, the iML1515 and eciML1515 models were efficiently sampled, simulating all considered conditions.

3.4.1 Sampling in Chemostat

The models were sampled constraining the glucose exchange and biomass reactions to replicate the metabolic fluxes included in the Chemostat dataset. The glucose uptake was set according to the FBA results presented in section 3.2. The biomass was set to 0.2 mmol gDW^{-1} h^{-1} in accordance with the

experimental dilution rate. The sampling results are presented in the following scatter plots, where the measured fluxes are plotted against the predicted fluxes. The exchange and intarcellular fluxes are presented separately.

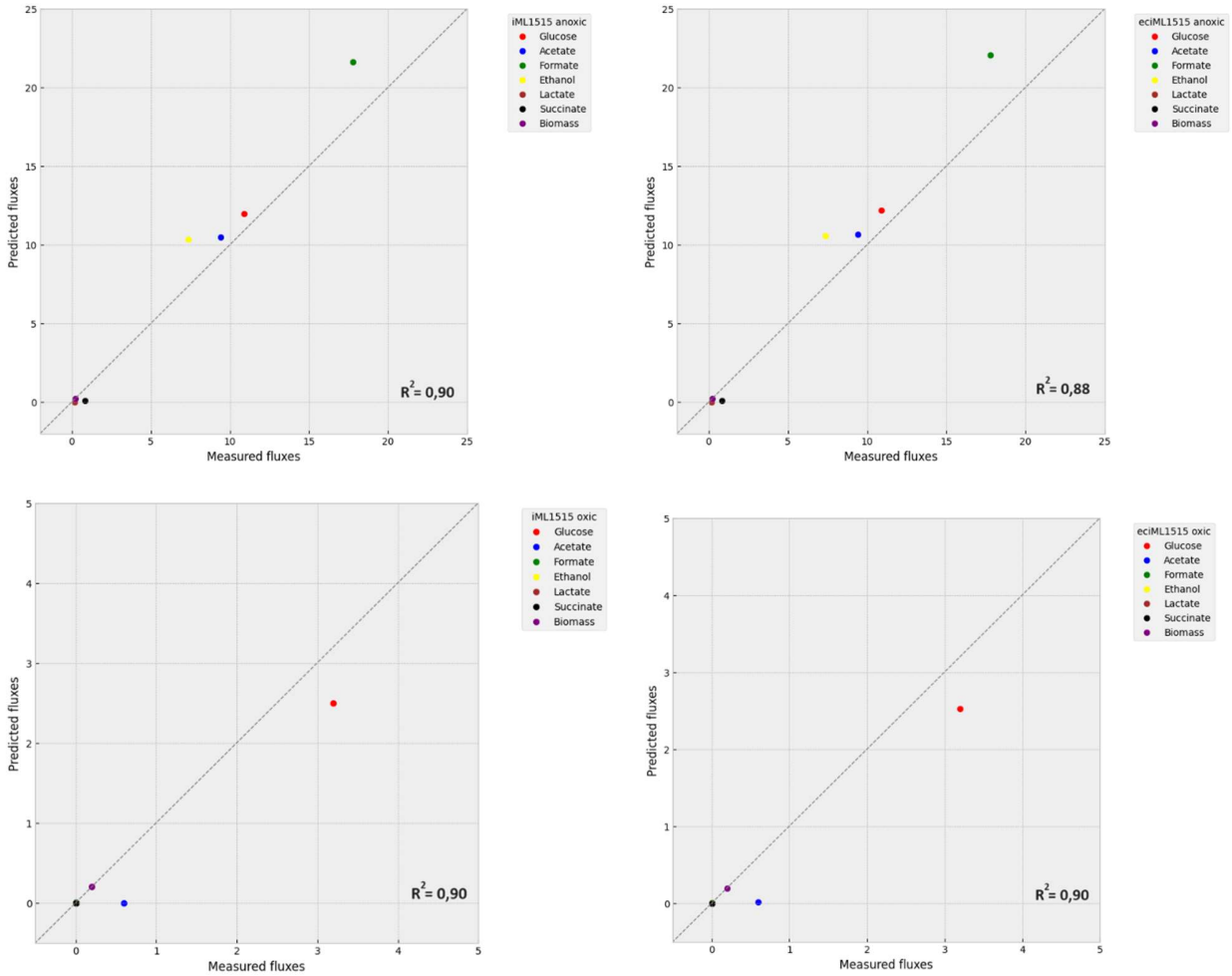

Figure 1- Scatter plot of Exchange reaction fluxes in chemostat growth mode under anoxic and oxic conditions.

The computed R^2 scores suggested comparable performances for eciML1515 and iML1515 in predicting the exchange fluxes, demonstrating a higher predictive accuracy under anoxic conditions. The scatter plots suggest that both models were able to predict the absence of fermentative products under oxic conditions in agreement with the data. Noteworthy, the two models did not predict the acetate production which, indeed, appeared remarkably low also in the dataset. The main discordance concerns the glucose uptake flux that appeared underestimated by both models compared to the dataset. Under

anoxic conditions, the eciML1515 and iML1515 demonstrated comparable accuracy with R^2 scores of 0.90 and 0.88 respectively.

To get insights on the predictions, the following bar charts show the direct comparison between the predictions computed by both eciML1515 and iML1515 models, together with the data highlighting the differences across measured and predicted exchange fluxes.

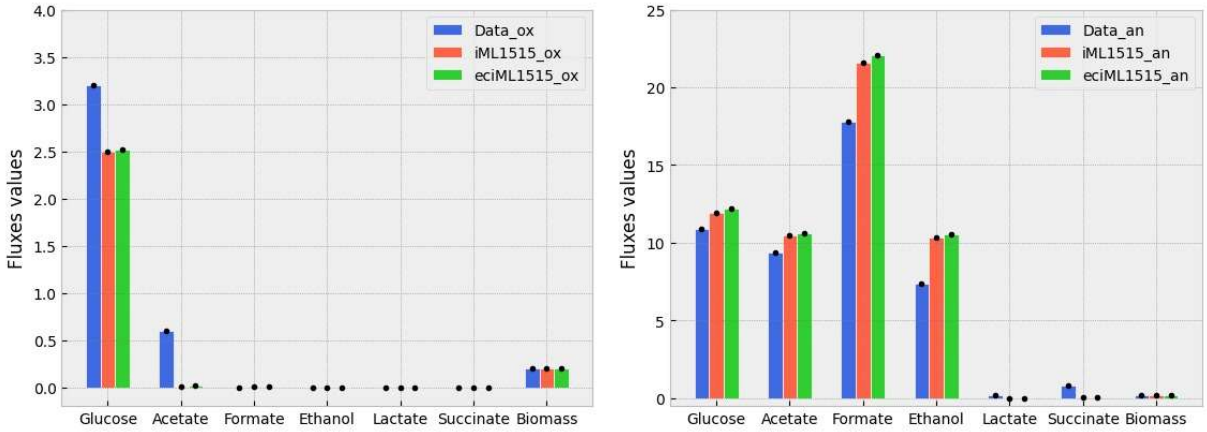

Figure 2 – Bar charts of Exchange reaction fluxes under oxic and anoxic conditions.

The fluxes comparison indicates that, under oxic conditions, the two models returned a notably similar values for glucose uptake fluxes close to 2.5 mmol gDW^{-1} h^{-1}. This shows a slight underestimation compared to the experimental flux of 3.2 mmol gDW^{-1} h^{-1}. Then, the bars show the presence of a limited production of acetate settled to 0.6 mmol gDW^{-1} h^{-1}. As displayed in the plot, iML1515 did not predict the acetate flux while the eciML1515 computed an extremely low flux, which is not visible in the plot but that is reported in Table A1 in the appendix. Under anoxic conditions, the models simulated fluxes that were overall in agreement with the data, even though they moderately increased in the glucose uptake, acetate, formate and ethanol fluxes. Moreover, the succinate flux was not predicted by either model.

