Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

Article

ECMpy, a simplified workflow for constructing enzymatic constrained metabolic network model

Zhitao Mao[1, †], Xin Zhao[1, †], Xue Yang[1], Peiji Zhang[1], Jiawei Du[1], Qianqian Yuan[1], and Hongwu Ma[1, *]

[1] Biodesign Center, Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China
* Correspondence: ma_hw@tib.cas.cn
† These authors equally contributed to this work as co-first.

Abstract:

Genome-scale metabolic models (GEMs) have been widely used for phenotypic prediction of microorganisms. However, the lack of other constraints in the stoichiometric model often leads to a large metabolic solution space inaccessible. Inspired by previous studies that take allocation of macromolecule resources into account, we developed a simplified Python-based workflow for constructing enzymatic constrained metabolic network model (ECMpy) and constructed an enzyme-constrained model for *Escherichia coli* (ec*i*ML1515) by directly adding a total enzyme amount constraint in the latest version of GEM for *E. coli* (*i*ML1515), considering the protein subunit composition in the reaction, and automated calibration of enzyme kinetic parameters. Using ec*i*ML1515, we predicted the overflow metabolism of *E. coli* and revealed that redox balance was the key reason for the difference between *E. coli* and *Saccharomyces cerevisiae* in overflow metabolism. The growth rate predictions on 24 single-carbon sources were improved significantly when compared with other enzyme-constrained models of *E. coli*. Finally, we revealed the tradeoff between enzyme usage efficiency and biomass yield by exploring the metabolic behaviors under different substrate consumption rates. Enzyme-constrained models can improve simulation accuracy and thus can predict cellular phenotypes under various genetic perturbations more precisely, providing reliable guidance for metabolic engineering.

Keywords: Enzyme-constrained model; *Escherichia coli*; Enzyme kinetics; Protein subunit; Overflow metabolism;

1. Introduction

Accurate prediction of metabolic phenotypes of an organism is a key goal of computational biology and has attracted more and more attention from researchers. For this purpose, many genome-scale metabolic models have been developed [1, 2] and successfully applied for guiding metabolic engineering based on flux balance analysis (FBA) and other stoichiometry-based methods [3, 4]. However, in many cases, a microorganism shows suboptimal metabolism [5, 6] that is inconsistent with the optimal solution of FBA [7], implying that the metabolic capacity of an organism is also constrained by other factors. For example, overflow metabolism, involving incomplete oxidation of glucose to fermentation byproducts such as acetate and ethanol instead of using respiratory pathway even in the presence of oxygen [8] cannot be properly explained by models only considering reaction stoichiometries. Studies suggested that it is likely to be caused by the limited amount of protein molecules within the cell [9].

In recent years, researchers proposed several new methods that introduced new constraints such as cell volume limitation [10], enzyme activity and total protein mass [11, 12], thermodynamics [13] into the model along with the stoichiometric constraints. FBA with Molecular Crowding (FBAwMC) [10] introduced both the crowding coefficient and cell volume constraint to limit the space occupied by enzymes. With the new constraints,

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

the method successfully simulated the substrate hierarchy utilization in *E. coli* [10]. Adadi et al. further extended FBAwMC by introducing known enzyme kinetic parameters and proposed a new method called MOMENT (Metabolic modeling with enzyme kinetics), which improved the prediction accuracy of intracellular fluxes and enzyme gene expression values [14]. In 2017, Sanchez et al. proposed a new construction workflow of enzyme-constrained model (GECKO, Genome-scale model to account for Enzyme Constraints, using Kinetics and Omics), which used an average enzyme saturation coefficient and determined the fraction of enzyme proteins from proteomics data [15]. They developed an enzyme-constrained model for *S. cerevisiae* using GECKO and made accurate prediction of several metabolic phenotypes [15]. However, introducing the enzyme constraints into the original metabolic model using GECKO needs to be extensively revised by modifying every metabolic reaction with a pseudo-metabolite representing an enzyme and adding hundreds of exchange reactions for enzymes, which is complex and significant increasing the model size. Bekiaris et al. further provided an automatic workflow (AutoPACMEN) for construction of enzyme-constrained model inspired by MOMENT and GECKO, which only introduced one pseudo-reaction and pseudo-metabolite [16]. These two construction processes, GECKO and AutoPACMEN, have greatly facilitated the construction of enzyme-constrained models for each species, and successfully constructed for *S. cerevisiae* [15], *Bacillus subtilis* [17], *Bacillus coagulans* [18], *E. coli* [19] and *Streptomyces coelicolor* [20], which have successfully applied to target prediction for enhancing the yield of products [17, 19, 20].

In current study, we propose a simpler workflow called ECMpy by explicitly introducing an enzyme constraint without modifying existing metabolic reactions or adding new reactions. Using ECMpy workflow, we constructed a high-quality enzyme-constrained model for *E. coli* (eci*i*ML1515) based on its latest metabolic model *i*ML1515 [21], high coverage of enzyme kinetics data gathering from literature [22], and automated enzyme kinetic parameter calibration process. We demonstrated that eci*i*ML1515 could simulate the sub-optimal metabolism such as overflow metabolism and the maximal growth rates under different carbon sources. The whole process for model construction and simulation is available at GitHub (https://github.com/tibbdc/ECMpy) for users to easily reproduce the results and use it as a reference to build enzyme-constrained model for other organisms.

