

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

ecBSU1: A Genome-scale Enzyme-constrained Model of *Bacillus subtilis* based on the ECMpy Workflow

Ke Wu ^{1,2,3,†}, Zhitao Mao ^{2,3,†}, Yufeng Mao ^{2,3}, Jinhui Niu ^{2,3}, Jingyi Cai ^{2,3}, Qianqian Yuan ^{2,3}, Lili Yun ⁴, Xiaoping Liao ^{2,3}, Zhiwen Wang ^{1,*}, and Hongwu Ma ^{2,3,*}

¹ Key Laboratory of Systems Bioengineering (Ministry of Education); Frontier Science Center for Synthetic Biology (Ministry of Education); Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China

² Biodesign Center, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China

³ National Technology Innovation Center of Synthetic Biology, Tianjin, China

⁴ Tianjin Medical Laboratory, BGI-Tianjin, BGI-Shenzhen, Tianjin, China

* Correspondence: zww@tju.edu.cn; ma_hw@tib.cas.cn

† These authors contributed equally to this work

Abstract: Genome-scale metabolic models (GEMs) play an important role in the phenotype prediction of microorganisms, and their accuracy can be further improved by integrating other types of biological data such as enzyme concentrations and kinetic coefficients. Enzyme-constrained models (ecModels) have been constructed for several species and were successfully applied to increase the production of commodity chemicals. However, there was still no genome-scale ecModel for the important model organism *Bacillus subtilis* prior to this study. Here, we integrated enzyme kinetic and proteomic data to construct the first genome-scale ecModel of *B. subtilis* (ecBSU1) using the ECMpy workflow. We first used ecBSU1 to simulate overflow metabolism and explore the trade-off between biomass yield and enzyme usage efficiency. Then, we simulated the growth rate on eight previously published substrates and found that the simulation results of ecBSU1 were in good agreement with the literature. Finally, we identified target genes that enhance the yield of commodity chemicals using ecBSU1, most of which were consistent with the experimental data, and some of which may be potential novel targets for metabolic engineering. This work demonstrates that the integration of enzymatic constraints is an effective method to improve the performance of GEMs. The ecModel can predict overflow metabolism more precisely and can be used for the identification of target genes to guide the rational design of microbial cell factories.

Keywords: enzyme-constrained model; *Bacillus subtilis*; metabolic engineering

1. Introduction

Bacillus subtilis is a model organism of choice for the industrial production of various valuable compounds, such as biopolymers and proteins, due to its excellent capacity for protein secretion, good growth characteristics, distinct endogenous metabolism, and robustness in industrial fermentation [1]. Genome-scale metabolic network models (GEMs) of *B. subtilis* have been successfully applied to guide the production of riboflavin [2], isobutanol [2], 2,3-butanediol [3], and 3-hydroxypropionic acid [4]. The first *B. subtilis* GEM was published in 2008 [5], and several models were subsequently reported [2, 6–8], which were updated in terms of reactions, metabolites and genes. The iBsu1147 model constructed by our team [2] has the highest number of reactions, metabolites and genes of all models reported to date (Fig. S1). However, GEMs only consider stoichiometric constraints, making them unable to reflect the true state of the cell and locate kinetic bottlenecks limiting the flux through specific product synthesis pathways.

By contrast, enzyme-constrained models (ecModels) introduce enzyme kinetic information into a GEM, thus reflecting the protein resource limitation faced during cell growth, enabling them to identify the rate-limiting enzymes in the pathway and further

guide rational metabolic engineering. As a consequence, ecModels have been successfully applied to guide the production of L-lysine [9], poly-glutamic acid [10], heme [11] and recombinant proteins [12]. Currently, three methods exist to automate the construction of ecModels, including GECKO [13], AutoPACMEN [14] and ECMpy [15]. GECKO is the earliest automated approach for the introduction of protein resource constraints into GEMs, which introduces average enzyme saturation coefficients and enzyme usage constraints from proteomic data [13]. However, GECKO adds many pseudo-metabolites representing enzymes, significantly increasing the complexity and scale of the model. Inspired by GECKO, Bekiaris et al. developed the AutoPACMEN automated workflow for the construction of ecModels, which introduces only one pseudo-reaction and pseudo-metabolite [14]. Recently, we developed the simplified Python-based workflow ECMpy, which allows the construction of an ecModel by directly adding a total enzyme amount constraint into a GEM [15]. Recently, ecModels have been constructed for several species, including *Escherichia coli* [9, 12, 15], *Saccharomyces cerevisiae* [13], *Aspergillus niger* [16], *Corynebacterium glutamicum* [17] and *B. subtilis* [10]. The first ecModel for *B. subtilis* (ec_iYO844) only integrated enzyme kinetic parameters for 17 reactions located in the central carbon metabolism using the GECKO method, but this model allowed more accurate prediction of the flux distribution and growth rate of wild-type and single-gene/operon deletion strains compared to the GEM [10].