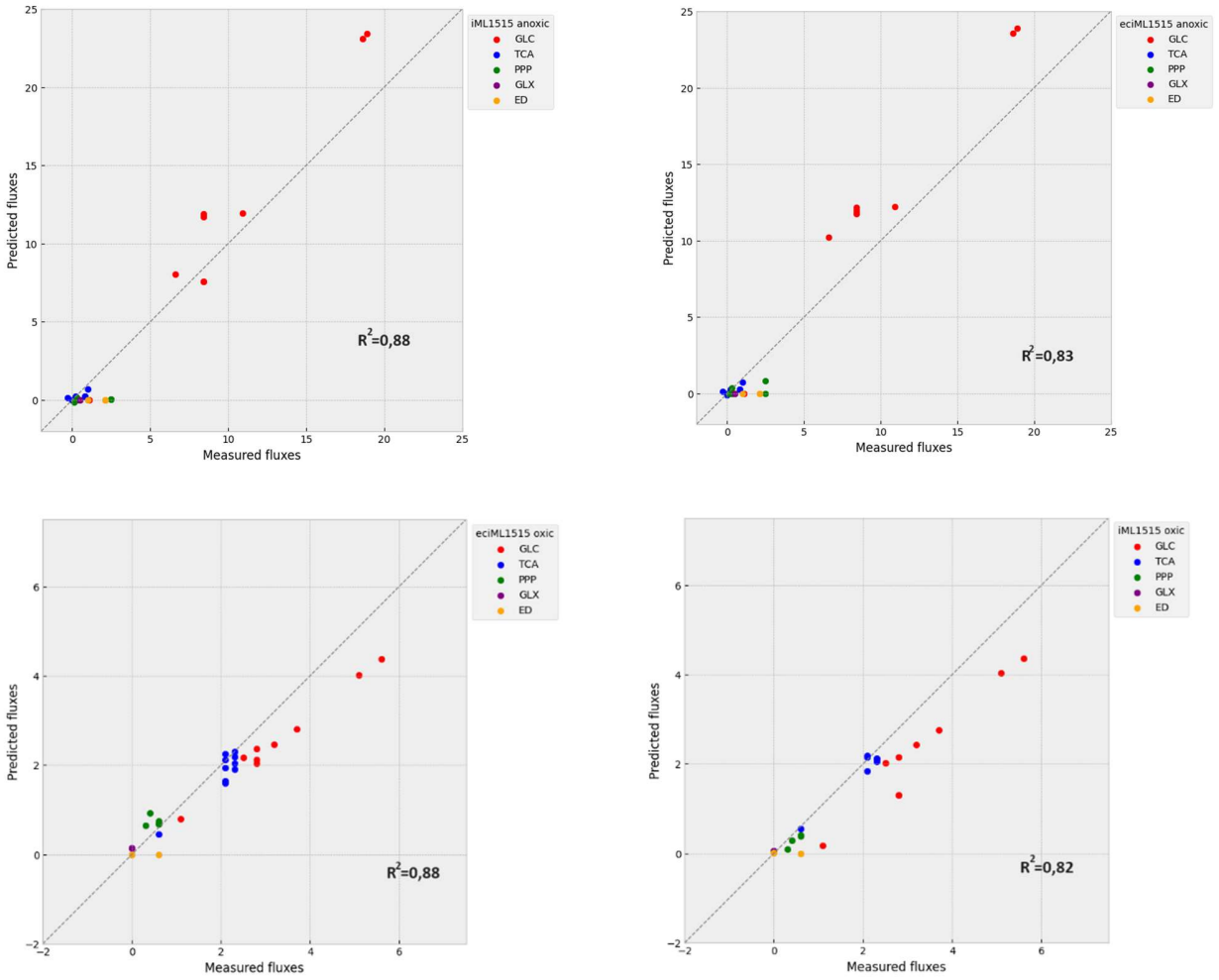

Figure 3 – Scatter plots of Intracellular fluxes chemostat under anoxic and oxic conditions.

Initially, the R^2 computed for the iML1515 predictions were characterized by remarkably negative values both under oxic and anoxic conditions as shown in Table A2 of the appendix, suggesting a complete inability of iML1515 of simulating the metabolic fluxes of interest. However, the fluxes predicted by iML1515 did not display notable discrepancies compared to the data, apart from the two fluxes SUCOAS and SUCDi. Thus, these two fluxes were removed from the dataset and the R^2 were recalculated. The new R^2 scores provided a more realistic indication of iML1515 accuracy.

The computed R^2 scores suggest that iML1515 and eciML1515 showed a comparable accuracy in the prediction of intracellular fluxes under oxic and anoxic conditions.

Interestingly, the scatter plots highlight how fluxes belonging to different pathways were characterized by different accuracy levels. Under oxic conditions, the pentose phosphate pathway fluxes computed by iML1515 were more consistent to data compared to eciML1515. As indicated by the plots, the former model returned underestimated fluxes, while the latter overestimated them. Then, the glycolytic pathway and TCA cycle fluxes were underestimated by both models compared to measured data. Contrastingly, under anoxic conditions, the glycolytic fluxes suggest a reduced accuracy in the fluxes' computation. In fact, the values appeared overestimated by both models, whereas the other pathways were underestimated compared to the data.

The following tables provide an overview of the of intracellular fluxes. The experimental data are compared to the iML1515 and eciML1515 predictions.

Table 7 – Comparison of experimental and predicted intracellular fluxes under anoxic and oxic conditions.

		Anoxic conditions			Oxic conditions		
Pathways	ID	Data	iML1515	eciML1515	Data *	iML1515	eciML1515
Glycolitic	GLCptspp	*10,9 ± 0*	*11,93 ± 0,02*	*12,2 ± 0,42*	*3.2*	*2,43 ± 0,03*	*2,46 ± 0,07*
Glycolitic	PGI	*8,37 ± 0,26*	*11,91 ± 0,02*	*12,19 ± 0,02*	*2.5*	*2,03 ± 0,06*	*2,16 ± 0,03*
Glycolitic	PFK	*8,41 ± 0,31*	*7,6 ± 0,74*	*11,75 ± 0,11*	*2.8*	*1,31 ± 0,13*	*2,03 ± 0,01*
Glycolitic	FBA	*8,41 ± 0,31*	*7,59 ± 0,74*	*11,82 ± 0,09*	*2.8*	*1,3 ± 0,12*	*2,12 ± 0,06*
Glycolitic	TPI	*8,41 ± 0,31*	*11,73 ± 0,01*	*11,98 ± 0,06*	*2.8*	*2,16 ± 0,02*	*2,37 ± 0,29*
Glycolitic	GAPD	*18,92 ± 0,31*	*23,41 ± 0,01*	*23,89 ± 0,01*	*5.6*	*4,37 ± 0,02*	*4,37 ± 0,05*
Glycolitic	ENO	*18,62 ± 0,31*	*23,09 ± 0,01*	*23,56 ± 0,01*	*5.1*	*4,04 ± 0,02*	*4,02 ± 0,04*
Glycolitic	PYK	*6,63 ± 0,31*	*8,02 ± 0,44*	*10,23 ± 0,11*	*1.1*	*0,17 ± 0,04*	*0,79 ± 0,03*
Glycolitic	PDH	*1,1 ± 0,08*	*0,01 ± 0,01*	*0 ± 0*	*3.7*	*2,76 ± 0,03*	*2,81 ± 0,02*
Glycolitic	PFL	*17,76 ± 0*	*21,64 ± 0,03*	*22,04 ± 0,14*	-	*0,03 ± 0,01*	*0 ± 0*
TCA cycle	CS	*0,75 ± 0,08*	*0,21 ± 0*	*0,28 ± 0*	*2.3*	*2,12 ± 0,01*	*2,04 ± 0,01*
TCA cycle	ACONTa	*0,75 ± 0,08*	*0,21 ± 0*	*0,29 ± 0,03*	*2.3*	*2,12 ± 0,01*	*2,18 ± 0,05*
TCA cycle	ACONTb	*0,75 ± 0,08*	*0,21 ± 0*	*0,28 ± 0,01*	*2.3*	*2,12 ± 0,01*	*2,3 ± 0,22*
TCA cycle	ICDHyr	*0,25 ± 0,07*	*0,21 ± 0*	*0,28 ± 0*	*2.3*	*2,06 ± 0,02*	*1,91 ± 0,04*
TCA cycle	AKGDH	*0 ± 0,07*	*0 ± 0*	*0 ± 0*	*2.1*	*1,84 ± 0,02*	*1,65 ± 0,03*
TCA cycle	SUCOAS	*0 ± 0,07*	*121,87 ± 110,93*	*-0,09 ± 0,01*	*2.1*	*103,58 ± 41,26*	*1,59 ± 0,03*
TCA cycle	SUCDi	*0,33 ± 0,07*	*0 ± 0*	*0 ± 0*	*2.1*	*7,84 ± 0,46*	*1,94 ± 0,05*
TCA cycle	FUM	*0,33 ± 0,07*	*0,13 ± 0,01*	*0,13 ± 0*	*2.1*	*2,15 ± 0,02*	*2,13 ± 0,03*
TCA cycle	MDH	*0,17 ± 0,08*	*0,13 ± 0,01*	*0,11 ± 0,01*	*2.1*	*2,19 ± 0,02*	*2,25 ± 0,01*
TCA cycle	PPC	*0,98 ± 0,02*	*0,69 ± 0,01*	*0,73 ± 0,01*	*0.6*	*0,55 ± 0,01*	*0,46 ± 0,02*
PPP	PGL	*2,49 ± 0,26*	*0,03 ± 0,01*	*0,01 ± 0,01*	*0.6*	*0,4 ± 0,04*	*0,68 ± 0,02*
PPP	GND	*0,34 ± 0,2*	*0,02 ± 0,01*	*0,01 ± 0,01*	*0.6*	*0,39 ± 0,04*	*0,68 ± 0,02*
PPP	RPI	*0,27 ± 0,07*	*0,16 ± 0,01*	*0,35 ± 0,11*	*0.4*	*0,28 ± 0,01*	*0,93 ± 0,44*
PPP	RPE	*0,07 ± 0,13*	*-0,15 ± 0,01*	*0,01 ± 0,01*	*0.3*	*0,1 ± 0,02*	*0,65 ± 0,03*
Exchange	EX_lac__D_e	*0,19 ± 0*	*0,01 ± 0,01*	*0 ± 0*	*0*	*0 ± 0*	*0 ± 0*
Exchange	EX_etoh_e	*7,37 ± 0*	*10,33 ± 0,02*	*10,54 ± 0,01*	*0*	*0 ± 0*	*0 ± 0*
Exchange	EX_ac_e	*9,4 ± 0*	*10,47 ± 0,03*	*10,65 ± 0,01*	*0.6*	*0 ± 0*	*0,02 ± 0*