2. Materials and Methods

2.1. The workflow of ECMpy

Metabolic network (like *i*ML1515 model in this study) was used as the initial model for the construction of enzyme-constrained model according to the workflow shown in Figure 1. Firstly, reversible reactions in model were divided into two irreversible reactions because of different k_{cat} values. The stoichiometric constraints (Eq. 1) and reversibility constraints (Eq. 2) used were the same as in flux balance analysis [23]. A new enzymatic constraint (Eq. 3) was introduced into the model, where *ptot* and *f* represent the total protein fraction in *E. coli* and the mass fraction of enzymes, respectively. The enzyme mass fraction *f* was calculated based on Eq. 4 where A_i and A_j represented the abundances (mole ratio) of the i-th protein (*p_num* represented proteins expressed in the model) and j-th protein (*g_num* represented proteins expressed in the whole proteome). MW_i and $k_{cat,i}$ were molecular weight and turnover number of an enzyme catalyzing reaction *i*. For reactions catalyzed by multiple isoenzymes, a reaction can be split into multiple reactions. For reactions catalyzed by enzyme complex, using the minimum value of protein in complex ($\frac{k_{cat,i}}{\mathrm{MW}_i} = min(\frac{k_{cat,ij}}{\mathrm{MW}_{ij}}, j \in m)$, *m* is the number of proteins in complex). σ_i was the saturation coefficient of i-th enzyme.

$$S \cdot v = 0 \tag{1}$$

$$v_{lb} \leq v \leq v_{ub} \tag{2}$$

$$\sum_{i=1}^{n} \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \leq ptot \cdot f \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

$$f = \sum_{i=1}^{p\_num} A_i MW_i \Big/ \sum_{j=1}^{g\_num} A_j MW_j \quad (4)$$

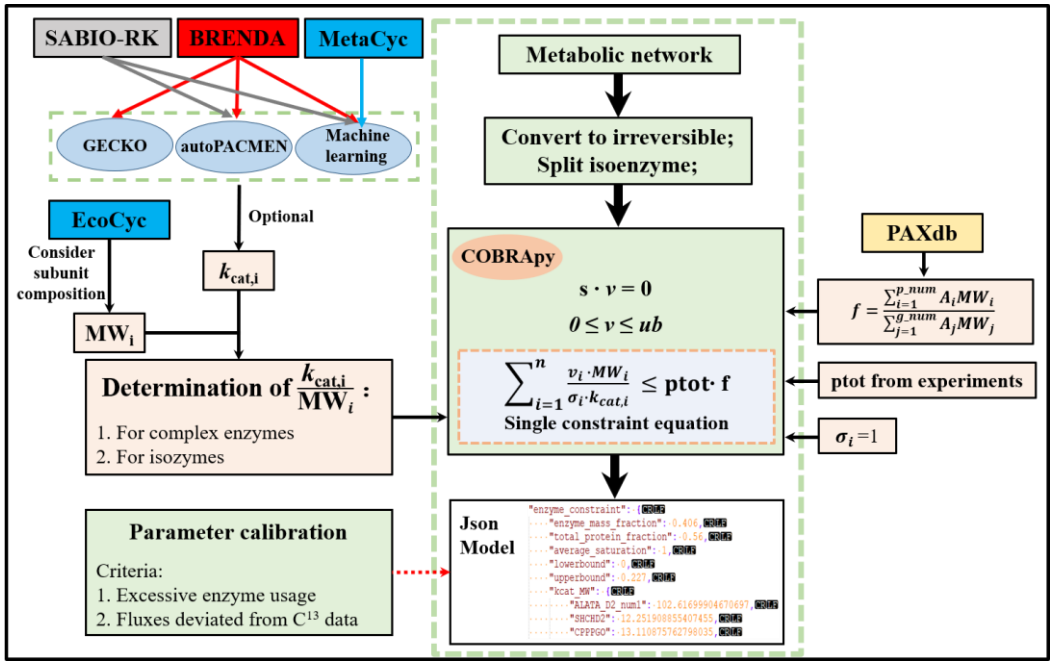

Figure 1. The ECMpy workflow for construction of enzyme-constrained model.

2.2. Calibration of the original kcat values

Generally, enzyme-constrained models need model validation (e.g. adjust the original k_{cat} values to some extent to improve the agreement of model predictions with experimental data), like the way in GECKO and AutoPACMEN [16]. We proposed two principles (enzyme usage and C^{13} flux consistency) to adjust the original k_{cat} values, as follows: First, a reaction with an enzyme usage exceeding 1% of the total enzyme content requires parameter correction; Second, a reaction with the k_{cat} multiplied by 10% of the total enzyme amount ($v_i = \frac{10\% * E_{total} * \sigma_i * k_{cat,i}}{MW_i}$) is less than the flux determined by C^{13} experiment needs to be corrected. All the k_{cat} data used for correction comes from BREDNA and SABIO-RK databases (using the maximum k_{cat} value).

2.3. Simulation

We stored enzyme constraint information and metabolic network into json format, as SBML format cannot save enzyme constraints duo to COBRApy [24] limitations. Then, we directly read json file to get enzyme-constrained model using 'get_enzyme_constraint_model' function written by us. This transformed enzyme-constrained model is consistent with classical constraint-based models in format, which means that functions in COBRApy can be used directly on this model.

To evaluate ec*i*ML1515's ability to predict growth rates, we compared the predicted results of *i*ML1515 and ec*i*ML1515 with experimental results performed by Adadi et al.[14], respectively. Specially, we set the upper bound of substrate uptake rate to 10 mmol/gDW/h and measured *E. coli*'s growth rates on 24 single carbon sources (e.g., acetate, fructose, fumarate and et.al.). For comparison of each methods on 24 single carbon sources, the model and experimental results were used to calculate estimation error of the growth rate (Eq. 5) [25] and normalized flux error (Eq. 6) [26].

$$\text{estimation error} = \frac{\left| v_{growth,sim} - v_{growth,exp} \right|}{v_{growth,exp}} \quad (5)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

$$\text{normalized flux error} = \frac{\sqrt{\sum_i^n (v_{growth,sim_i} - v_{growth,exp_i})^2}}{\sum_i^n (v_{growth,exp_i})^2} \quad (6)$$

In addition to the maximal growth rates under different carbon sources, we also explored the overflow metabolic behaviors of *E. coli*. Specially, the growth rate is fixed (from 0.1 h^{-1} to 0.65 h^{-1}) and glucose is supplied infinitely. Besides, we calculated the reaction enzyme cost (Eq. 7), energy synthesis enzyme cost (Eq. 8) and oxidative phosphorylation ratio (Eq. 9) to explore the adjustment strategy of *E. coli*'s overflow metabolic pathway.

$$\text{reaction enzyme cost}_i = \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \quad (7)$$

$$\text{energy enzyme cost}_i = \sum_{i=1}^{n} \text{reaction enzyme cost}_i \,/ v_{ATP} \quad (8)$$

$$\text{oxidative phosphorylation ratio} = \frac{v_{O_2}}{v_{glucose}} \quad (9)$$

In order to obtain the trade-off between yield ($\frac{v_{biomass}}{v_{glucose} * MW_{glucose}}$) and enzyme usage efficiency ($\frac{v_{biomass}}{E_{min}}$), we developed a new method (Eq. 10-14) to calculate the minimum enzyme amount (E_{min}) inspired by pFBA (Parsimonious FBA) [27]. When simulation, we set the concentration of glucose from 1 mmol/gDW/h to 10 mmol/gDW/h).

$$obj: minimize \sum_{i=1}^{n} \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \quad (10)$$

$$S \cdot v = 0 \quad (11)$$

$$v_{lb} \leq v \leq v_{ub} \quad (12)$$

$$\sum_{i=1}^{n} \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \leq ptot \cdot f \quad (13)$$

$$v_{biomass} = \max (growth\ rate) \quad (14)$$

3. Results

3.1. Construction of enzyme-constrained model of iML1515 by ECMpy

The *i*ML1515 model was used as the initial model for the construction of enzyme-constrained model. During the process we observed that some errors in *i*ML1515 (e.g., GPR relationships, reaction direction and EC number, et al.) and corrected based on information from EcoCyc (See Table S1 for details). Then, we divided reversible reactions in *i*ML1515 into two irreversible reactions and split reactions catalyzed by multiple isoenzymes into different reactions (append num in reaction id, e.g., ALATA_D2_num1). The molecular weights and subunit composition of enzymes in *i*ML1515 were obtained from EcoCyc database [28]. GECKO and sMOMENT (AutoPACMEN for *E. coli*) used the in vitro k_{cat} which obtained in labor-intensive, low-throughput in vitro assays and resulted in only a small fraction of cellular enzymes has a measured k_{cat} even in model organisms [29]. That is why we used the k_{cat} values derived from machine learning methods performed by Heckmann et al. [22]. In the model, k_{cat} values were assigned to 2432 enzymatic reactions, and the coverage exceeds 60% (including isozyme split reaction and reversible split reaction, exclude exchange reaction), which is larger than the GECKO and sMOMENT (the number of reactions that matched EC number and substrate at the same time was only about 387). The protein fraction *ptot* was set at 0.56 g protein /gDW based on experimentally measured macromolecular composition of *E. coli* cells [30, 31]. The *E. coli* protein abundance values were obtained from PAXdb database [32] and the 'Whole organism (Integrated)' dataset with the highest coverage and credibility was selected. According to Eq. 4, *f* was calculated to be 0.406 g enzyme /g protein. However, the flux of growth rate predicted by this initial model is low and the conversion of phosphoenolpyruvate to TCA pathway is abnormal (Figure S1). We first calibrated the reaction according

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

to the enzyme usage, and totally changed 14 reactions (See Table S2 for details). The flux of growth rate predicted by the calibrated model increased to 0.5594 h^{-1}, but the conversion of phosphoenolpyruvate to TCA pathway was still abnormal (Figure S1). Subsequently, we compared with the C^{13} experimental data [33] and found that the k_{cat} value of two reactions (PDH: pyruvate to acetyl-CoA and AKGDH: 2-oxoglutarate to succinyl-CoA) is low, which mainly caused by the subunit composition of these two reactions is complicated and the protein molecular weight is very large. After calibration using C^{13} data (changed 2 reactions, Table S2), the growth rate increased to 0.6802 h^{-1}, and the consistency with the pathway obtained by C^{13} data reached 92.1% (Figure S2). Different from other methods for constructing enzyme-constrained models, our method considers the composition of protein subunits and realizes enzyme constraint by simply adding the total enzyme amount equation (Table 1). Therefore, the enzyme-constrained model we constructed does not change the stoichiometric matrix format (because the isoenzyme reaction and reversible reaction were split, the number of reactions increased), and the solution and subsequent operations of the entire model are consistent with the classical constraint-based model. We used AutoPACMEN to build the GECKO and sMOMENT model of *i*ML1515, and compared them with ECMpy. We found that when considering the subunit composition of protein, the growth rate predicted by GECKO and sMOMENT model is lower, and the flux distribution of the pathway is obviously abnormal from the C^{13} data, especially the TCA pathway (Figure 2).