The fluxes reported in Table 7 indicate that the fluxes predicted by the two models were congruent to the measured data that were included in the Chemostat dataset. Overall, the models were able to predict the primary features of metabolism under oxic and anoxic conditions.

More specifically, the fluxes associated with the glycolytic pathway predicted by the two models agree with the data both under oxic and anoxic conditions. However, the PFK, FBA and PYK fluxes predicted by iML1515 are underestimated compared to measured fluxes, and this also occurs under both conditions. Similarly to the glycolytic pathway, the fluxes related to the TCA cycle computed by the two models are also in accordance with the data under both conditions. Some abnormal fluxes are present: iML1515, predicts extremely elevated fluxes for SUCOAS and SUCDi with values of 103.6 mmol gDW^{-1} h^{-1} and 7.8 mmol gDW^{-1} h^{-1} respectively under oxic conditions, and only related to SUCOAS with a value of 121.9 mmol gDW^{-1} h^{-1} under anoxic conditions. Lastly, the pentose phosphate pathway fluxes are overall underestimated by both eciML1515 and iML1515, and especially under anoxic conditions.

Moreover, by comparing the different pathways under anoxic conditions, the values of predicted fluxes of acetate and ethanol are in accordance with the measured data, even though the acetate flux is moderately overestimated. Interestingly, the Chemostat dataset includes a notably low acetate flux under oxic conditions, indicated only a marginal overflow metabolism. As expected, iML1515 did not predict the acetate production while eciML1515 computed an extremely low flux of 0.2 mmol gDW^{-1} h^{-1}.

3.4.2 Flux Sampling at the exponential growth phase

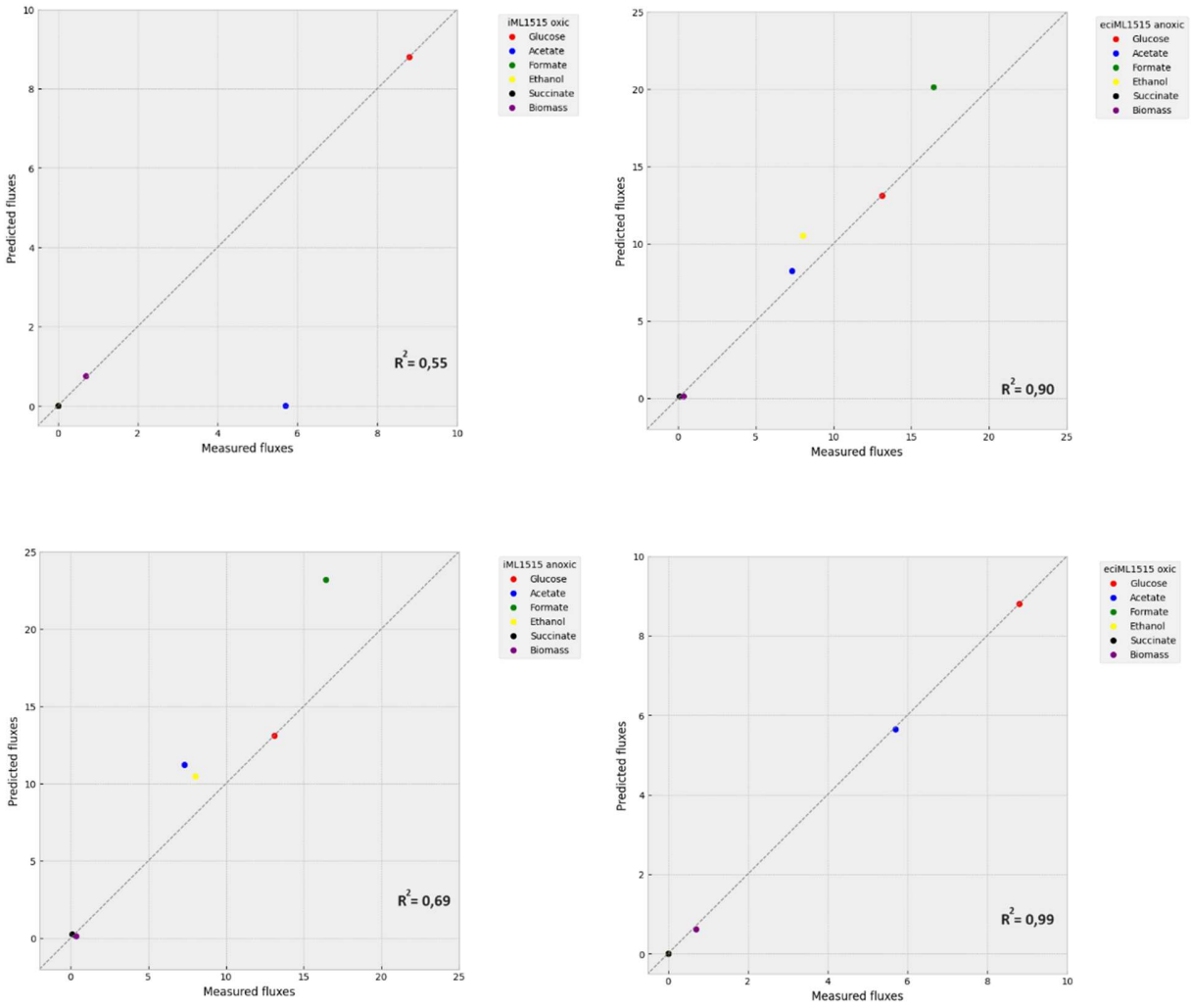

Figure 4 – Scatter plots Exchange fluxes exponential growth phase under anoxic and

The computed R^2 scores indicate that eciML1515 outperforms iML1515 in simulating the exchange fluxes collected at the exponential growth phase, returning predicted fluxes that match the experimental data more accurately. This result is supported by the remarkably increased values of the R^2 associated to eciML1515 in relation to iML1515 predictions, under all conditions. The comparison of scatter plots clearly displays that, under anoxic conditions, the fluxes predicted by eciML1515 describe more accurately the measured fluxes as displayed by the R^2 of 0.9; and that the predictive accuracy of acetate and formate fluxes is notably increased compared to iML1515 predictions.