Table 1. Comparison of the construction methods of enzyme-constrained model

	MOMENT	GECKO	AutoPACMEN	ECMpy
Subunit number	×	√	× (provide interface)	√
Proteomics	×	√	√	√
Saturation	1	0.46	1	1
Initial model	*i*AF1260	Yeast7	*i*JO1366	*i*ML1515
Mass fraction of enzymes	0.56	0.448	0.095	0.227
Total enzyme constraint method	add enzyme concentrations for each reaction and add the enzymes solvent capacity constraint	change stoichiometric matrix, and introduce a large number of pseudo-reaction and pseudo-metabolite	change stoichiometric matrix, and introduce one pseudo-reaction and pseudo-metabolite	only add a total enzyme constraint
Reaction reversibility	not split	split	part split	split
Isozyme	a reaction can be catalyzed by multiple enzymes	a reaction can be catalyzed by multiple enzymes	always assumes that the enzyme with the minimal cost is used	a reaction can be catalyzed by multiple enzymes
If Missing kcat	median turnover number across all reactions	match the kcat value to other substrates, organisms, or even introducing wild cards in the EC number.	similar to GECKO	enzyme cost=0
Model calibration	×	√	√	√
Model type	Not given	XML	XML	Json

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

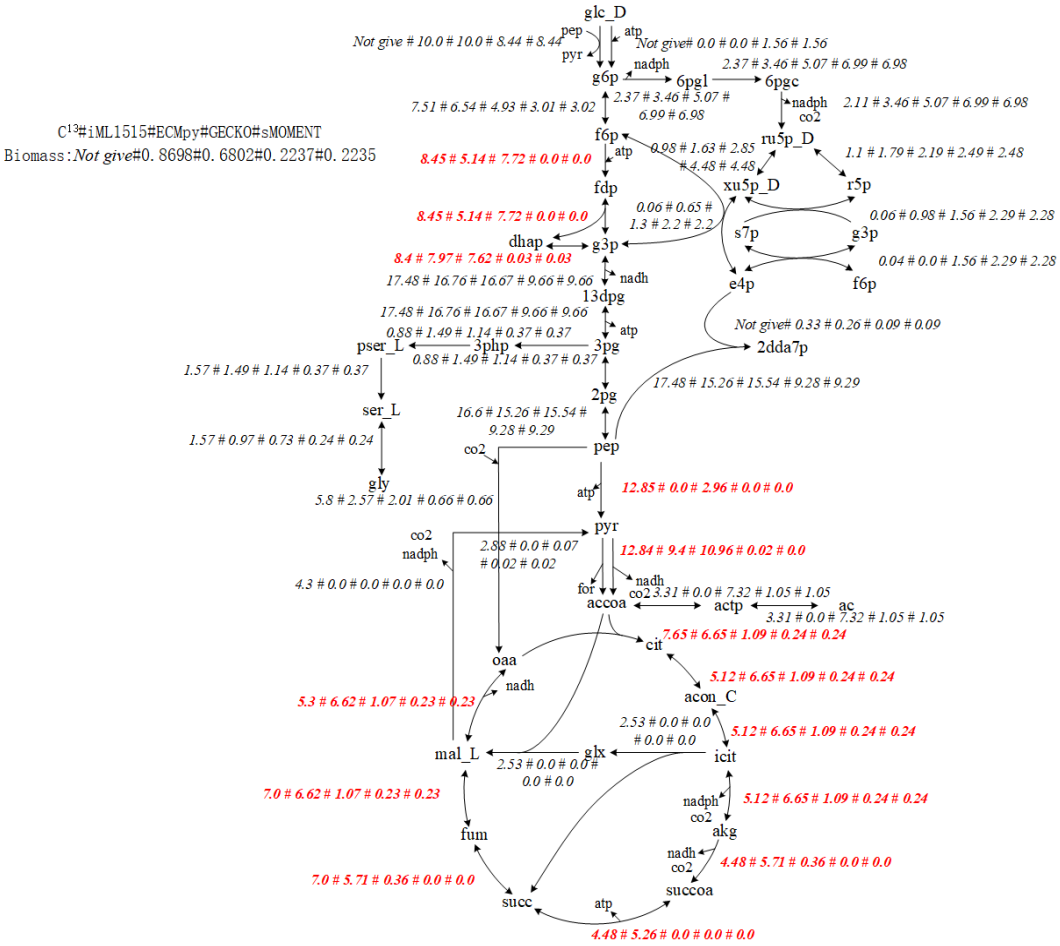

Figure 2. Flux comparison of *i*ML1515, ECMpy, GECKO and sMOMENT. From left to right: C^{13} experimental data, prediction results of *i*ML1515 model, prediction results of ec*i*ML1515 constructed by ECMpy, prediction results of ec*i*ML1515 constructed by GECKO, and prediction results of ec*i*ML1515 constructed by sMOMENT.

3.2. Overflow metabolism of E. coli

Overflow metabolism describes a phenomenon in which cells produce fermentation products even in the presence of oxygen that led to waste of carbon sources [9]. Enzyme-constrained metabolic models have been used to simulate the overflow metabolism in *S. cerevisiae* [15, 34]. To test our model, we applied it to simulate the overflow metabolism reported by literature [35], in which *E. coli* secreted acetate at high growth rates (above 0.5 h^{-1}). As shown in Figure 3a and b, eciML1515 model (the kinetic parameters for each reaction see Table S3) could precisely simulate the switch point where acetate production started. The simulation results indicated that at high growth rates, the acetate producing fermentation pathway was activated due to its low enzyme cost in comparison with the energetically-efficient oxidative respiratory pathway (0.62 g vs 2.38 g enzyme for 1 mol ATP /h, Table S4).