Noteworthy, under oxic conditions, eciML1515 displays a high predictive power highlighted by a R^2 of 0.99, which suggests that all computed fluxes are congruent to the actual fluxes. Conversely, iML1515 predictions are associated with low R^2 scores, especially under oxic conditions (R^2 = 0.55). The low accuracy of iML1515 is caused by its inability to predict the overflow metabolism, whereas eciML1515 predicts the acetate production accurately.

The bar charts in figure 5 provide a visual representation of predicted and measured exchange flux values at the exponential growth phase, facilitating a direct comparison between the predictions computed by both eciML1515 and iML1515 models.

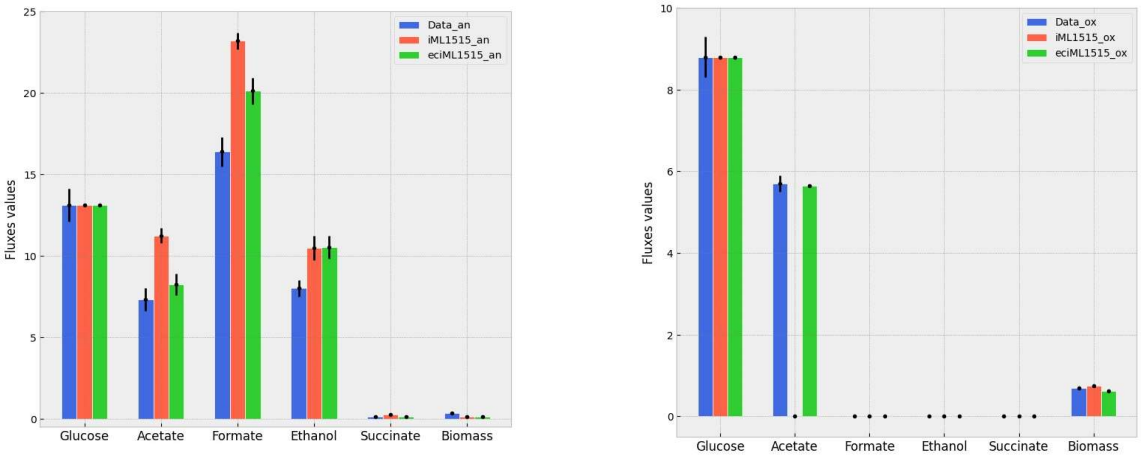

Figure 5 – Bar charts of Exchange fluxes at the exponential growth phase under anoxic and oxic conditions.

The bar charts comparison suggests that, under oxic conditions, both eciML1515 and iML1515 models accurately predicted fluxes to zero for formate, ethanol, and succinate, consistently with experimental data. Additionally, the biomass fluxes computed by models closely matched the measured flux. The most notable distinction involves the acetate flux: while eciML1515 computed an acetate flux equal to 5.64 mmol gDW^{-1} h^{-1} matching the measured flux of 5.7 mmol gDW^{-1} h^{-1} iML1515 erroneously predicted zero acetate production. This discrepancy appears to be essential to achieve the elevated eciML151 accuracy. On the other hand, under anoxic conditions, the two models tend to overestimate the fluxes referred to acetate, formate, and ethanol production compared to the corresponding experimental data. However, the fluxes predicted by eciML1515 are more in agreement to the measurements.

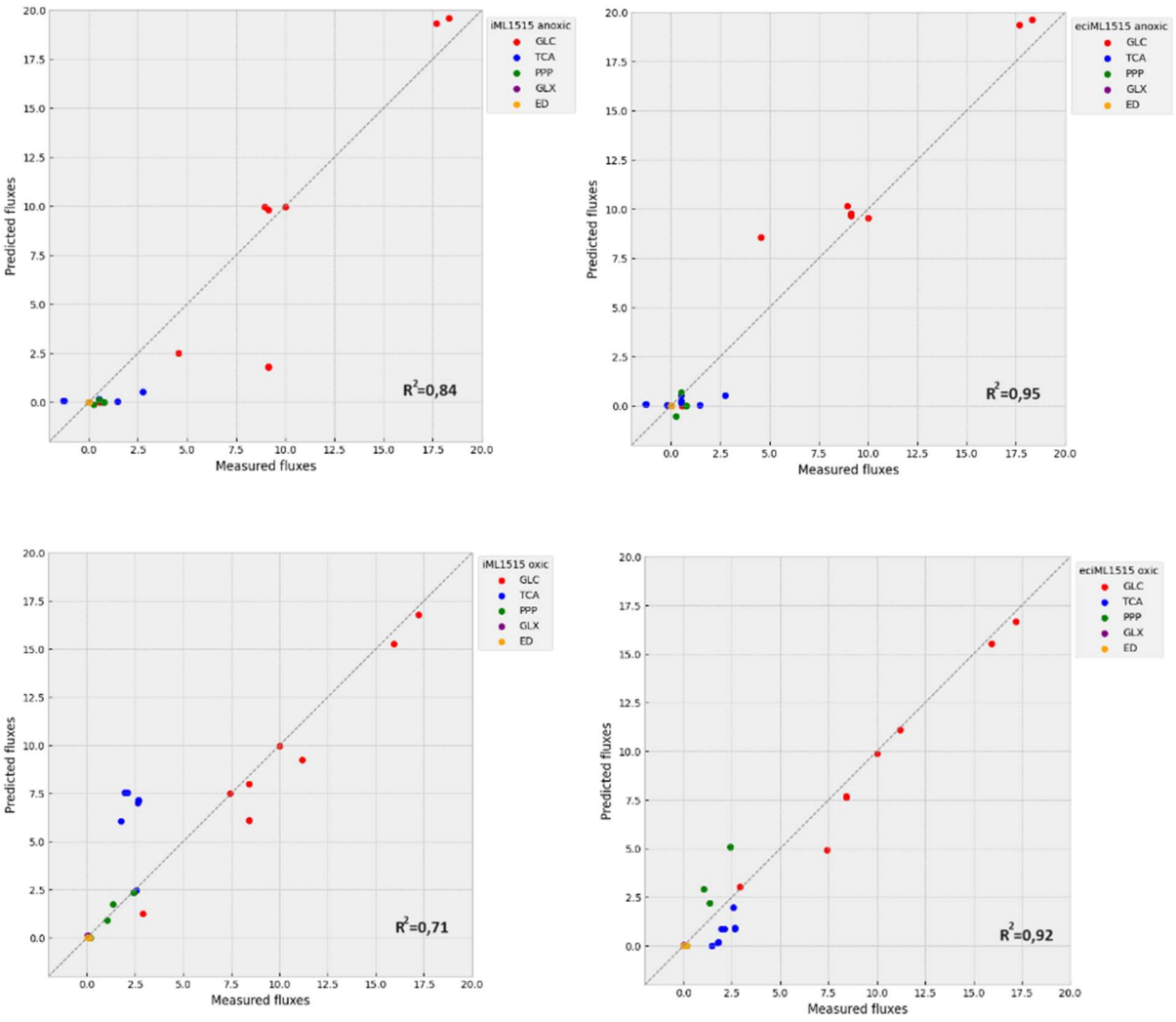

Figure 6 – Scatter plots of Intracellular fluxes under anoxic and oxic conditions.

The comparison of predicted and measured intracellular fluxes at the exponential growth phase suggests that eciML1515 is characterized by an overall higher accuracy than iML1515, in accordance with the previous evaluation of the exchange fluxes accuracy. Initially, the R^2 computed for the iML1515 predictions were characterized by remarkably negative values both under oxic and anoxic conditions as shown in Table A5 of the appendix, suggesting a complete inability of iML1515 of simulating the metabolic fluxes of interest. However, the fluxes predicted by iML1515 did not display notable discrepancies compared to the data, apart from the two fluxes SUCOAS and SUCDi. Thus, these two fluxes were removed from the dataset and the R^2 were recalculated. The new R^2 scores provided a more realistic indication of

iML1515 accuracy. A close observation of scatter plots suggests that intracellular fluxes involved in different metabolic pathways are characterized by different accuracy levels.