The model also predicted a notable difference in the overflow metabolism between *E. coli* and *S. cerevisiae* **(**Figure 3c**)**. In *S. cerevisiae*, the oxygen-consuming high-yield respiratory pathway was decreased to a very low value [36], whereas in *E. coli* the respiratory pathway was maintained at a high level (Figure 3a) even though the acetate production pathway was activated. A logical explanation for this is that the fermentation products of these two organisms are different. In *S. cerevisiae* ethanol was produced and NADH was balanced in the fermentation pathway. In *E. coli*, however, acetate was produced and the excess NADH produced in the fermentation pathway needs to be balanced through the oxidative respiratory pathway (Figure 3d). This result was in agreement with the finding of a previous study [37].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

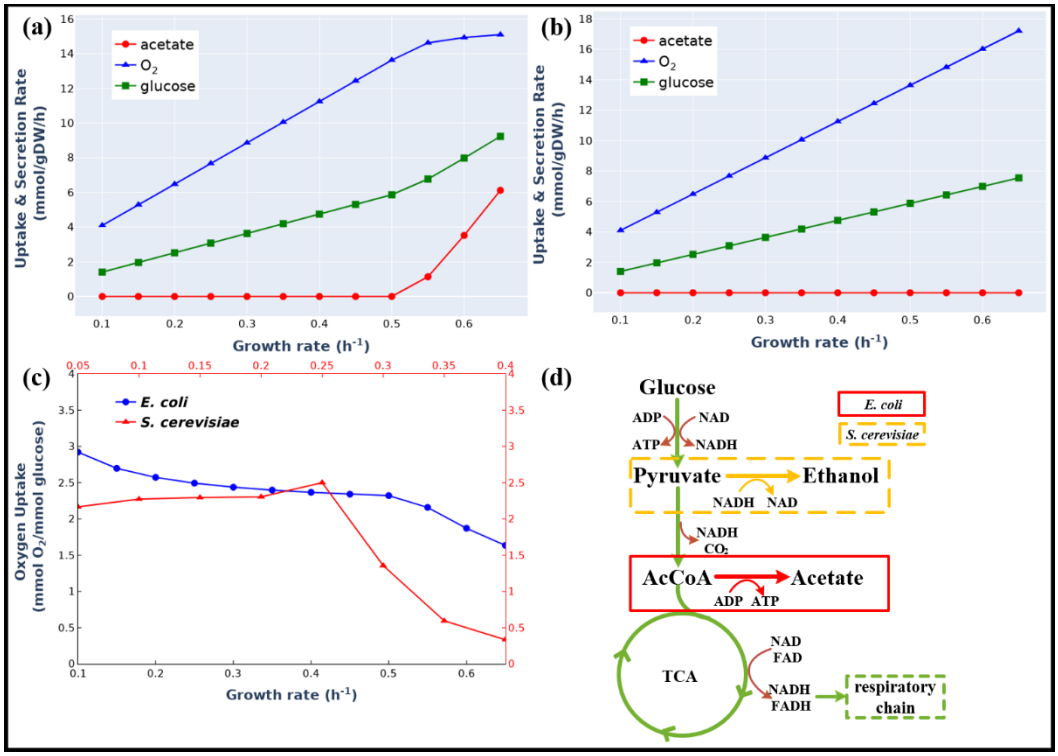

Figure 3. Comparison of simulation results of the enzyme-constrained model ec-*i*ML1515 and the stoichiometric model *i*ML1515. Simulation of overflow metabolism at different growth rates using ec*i*ML1515 (a) and *i*ML1515 (b). c. Simulated different overflow metabolism of *E. coli* and *S. cerevisiae*. d. The different overflow metabolic pathways of *E. coli* and *S. cerevisiae*.

3.3. Maximum growth rate of E. coli on different carbon sources

We simulated the maximum growth rates of *E. coli* on 24 different carbon sources, and observed that certain other fermentation byproducts (e.g., pyruvate and fumarate) in addition to acetate could also be produced at the maximal growth rates. The predicted results were in good agreement with previously reported experimental results [14] as shown in Figure 4a (the normalized flux error is 0.062) and Table S5. On the other hand, the calculated growth rates using *i*ML1515 (the substrate uptake rates were set as same with those for ec*i*ML1515) were significantly higher than the measured values (the standard flux error is 0.205, Figure 4b). The prediction results for most of substrates (e.g., N-Acetyl-D-glucosamine and glucose) from ec*i*ML1515 were closer to (estimation error is 0.01 and 0.03) experimental values than those from *i*ML1515 model. In stoichiometric model like *i*ML1515, the substrate uptake rate needs to be preset to calculate the growth rate and there is a linear relationship between the growth rate and substrate consumption rate. Whereas in the enzyme-constrained model, the maximal growth rate is limited by enzyme resources and thus there is no need to preset a substrate consumption rate. This means that at the maximal growth rate, a considerable quantity of substrates was actually utilized through the fermentation pathways with the secretion of fermentation products. Therefore, the predicted growth rates from the enzyme-constrained model were significantly lower than those from *i*ML1515 but much closer to the experimental findings. One exception for acetate as the carbon source is that the predicted results were same for both models as no acetate producing fermentation pathway was activated in this case. From the results shown in Figure 4a, we can also see that for most carbon sources the predicted growth rates were still higher than the experimentally measured rates. This may imply that there are other constraints along with enzyme constraints limiting cellular growth, such as the regulatory or thermodynamic constraints. New models integrating these new constraints in proper formula can further improve the prediction accuracy [38]. For xylose

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

and glycerol, the predicted rates were smaller than the experimental values, implying that the k_{cat} values of enzymes in the uptake pathways of these two substrates may be underestimated. Besides, we found that ECMpy is better than GECKO and sMOMENT for the simulation of growth rate on 24 different carbon sources (all consider protein subunits, but ECMpy corrected for enzyme kinetic parameters), and the simulation results of all enzyme-constrained models are also better than non-enzyme-constrained models (Figure 4a-d). This may also mean more precise measurement of the enzyme kinetic parameters could improve model prediction.