Under oxic conditions, the two models predicted the fluxes within the glycolytic pathway in a similar way, that is, underestimating the values compared to the experimental dataset. As for the pentose phosphate fluxes, they were better predicted by iML1515 compared to eciML1515 predictions that appear overestimated in relation to the measured fluxes. Under anoxic conditions, the accuracy related to the glycolytic fluxes was overall lower, especially for iML1515, while the predictions referred to the TCA cycle and pentose phosphate pathway appear similar across the two models.

Figure 8 – Comparison of predicted and measured intracellular fluxes at the exponential growth phase under anoxic and oxic conditions.

		Anoxic conditions			Oxic conditions		
Pathways	ID	Data	iML1515	eciML1515	Data	iML1515	eciML1515
Glycolitic	GLCptspp	*13,1 ± 0,04*	*13,03 ± 0,05*	*12,87 ± 1,32*	*8,8 ± 0,02*	*8,69 ± 0,06*	*8,78 ± 0,34*
Glycolitic	PGI	*11,69 ± 0,24*	*12,84 ± 0,09*	*12,93 ± 0,11*	*6,53 ± 0,15*	*7,04 ± 0,06*	*4,47 ± 0,03*
Glycolitic	PFK	*11,98 ± 0,27*	*4,52 ± 0,66*	*9,5 ± 1,26*	*7,41 ± 0,11*	*4,54 ± 0,25*	*6,71 ± 0,02*
Glycolitic	FBA	*11,98 ± 0,27*	*4,51 ± 0,66*	*9,88 ± 0,92*	*7,41 ± 0,11*	*4,54 ± 0,26*	*6,79 ± 0,02*
Glycolitic	TPI	*11,98 ± 0,27*	*12,65 ± 0,13*	*10,18 ± 1,23*	*7,41 ± 0,11*	*8,26 ± 0,07*	*6,72 ± 0*
Glycolitic	GAPD	*24,01 ± 0,5*	*25,4 ± 0,09*	*22,48 ± 1,18*	*15,14 ± 0,18*	*15,21 ± 0,02*	*14,64 ± 0*
Glycolitic	ENO	*23,14 ± 0,71*	*25,11 ± 0,08*	*22,28 ± 1,14*	*14,01 ± 0,26*	*13,89 ± 0,02*	*13,57 ± 0*
Glycolitic	PYK	*5,96 ± 1,4*	*8,98 ± 0,39*	*7,69 ± 1,11*	*2,56 ± 0,57*	*1,76 ± 0,04*	*1,71 ± 0,08*
Glycolitic	PDH	*0,73 ± 1,98*	*0,12 ± 0,11*	*0,11 ± 0,01*	*9,85 ± 0,54*	*8,9 ± 0,04*	*9,86 ± 0,02*
Glycolitic	PFL	*16,86 ± 2,94*	*23,27 ± 0,5*	*20,15 ± 0,81*	-	*0,04 ± 0,01*	*0 ± 0*
TCA cycle	CS	*0,7 ± 0,18*	*0,21 ± 0,03*	*0,92 ± 0,07*	*2,34 ± 0,43*	*6,48 ± 0,02*	*1,77 ± 0,01*
TCA cycle	ACONTa	*0,7 ± 0,18*	*0,19 ± 0,03*	*0,93 ± 0,06*	*2,34 ± 0,43*	*6,47 ± 0,02*	*1,76 ± 0*
TCA cycle	ACONTb	*0,7 ± 0,18*	*0,19 ± 0,03*	*0,94 ± 0,05*	*2,34 ± 0,43*	*6,47 ± 0,02*	*1,76 ± 0*
TCA cycle	ICDHyr	*0,68 ± 0,19*	*0,18 ± 0,01*	*0,91 ± 0,08*	*2,33 ± 0,56*	*6,42 ± 0,04*	*1,32 ± 0,02*
TCA cycle	AKGDH	*0,01 ± 0,05*	*0 ± 0*	*0 ± 0*	*1,56 ± 0,58*	*5,59 ± 0,04*	*0,64 ± 0,02*
TCA cycle	SUCOAS	*-0,25 ± 0,09*	*206,12 ± 73,19*	*0,08 ± 0,02*	*1,27 ± 0,59*	*494,82 ± 36,96*	*0 ± 0*
TCA cycle	SUCDi	*-1,92 ± 0,71*	*0 ± 0*	*0 ± 0*	*1,58 ± 0,44*	*30,06 ± 0,87*	*1,12 ± 0,01*
TCA cycle	FUM	*-1,69 ± 0,73*	*0,02 ± 0,06*	*0,06 ± 0*	*1,85 ± 0,44*	*6,51 ± 0,03*	*1,74 ± 0*
TCA cycle	MDH	*-1,67 ± 0,86*	*0 ± 0,07*	*0,03 ± 0,02*	*1,71 ± 0,48*	*6,55 ± 0,03*	*2,19 ± 0,02*
TCA cycle	PPC	*3,62 ± 0,88*	*0,76 ± 0,17*	*1,36 ± 0,03*	*2,26 ± 0,32*	*2,23 ± 0,03*	*1,39 ± 0,02*
PPP	PGL	*1,03 ± 0,18*	*0,17 ± 0,1*	*0,3 ± 0,02*	*2,14 ± 0,15*	*1,64 ± 0,04*	*4,37 ± 0,01*
PPP	GND	*0,99 ± 0,21*	*0,03 ± 0,01*	*0,07 ± 0,01*	*2,11 ± 0,16*	*1,63 ± 0,04*	*4,37 ± 0,01*
PPP	RPI	*0,66 ± 0,12*	*0,2 ± 0,03*	*25,33 ± 13,49*	*1,18 ± 0,07*	*-1,11 ± 0,01*	*1,93 ± 0,01*
PPP	RPE	*0,33 ± 0,14*	*-0,17 ± 0,03*	*0,36 ± 0,07*	*0,93 ± 0,11*	*0,49 ± 0,03*	*2,46 ± 0,03*
Exchange	EX_for_e	*16,38 ± 3,28*	*23,19 ± 0,51*	*20,11 ± 0,82*	-	*0 ± 0*	*0 ± 0*
Exchange	EX_etoh_e	*7,97 ± 1,71*	*10,48 ± 0,74*	*10,51 ± 0,69*	-	*0 ± 0*	*0 ± 0*
Exchange	EX_ac_e	*7,32 ± 1,64*	*11,22 ± 0,46*	*8,24 ± 0,66*	*5,69 ± 0,73*	*0 ± 0*	*5,65 ± 0,01*

Table 8 compares the predicted to the measured fluxes reported in the Exponential growth phase dataset. The results indicate that under anoxic conditions, the two models' predicted fluxes were congruent with the data, especially related to the glycolytic pathway. Similarly, the predicted fluxes concerning the pentose phosphate pathway also appeared in agreement to the data, even though both models tended to underestimate the experimental values. Besides, the fluxes referred to the TCA cycle present some discrepancies. Most importantly, the experimental fluxes of SUCDi SUCOAS, FUM and MDH suggested that, under the considered conditions, the reactions ran in the reverse reaction direction. However, this behaviour was not properly described by either eciML1515 or iML1515, which returned fluxes values relative to the forward reaction close to zero.

Under oxic conditions, the fluxes of the TCA cycle predicted by eciML1515 present similar values and are all located below 2 mmol gDW^{-1} h^{-1} in agreement with the experimental data, which are settled near 2.5 mmol gDW^{-1} h^{-1}. The eciML1515 models predicted the acetate exchange reaction to present a flux of 5.6 mmol gDW^{-1} h^{-1}, notably similar to the actual flux that was found to range at 5.69 mmol gDW^{-1} h^{-1} in the exponential growth phase dataset. Simultaneously, the pentose phosphate fluxes presented remarkably higher fluxes compared to the dataset. In contrast, under the same conditions, iML1515 predicts the TCA cycle fluxes to be notably high, thus bringing the acetate production down to zero. Lastly, in this model the pentose phosphate fluxes are quite consistent to the data, as opposed to the eciML1515.

Noteworthy, iML1515 predict extremely elevated fluxes for SUCOAS and SUCDi with values of 494.82 mmol gDW^{-1} h^{-1} and 30.6 mmol gDW^{-1} h^{-1}, under oxic conditions, and only related to SUCOAS with a value of 206.12 mmol gDW^{-1} $h^{-}1$ under anoxic conditions, suggesting the plausible presence of a looping reaction. These anomalous flux values are not observed with eciML1515.

3.4.3 Batch simulations

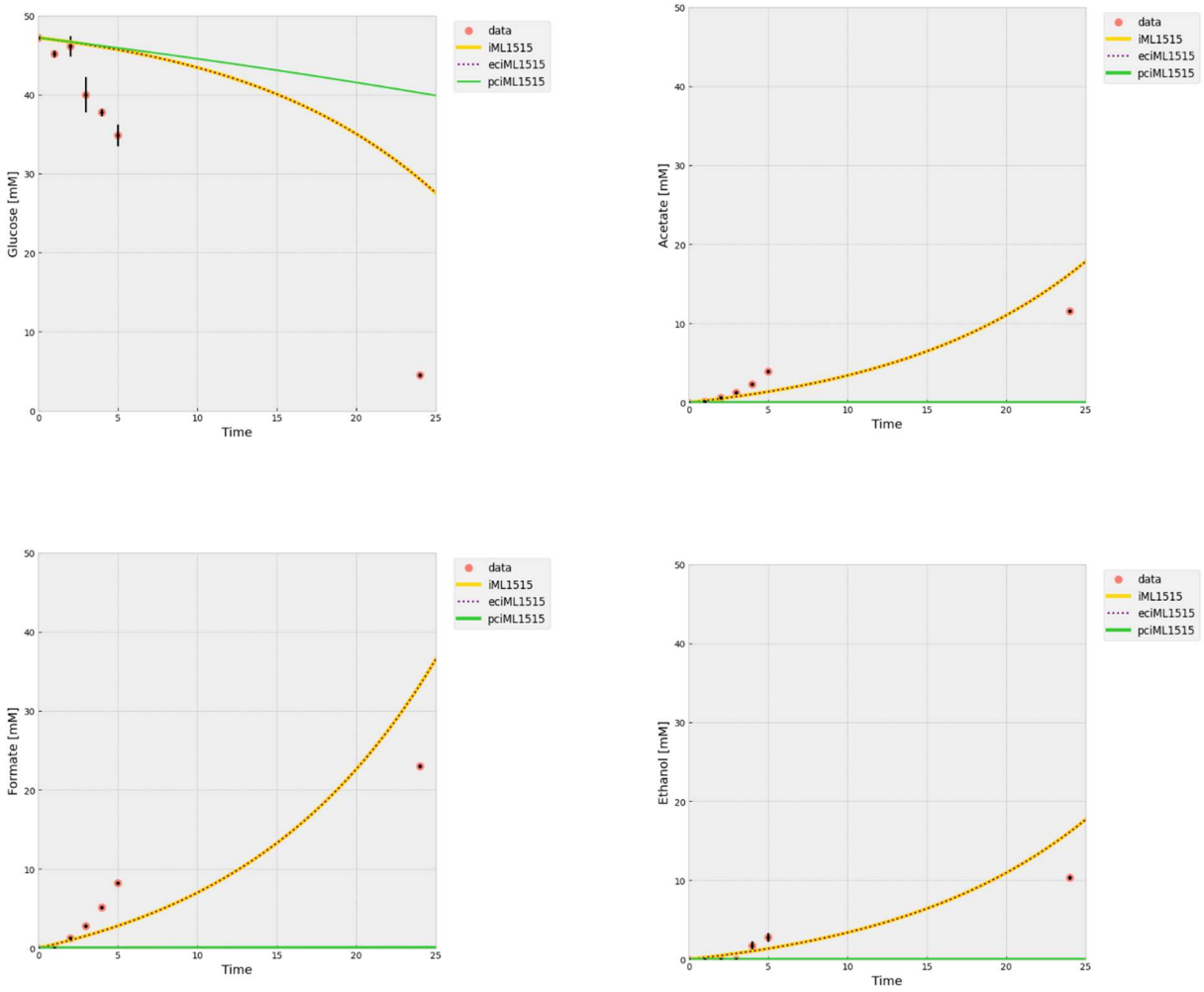

Figure 7 – Batch simulation under anoxic conditions (glucose,formate, ethanol, acetate)

The plots in figures 7 and 8 represent the comparison between the predicted extracellular concentrations of glucose and the primary fermentative concentration of the measured data contained in the Batch dataset. Specifically, figure 7 includes glucose together with acetate, formate and ethanol; while figure 8 shows succinate, lactate and pyruvate. The predictions were obtained by iML1515, eciML1515 and pciML1515 setting the computational process settings to specifically replicate the Batch dataset.

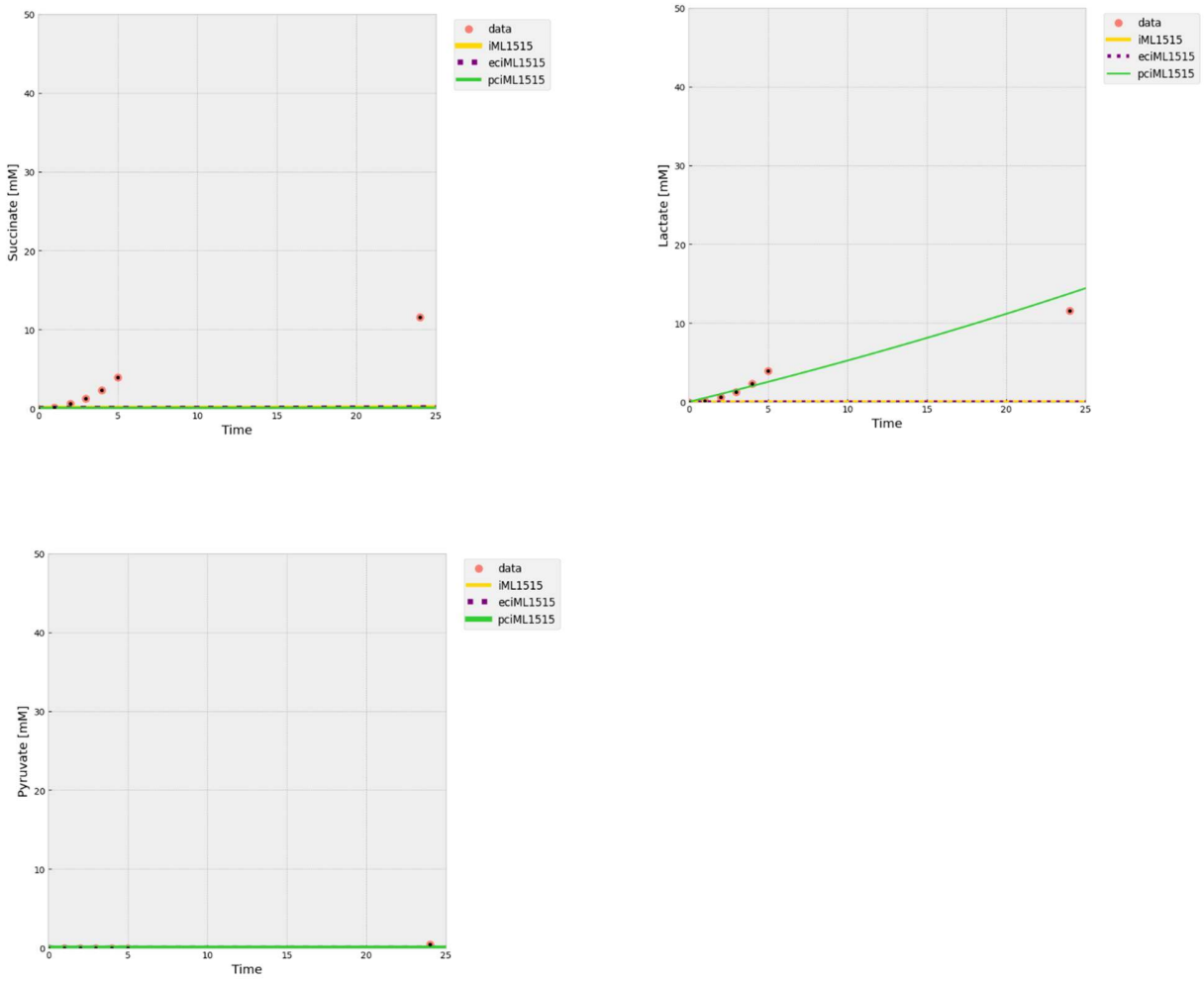

Figure 8 – Batch simulation under anoxic conditions (pyruvate, succinate, lactate)

The plots show that the eciML1515 and iML1515 returned overlapping curves, suggesting that under the simulated conditions, the enzyme constraint of eciML1515 did not have an evident effect on the predictions. Overall, the curves indicate that none of the models returned accurate predictions; moreover, pciML1515 provided the most discordant results, as well as the worst predictive accuracy. This model, indeed, did not predict the production of acetate, ethanol, succinate, or formate, which differs from eciML1515 and iML1515. Contrarily, pciML1515 predicted the production of lactate. The curve of its predicted concentration did not match the data trend, but the concentration of lactate at the end of the predicted batch process agreed with the experimental data. The eciML1515 and iML1515 models computed the production of ethanol, acetate and formate; however, their predicted concentrations did not match the data trend. This suggests that the concentration of metabolites grew faster than the data,

reaching concentrations considerably higher compared to the measured concentrations at the end of the simulation. Similarly, the glucose consumption was not properly described by any of the models that simulated this process in a slower way, and none of the considered models was able to predict the succinate production included in the Batch dataset.