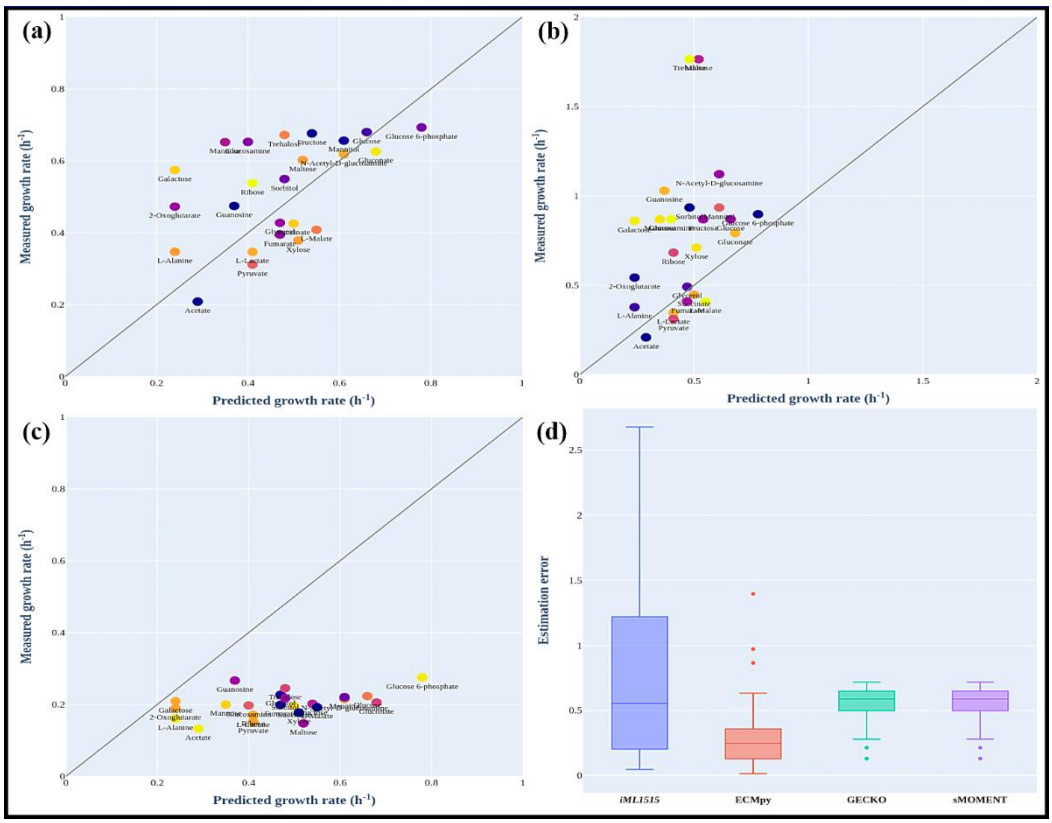

Figure 4. Predicted *E. coli* growth rates on different carbon sources using ECMpy (a), *i*ML1515 (b), GECKO and sMOMENT (c). d. Estimation fluxes error of different model (GECKO and sMOMENT consider protein subunits).

3.4. Simulation of trade-off between enzyme usage efficiency and biomass yield

In addition to the maximal growth rates under different carbon sources, we also explored the metabolic behaviors of *E. coli* at different substrate (glucose as an example) uptake rates. As shown in Figure 5a and b, the metabolism processes can be divided into three stages: substrate-limited stage, overflow switching phase and overflow stage. At the first stage, the glucose uptake rate is low and has a linear relationship with growth rates. The biomass yield is almost constant (not exactly the same as a small number of substrates are used for non-growth-related maintenance). At the second stage, the cell redistributes the intracellular fluxes toward pathways with high enzyme usage efficiency but low biomass yield, and acetate gradually becomes a byproduct of the newly activated pathways. In contrast, at the overflow stage, the organism has to activate the less energy efficient but higher enzyme usage efficiency fermentation pathway to produce energy required for growth, leading to a sharp drop of biomass yield due to a big fraction of substrates used in the fermentation pathway. There was a clear trade-off between yield and enzyme usage efficiency (Figure 5b). These predicted metabolic behaviors were consistent with long-standing empirical models of microbial growth [39]. This trade-off phenomenon was also predicted by the *E. coli* ME-model [40], indicating that the enzyme-constrained model

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

could accurately predict the same phenomenon as ME-model but without introducing thousands of new reactions involved in transcription and translation process in the model.

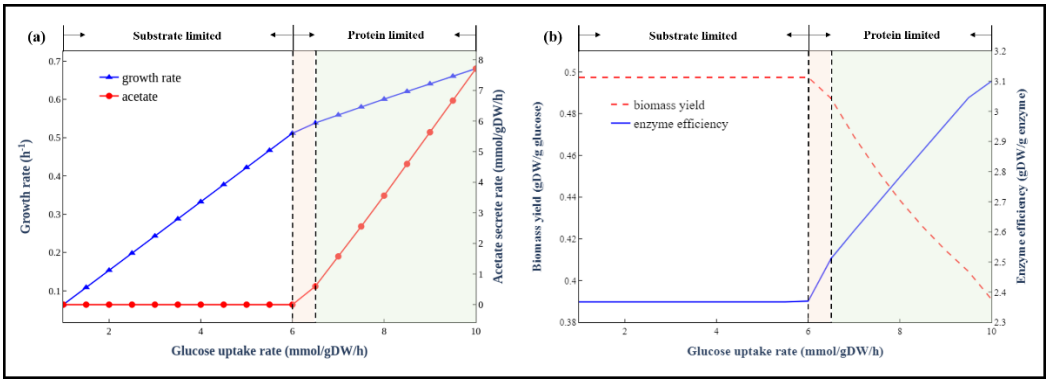

Figure 5. (a) Simulated growth rates at different glucose uptake rates. (b) The trade-off between biomass yield and enzyme efficiency.

4. Discussion

We constructed a genome scale enzyme-constrained model ec*i*ML1515 for *E. coli* using the simplified Python-based ECMpy workflow. The new model was validated with various experimental data from literature including metabolic overflow data and the growth rates under different carbon sources. The prediction results were better than GECKO and sMOMENT, and those enzyme-constrained models also better than original *i*ML1515, indicating in these conditions enzyme availability rather than network stoichiometry is the key constraints. The enzyme-constrained model also showed a clear trade-off between biomass yield and enzyme usage efficiency. Switching from a high yield pathway to a high-rate pathway could be a general principal in metabolic regulation. This provides new insight in engineering organisms for production of valuable biochemicals. In organism using a high yield and high enzyme cost biosynthesis pathway, improving enzyme specific activity could be more effective than enzyme overexpression.