4. Discussion

Genome-metabolic models are commonly employed to guide metabolic engineering approaches as they simulate metabolic phenotypes in silico according to specific growth conditions or genetic manipulation. This aspect renders GEMs powerful computational tools that considerably speed up the design of optimised organisms, especially for industrial applications (Moreno-Paz et al., 2022; Yang et al., 2021)The restricted predictive accuracy, confirmed by several published studies, represents the primary limitation of GEMs (Grigaitis et al., 2021). Various studies have led to the construction of extended GEMs that include additional factors essential to metabolism, such as the enzymatic abundance available in the cells. The introduction of extended GEMs, especially enzyme-constrained models, has markedly advanced constraint-based modelling(Grigaitis et al., 2021; Mao et al., 2022; Sánchez et al., 2017).

However, diverse extended models are typically validated in oxic conditions, making an adequate comparison difficult. This thesis examined the predictive performance of three metabolic models of *E. coli*. The main focus was on the comparison between the conventional GEM, iML1515, and the two enzyme constraint models, eciML1515 and pciML1515. In the following section, the predictive accuracy of each model under different growth conditions will be addressed, especially focusing on the models' capability of predicting the overflow metabolism of *E. coli*.

Initially, the models were analysed with FVA under unlimited glucose availability. This unconventional procedure was explicitly applied to evaluate the limiting factor in the models. The FVA results highlighted that iML1515 is only constrained by the stoichiometry and boundaries of reactions, while in eciML1515 and pciML1515, the limited enzyme availability is the constraining factor. The careful evaluation of fluxes' ranges may suggest that eciML1515 is able to predict realistic fluxes similar to the fluxes' intervals predicted setting the glucose uptake to 10 mmol gDW^{-1} h^{-1} to simulate physiological glucose consumption. Contrarily, the ranges of fluxes returned by pciML1515 do not align as expected with the two growth conditions simulated. The fluxes associated with the fermentative products are similar under oxic and anoxic conditions. Furthermore, the exchange lactate, PYK, ACONTa, and CS fluxes are associated with broad ranges incompatible with a realistic situation. It should be noted that the FVA performed under unlimited glucose uptake conditions does not return realistic outputs, which are represented by considerably large flux ranges. Moreover, it was striking to see that the minimal flux through the acetate exchange reaction was high (>300 units) for iML1515, while for all other simulated conditions this model was not able to simulate overflow metabolism.

The fact that it was not possible to obtain the results of the pciML1515 sampling process prevents the in-depth comparison of the model's accuracy with the one of iML1515 and eciML1515. Furthermore, it may suggest complications in performing pciML1515 simulations. However, since it was possible to run an FVA, the models were compared considering this output. Here, it was seen that the pciML1515 predicted

notably lower biomass fluxes compared to the other two models, both under oxic and anoxic conditions. This discrepancy appears remarkable especially under anoxic conditions with a predicted value of 0.03 mmol gDW^{-1} h^{-1}, suggesting the presence of a possible bottleneck.

On another note, the comparison of models' predictions to the Chemostat dataset indicates that eciML1515 and iML1515 display a comparable accuracy both under oxic and anoxic conditions. However, the conducted analyses do not conclusively assess the models' accuracy because of the absence of data in the Chemostat dataset concerning the overflow metabolism under oxic conditions. The ability to predict the acetate production is essential, being one of the primary limitations of conventional GEMs.

Differently, the comparison against the Exponential growth phase, shows that eciML1515 is capable of predicting the overflow metabolism, in agreement with the previous findings of Mao et al., 2022. Furthermore, the enzyme constrained model presents a greater accuracy in the prediction of intracellular and exchange fluxes with R^2 of 0.9. The model also computed fermentative and biomass fluxes conform to the experimental dataset.

A detailed observation of intracellular fluxes highlighted how iML1515 and eciML1515 predict the metabolic fluxes belonging to different pathways with different accuracy levels. Indeed, under oxic conditions, eciML1515 computed the fluxes of pentose phosphate pathway less accurately compared to iML1515. As for the TCA cycle and acetate fluxes, eciML1515 presents these in accordance with the data, while the predictions made by iML1515 are respectively overestimated and zero. This result suggests that the models distribute the carbon flux differently across the pathway. IM1515, in fact, drives the carbon flux mainly toward the TCA cycle and not to the acetate producing reactions. Differently, eciML1515 directs the carbon flux partially toward the acetate production, predicting the switch of metabolism due to the limited availability of enzymes involved in the TCA cycle that push the model to activate alternative metabolic pathways. Indeed, the TCA cycle requires a larger number of proteins compared to the acetate production.

Noteworthy, the SUCOAS and SUCDi fluxes predicted by iML1515 present extremely elevated values, probably caused by looping reactions not constrained in the conventional GEMs but limited by the enzyme constraint in eciML1515. To confirm this hypothesis and probably improve the iMl1515' accuracy, parsimonious FBA (*pFBA*) can be performed, permitting the computation of the flux distribution without any futile cycle.

Subsequently, the comparison of batch simulations to the Batch dataset highlighted that none of the three models returned predictions congruent to the experimental data. The predictions of eciML1515 and iML1515 appeared extremely similar, as suggested by the fact that the computed concentrations overlapped. This suggests that under the simulated conditions the enzyme constraint did not impact the predictions of eciML1515. However, the two models were able to predict the production of acetate, ethanol and succinate. The pciML151 displayed the lower predictive performance in replicating the

measured data. Indeed, the pciML1515 only predicted the lactate production, while missing in the other two models.

The evaluation of models undertaken in this project would be persecuted, focusing on the evaluation of model accuracy when the chemostat growth process is simulated. The performed comparison led to inconclusive results, primarily due to the impossibility of comparing the models' ability to predict the overflow metabolism of *E. coli*. The identification of further datasets, including data related to acetate production under oxic conditions, could permit the repeat of the analyses, potentially leading to a valid evaluation of the accuracy of models simulated to replicate the chemostat mode.

Furthermore, the three models would be integrated with the adjustment proposed by Bernsteinet al., 2023, which was observed to lead to an increase in accuracy. The adjustment involved the addition of vitamins and cofactors during the model's simulations, correcting the reversibility of specific reactions and mapping isoenzymes.

According to the results, eciML1515 displays an increased accuracy compared to iML1515 and pciML1515. However, under certain circumstances, the model shows a decreased accuracy. A possible perspective is a further improvement of the model, modifying the model parameters related to the enzyme constraint, such as kinetics parameters or providing the model with proteomics data collected under diverse growth conditions.

5. Conclusion

The benchmarking progress conducted in this project highlighted that the enzyme constraint ecML1515 shows a higher predictive accuracy compared to the conventional GEM iML1515, especially under oxic conditions, ascribed to its capability of predicting overflow metabolism. The accuracy comparison of eciML1515 and pciML1515 was not performed adequately as a result of the impossibility of sampling pciML1515. However, preliminary analyses suggested that pciML1515 returns anomalous flux predictions, indicating that eciML1515 might display the most elevated predictive accuracy across the tested models.

Appendix

Table. A1 Comparison of exchange fluxes in chemostat process under anoxic and oxic conditions.

	Oxic conditions			Anoxic conditions		
ID	Data	eciML1515	iML1515	Data	eciML1515	iML1515
Glucose uptake	*3,2±0,01*	*2,52±0*	*2,5±0*	*10,9±0,02*	*12,18±0*	*11,96±3,91E-14*
Acetate	*0,6±0*	*0,017±4,83E-03*	*5,25E-04±2,98E-04*	*9,39±0,06*	*10,65±0,01*	*10,468±0,026*
Formate	*0±0*	*2,94E-03±1,03E-03*	*2,29E-03±1,75E-03*	*17,76±0,07*	*22,055±0,016*	*21,610±0,031*
Ethanol	*0±0*	*5,61E-16±1,88E-16*	*2,52E-04±2,46E-04*	*7,37±0,07*	*10,540±0,012*	*10,332±0,020*
Lactate	*0±0*	*1,62E-05±1,58E-05*	*0±0*	*0,19±0*	*8,36E-04±6,75E-04*	*0,011±0,007*
Succinate	*0±0*	*6,64E-06±4,02E-06*	*3,78E-04±2,17E-04*	*0,83±0,01*	*0,080±0,006*	*0,089±0,009*
Biomass	*0,2±0*	*0,194±1,48E-05*	*0,2±0*	*0,2±0*	*0,20±0*	*0,2±0*

Table.A1 above shows the values of exchange fluxes under oxic and anoxic conditions. The experimental data contained in the Chemostat datataset are associated with the experimental error. The predictions obtained from the models are provided with the mediam absolute deviation calculated during the sampling process.

Table. A2

R2_Intra_flux_Chemostat_			
ae_iML	an_iML	ae_eci	an_eci
-169.43	-17.87	0.88	0.83
0.82	0.88	0.89	0.83

Table.A3

R2_exch_flux_cehmostat			
ae_iML	an_iML	ae_eci	an_eci
0.9	0.9	0.9	0.88

Table.A2 Display the R^2 computed for the intracellular fluxes in chemostat. The first row contained the values obtained before the SUCOAS and SUCDi remotion while the bottom row contained the values computed after the remotion of those fluxes. Table- A3 reports the R2 associated to the exchange fluxes in chemostat.

Table. A4 Comparison of exchange fluxes at exponential growth phase under anoxic and oxic conditions.

	Oxic conditions			Anoxic conditions		
ID	Data	eciML1515	iML1515	Data	eciML1515	iML1515
Glucose uptake	*8,8±0,5*	*8,8±0*	*8,8±0*	*13,1±1*	*13,1±0*	*13,1±0*
Acetate	*5,7±0,2*	*5,646±0,011*	*2,34E-03±1,47E-03*	*7,3±0,7*	*8,24±0,66*	*11,22±0,46*
Formate	*0±0*	*1,62E-03±1,32E-03*	*3,78E-03±1,96E-03*	*16,4±0,9*	*20,11±0,82*	*23,19±0,51*
Ethanol	*0±0*	*3,23E-23±4,03E-24*	*6,79E-04±6,17E-04*	*8±0,5*	*10,51±0,69*	*10,48±0,74*
Succinate	*0±0*	*8,56E-04±6,05E-04*	*3,65E-05±2,79E-05*	*0,1±0,01*	*0,133±0,013*	*0,242±0,069*
Biomass	*0,7±0,01*	*0,62±1,15E-05*	*0,75E-05±1,13E-05*	*0,33±0,02*	*0,126±9,07E-06*	*0,129±7,04E-06*

Table.A4 above shows the values of exchange fluxes under oxic and anoxic conditions. The experimental data contained in the exponential growth phased datataset are associated with the experimental error. The predictions obtained from the models are provided with the median absolute deviation calculated during the sampling process.

Table.A5 R^2 scores exp_growth_phase

R2_Intra_flux_exp_growth_			
ae_iML	an_iML	ae_eci	an_eci
-136.11	-113.57	0.92	0.96
0.71	0.84	0.93	0.95

Table.A5 R^2 scores exp_growth_phase

R2_exch_flux_exp_growth_			
ae_iML	an_iML	ae_eci	an_eci
0.55	0.69	0.99	0.9

Table.A5 Display the R^2 computed for the intracellular fluxes in chemostat. The first row contained the values obtained before the SUCOAS and SUCDi remotion while the bottom row contained the values computed after the remotion of those fluxes. Table- A6 reports the R^2 associated to the exchange fluxes in chemostat.

Bibliography

Beg, Q. K., Vazquez, A., Ernst, J., de Menezes, M. A., Bar-Joseph, Z., Barabási, A.-L., & Oltvai, Z. N. (2007). 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. https://doi.org/10.1073/pnas.0609845104

Bekiaris, P. S., & Klamt, S. (2020). Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, *21*(1), 19. https://doi.org/10.1186/s12859-019-3329-9

Bernstein, D. B., Akkas, B., Price, M. N., & Arkin, A. P. (2023). Evaluating *E. coli* genome-scale metabolic model accuracy with high-throughput mutant fitness data. *Molecular Systems Biology*, *19*(12). https://doi.org/10.15252/msb.202311566

Crown, S. B., & Antoniewicz, M. R. (2013). Publishing 13C metabolic flux analysis studies: A review and future perspectives. *Metabolic Engineering*, *20*, 42–48. https://doi.org/10.1016/j.ymben.2013.08.005

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*(1), 74. https://doi.org/10.1186/1752-0509-7-74

Gonzalez, J. E., Long, C. P., & Antoniewicz, M. R. (2017). Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by 13C metabolic flux analysis. *Metabolic Engineering*, *39*, 9–18. https://doi.org/10.1016/j.ymben.2016.11.003

Grigaitis, P., Olivier, B. G., Fiedler, T., Teusink, B., Kummer, U., & Veith, N. (2021). Protein cost allocation explains metabolic strategies in Escherichia coli. *Journal of Biotechnology*, *327*, 54–63. https://doi.org/10.1016/j.jbiotec.2020.11.003

Herrmann, H. A., Dyson, B. C., Vass, L., Johnson, G. N., & Schwartz, J.-M. (2019). Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *Npj Systems Biology and Applications*, *5*(1), 32. https://doi.org/10.1038/s41540-019-0109-0

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–D522. https://doi.org/10.1093/nar/gkv1049

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, *83*(3), 1331–1340. https://doi.org/10.1016/S0006-3495(02)73903-9

Mao, Z., Zhao, X., Yang, X., Zhang, P., Du, J., Yuan, Q., & Ma, H. (2022). ECMpy, a Simplified Workflow for Constructing Enzymatic Constrained Metabolic Network Model. *Biomolecules*, *12*(1), 65. https://doi.org/10.3390/biom12010065

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M., & Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology*, *35*(10), 904–908. https://doi.org/10.1038/nbt.3956

Moreno-Paz, S., Schmitz, J., Martins dos Santos, V. A. P., & Suarez-Diez, M. (2022). Enzyme-constrained models predict the dynamics of *Saccharomyces cerevisiae* growth in continuous, batch and fed-batch bioreactors. *Microbial Biotechnology*, *15*(5), 1434–1445. https://doi.org/10.1111/1751-7915.13995

Müller, A. C., & Bockmayr, A. (2013). Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, *29*(7), 903–909. https://doi.org/10.1093/bioinformatics/btt059

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

Saa, P. A., & Nielsen, L. K. (2016). ll-ACHRB: a scalable algorithm for sampling the feasible solution space of metabolic networks. *Bioinformatics*, *32*(15), 2330–2337. https://doi.org/10.1093/bioinformatics/btw132

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, *13*(8). https://doi.org/10.15252/msb.20167411

Sara Moreno Paz. (2019). *Metabolic and reactor modelling for production of human milk oligosaccharides by E. coli*.

Toya, Y., Nakahigashi, K., Tomita, M., & Shimizu, K. (2012). Metabolic regulation analysis of wild-type and arcA mutant Escherichia coli under nitrate conditions using different levels of omics data. *Molecular BioSystems*, *8*(10), 2593. https://doi.org/10.1039/c2mb25069a

Van Rossum, G. and D. F. L. (2009). *Python 3 Reference Manual*.

Yang, X., Mao, Z., Zhao, X., Wang, R., Zhang, P., Cai, J., Xue, C., & Ma, H. (2021). Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metabolic Engineering*, *67*, 133–144. https://doi.org/10.1016/j.ymben.2021.06.005