Different from GECKO and sMOMENT, our method for enzyme constrained model construction just adding a constraint on the total amount of enzyme does not need to modify the reaction equations (e.g., introduce enzyme as reactants) and introduce over a thousand new enzyme exchange reactions (like GECKO). This greatly reduces the complexity in model construction and the model can be solved using COBRApy or other freely available python packages for constrained optimization. The whole model construction and simulation processes were written in Jupyter Notebook files available from internet. This enables people from anywhere to reproduce the work and construct their own enzyme constrained models for other organisms.

As we have shown that the quality of enzyme constrained model depended largely on the quantity and accuracy of enzyme parameters. Even for *E. coli*, the enzyme kinetic data coverage is low in databases such as BRENDA and kinetic parameters from different researches are often inconsistent. In this study, we make use of the predicted data from machine learning [22] to improve the data coverage. Besides, enzyme-constrained models need model validation to adjust the original k_{cat} values to some extent to improve the agreement of model predictions with experimental data [16]. A system kinetic parameter correction method has been presented in the sMOMENT workflow [16], which is helpful in identifying such unreliable parameters and improving model predication accuracy. However, this calibration workflow is time consuming, going through protein pool calibration, manual k_{cat} adjustment and automated k_{cat} calibration, and there are some unreasonable places, such as the manual correction is simply expanded by 10 times or reduced by 10 times. In recently, GECKO 2.0 provided an automatic procedure, in which the top enzymatic limitation on growth rate is identified and its correspondent k_{cat} is then iteratively replaced by the highest one available in BRENDA for the given enzyme class until the growth rate fit is normal [41]. Currently, we propose a simpler calibration process that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

requires only two steps (enzyme usage and C^{13} flux consistency, see method) to update the k_{cat} for a small number of reactions to achieve a better growth rate fit. This new calibration process will facilitate the construction of high-quality enzyme constraint models.

5. Conclusions

We presented ECMpy, a simple open-source Python-based workflow, for constructing enzyme-constrained models based on enzyme kinetic parameters and proteomics data. Using this method, we constructed an enzyme constrained model ec*i*ML1515 for *E. coli*. By introducing the enzyme constraints, the model can predict the overflow metabolism and growth under different carbon sources more precisely than the stoichiometric model *i*ML1515. The construction method can be applied to construct enzyme constrained models for other organisms and optimization framework can be extended to integrate other constraints such as thermodynamic feasibility to further reduce the solution space and subsequently improve model prediction accuracy.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title.

Author Contributions: Conceptualization, Zhitao Mao, Xin Zhao and Hongwu Ma; Data curation, Zhitao Mao and Xin Zhao; Formal analysis, Zhitao Mao and Xin Zhao; Funding acquisition, Hongwu Ma; Methodology, Zhitao Mao, Xin Zhao and Hongwu Ma; Project administration, Zhitao Mao and Hongwu Ma; Software, Zhitao Mao and Peiji Zhang; Validation, Xin Zhao, Xue Yang, Peiji Zhang, Jiawei Du and Qianqian Yuan; Writing – original draft, Zhitao Mao, Xin Zhao and Xue Yang; Writing – review & editing, Zhitao Mao, Xin Zhao, Xue Yang, Qianqian Yuan and Hongwu Ma.

Conceived the research, HM; Developed the automatic workflow of the enzyme-constrained model, ZM and XZ; Designed the enzyme-constrained model construction method and analyzed ec*i*ML1515, ZM, XZ, and XY. Wrote the manuscript, ZM and XZ. Further perfected the workflow, XY, PZ, JD and QY. All authors read and approved the final manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (2018YFA0900300, 2018YFA0901400); the International Partnership Program of Chinese Academy of Sciences (153D31KYSB20170121). Other than supplying funds, the funding agencies played no role in the development of this research, the analysis of results, or in the preparing of this manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The scripts and datasets generated during and/or analyzed during the current study can be found at: https://github.com/tibbdc/ECMpy.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Edwards JS, Palsson BO: **Systems properties of the Haemophilus influenzae Rd metabolic genotype**. *J Biol Chem* 1999, **274**(25):17410-17416.
2. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY: **Current status and applications of genome-scale metabolic models**. *Genome Biol* 2019, **20**(1):121.
3. O'Brien EJ, Monk JM, Palsson BO: **Using Genome-scale Models to Predict Biological Capabilities**. *Cell* 2015, **161**(5):971-987.
4. Kerkhoven EJ, Lahtvee PJ, Nielsen J: **Applications of computational modeling in metabolic engineering of yeast**. *FEMS Yeast Res* 2015, **15**(1):1-13.
5. Schuetz R, Kuepfer L, Sauer U: **Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli**. *Mol Syst Biol* 2007, **3**:119.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

6. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models**. *Metabolic engineering* 2003, **5**(4):264-276.
7. Lin Z, Zhang Y, Yuan Q, Liu Q, Li Y, Wang Z, Ma H, Chen T, Zhao X: **Metabolic engineering of Escherichia coli for poly(3-hydroxybutyrate) production via threonine bypass**. *Microb Cell Fact* 2015, **14**:185.
8. Veit A, Polen T, Wendisch VF: **Global gene expression analysis of glucose overflow metabolism in Escherichia coli and reduction of aerobic acetate formation**. *Applied microbiology and biotechnology* 2007, **74**(2):406-421.
9. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T: **Overflow metabolism in Escherichia coli results from efficient proteome allocation**. *Nature* 2015, **528**(7580):99-104.
10. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabasi AL, Oltvai ZN: **Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity**. *Proc Natl Acad Sci U S A* 2007, **104**(31):12663-12668.
11. Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E: **Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect**. *PLoS Comput Biol* 2011, **7**(3):e1002018.
12. Zeng H, Yang A: **Modelling overflow metabolism in Escherichia coli with flux balance analysis incorporating differential proteomic efficiencies of energy pathways**. *BMC systems biology* 2019, **13**(1):3.
13. Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, Liebermeister W: **The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization**. *PLoS Comput Biol* 2016, **12**(11):e1005167.
14. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T: **Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters**. *PLoS Comput Biol* 2012, **8**(7):e1002575.
15. Sanchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J: **Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints**. *Mol Syst Biol* 2017, **13**(8):935.
16. Bekiaris PS, Klamt S: **Automatic construction of metabolic models with enzyme constraints**. *BMC Bioinformatics* 2020, **21**(1):19.
17. Massaiu I, Pasotti L, Sonnenschein N, Rama E, Cavaletti M, Magni P, Calvio C, Herrgard MJ: **Integration of enzymatic data in Bacillus subtilis genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-gamma-glutamic acid production strains**. *Microbial cell factories* 2019, **18**(1):3.
18. Chen Y, Sun Y, Liu Z, Dong F, Li Y, Wang Y: **Genome-scale modeling for Bacillus coagulans to understand the metabolic characteristics**. *Biotechnology and Bioengineering* 2020, **117**(11):3545-3558.
19. Ye C, Luo Q, Guo L, Gao C, Xu N, Zhang L, Liu L, Chen X: **Improving lysine production through construction of an Escherichia coli enzyme-constrained model**. *Biotechnology and Bioengineering* 2020, **117**(11):3533-3544.
20. Sulheim S, Kumelj T, van Dissel D, Salehzadeh-Yazdi A, Du C, van Wezel GP, Nieselt K, Almaas E, Wentzel A, Kerkhoven EJ: **Enzyme-Constrained Models and Omics Analysis of Streptomyces coelicolor Reveal Metabolic Changes that Enhance Heterologous Production**. *iScience* 2020, **23**(9):101525.
21. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H *et al*: **iML1515, a knowledgebase that computes Escherichia coli traits**. *Nat Biotechnol* 2017, **35**(10):904-908.
22. Heckmann D, Lloyd CJ, Mih N, Ha Y, Zielinski DC, Haiman ZB, Desouki AA, Lercher MJ, Palsson BO: **Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models**. *Nat Commun* 2018, **9**(1):5252.
23. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nature Biotechnology* 2010, **28**(3):245-248.
24. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR: **COBRApy: COnstraints-Based Reconstruction and Analysis for Python**. *BMC Syst Biol* 2013, **7**:74.
25. Motamedian E, Mohammadi M, Shojaosadati SA, Heydari M: **TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data**. *Bioinformatics* 2017, **33**(7):1057-1063.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2021 doi:10.20944/preprints202112.0048.v1

26. Machado D, Herrgard M: **Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism**. *PLoS Comput Biol* 2014, **10**(4):e1003580.
27. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D *et al*: **Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models**. *Mol Syst Biol* 2010, **6**:390.
28. Karp PD, Ong WK, Paley S, Billington R, Caspi R, Fulcher C, Kothari A, Krummenacker M, Latendresse M, Midford PE *et al*: **The EcoCyc Database**. *EcoSal Plus* 2018, **8**(1).
29. Nilsson A, Nielsen J, Palsson BO: **Metabolic Models of Protein Allocation Call for the Kinetome**. *Cell Syst* 2017, **5**(6):538-541.
30. Bremer H: **Modulation of chemical composition and other parameters of the cell by growth rate**. *American Society for Microbiology* 1996:1553–1569.
31. Brunk E, Mih N, Monk J, Zhang Z, O'Brien EJ, Bliven SE, Chen K, Chang RL, Bourne PE, Palsson BO: **Systems biology of the structural proteome**. *BMC systems biology* 2016, **10**:26.
32. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, von Mering C: **PaxDb, a database of protein abundance averages across all three domains of life**. *Molecular & cellular proteomics : MCP* 2012, **11**(8):492-500.
33. Okahashi N, Kajihata S, Furusawa C, Shimizu H: **Reliable Metabolic Flux Estimation in Escherichia coli Central Carbon Metabolism Using Intracellular Free Amino Acids**. *Metabolites* 2014, **4**(2):408-420.
34. Chen Y, Nielsen J: **Energy metabolism controls phenotypes by protein efficiency and allocation**. *Proc Natl Acad Sci U S A* 2019, **116**(35):17592-17597.
35. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110**. *Applied and environmental microbiology* 1994, **60**(10):3724-3731.
36. Thomas TD, Ellwood DC, Longyear VM: **Change from homo- to heterolactic fermentation by Streptococcus lactis resulting from glucose limitation in anaerobic chemostat cultures**. *Journal of bacteriology* 1979, **138**(1):109-117.
37. van Hoek MJ, Merks RM: **Redox balance is key to explaining full vs. partial switching to low-yield metabolism**. *BMC systems biology* 2012, **6**:22.
38. Yang X, Mao Z, Zhao X, Wang R, Zhang P, Cai J, Xue C, Ma H: **Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models**. *Metabolic Engineering* 2021, **67**:133-144.
39. Koch AL: **Microbial physiology and ecology of slow growth**. *Microbiology and molecular biology reviews : MMBR* 1997, **61**(3):305-318.
40. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO: **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction**. *Mol Syst Biol* 2013, **9**:693.
41. Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N, Nielsen J: **Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0**. *bioRxiv* 2021:2021.2003.2005.433259.