In this study, we first systematically updated the iBsu1147 model through GPR update, biomass reaction standardization, etc., after which we established a comprehensive collection of parameters that affect the prediction accuracy of the ecModel (e.g., the enzyme kinetics data and quantitative information on enzyme subunit composition), and constructed the first genome-scale ecModel of *B. subtilis* (ecBSU1), using an updated ECMpy workflow [17]. Subsequently, we used ecBSU1 to accurately predict the growth rate of *B. subtilis* on different carbon sources, simulate the overflow metabolism, and explain the trade-off between biomass yield and enzyme usage efficiency. Finally, we predicted the target genes of *B. subtilis* for enhancing the production of industrial chemicals (e.g., riboflavin, menaquinone 7, and acetoin, etc.), and the predictions were in good agreement with the literature.

2. Methods

2.1. Model update

The model iBsu1147, which has the most reactions and genes, was selected as the initial model for the integration of enzymatic constraints. Since the iBsu1147 model was released in 2013, we performed quality control on the model, covering substrate utilization, redox balance, energy balance, biomass reaction standardization, and mass balance. Our previous results shown that the k_{cat} and molecular weight (MW) of an enzyme affect the prediction accuracy of the ecModel [15]. For example, the correctness of the EC number affects the correctness of the corresponding k_{cat} , and the GPR relationship affects the accuracy of MW calculation. Accordingly, we systematically corrected the EC number and GPR relationships. We used the GPRuler tool [18] and protein homology similarity to uncover the potential GPR errors present in the reaction (see [17] for details). To meet the requirements of AutoPACMEN processes for metabolic network format input, we converted most of the KEGG IDs and ModelSEED [19] IDs (both metabolites and reactions) into BiGG [20] IDs. Finally, we named the modified model iBsu1147^R (Revised iBsu1147).

2.2. Data acquisition

The molecular weight (MW) of each enzyme was downloaded from the UniProt database according to the gene ID, and the quantitative subunit information was obtained by parsing the 'Interaction information' in UniProt [17]. For example, P39119 is described in UniProt as a 'homodimer', so its subunit number is 2 (all corresponding information is listed in Table S1). The k_{cat} values were obtained from the BRENDA [21] and SABIO-RK [22] databases according to the EC numbers using AutoPACMEN. *B. subtilis* protein

abundance data were obtained from the PAXdb [23] database, and enzyme mass fraction was calculated according to Eq. 1:

$$f = \sum_{i=1}^{p_num} A_i MW_i / \sum_{j=1}^{g_num} A_j MW_j \quad (1)$$

where A_i and A_j represented the abundance of the i -th protein (p_num represents proteins expressed in the model) and j -th protein (g_num represents proteins expressed in the whole proteome).

2.3. Construction of ecBSU1

The enzyme-constrained model was constructed based on iBsu1147^R following the ECMpy workflow and named ecBSU1 (Fig. 1). Firstly, we divided reversible reactions in iBsu1147^R into pairs of irreversible reactions, and split reactions catalyzed by multiple iso-enzymes into different reactions (append num in reaction ID, e.g., GLCpts_num1), so that each reaction only has one corresponding enzyme. Then, we calculated the MW of each enzyme. For reactions catalyzed by enzyme complexes, we used the total sum of proteins in the complex ($MW = \sum_{j=1}^m N_j * MW_j$, where m is the number of different subunits in the enzyme complex and N_j is the number of j th subunits in the complex). Finally, a new enzymatic constraint ($\sum_{i=1}^n \frac{v_i * MW_i}{\sigma_i * k_{cat,i}} \leq ptot * f$) was introduced into the model, where $ptot$, f , and σ_i represent the total protein fraction in *B. subtilis*, the mass fraction of enzymes, and the saturation coefficient of the i -th enzyme, respectively (see [15] for details).

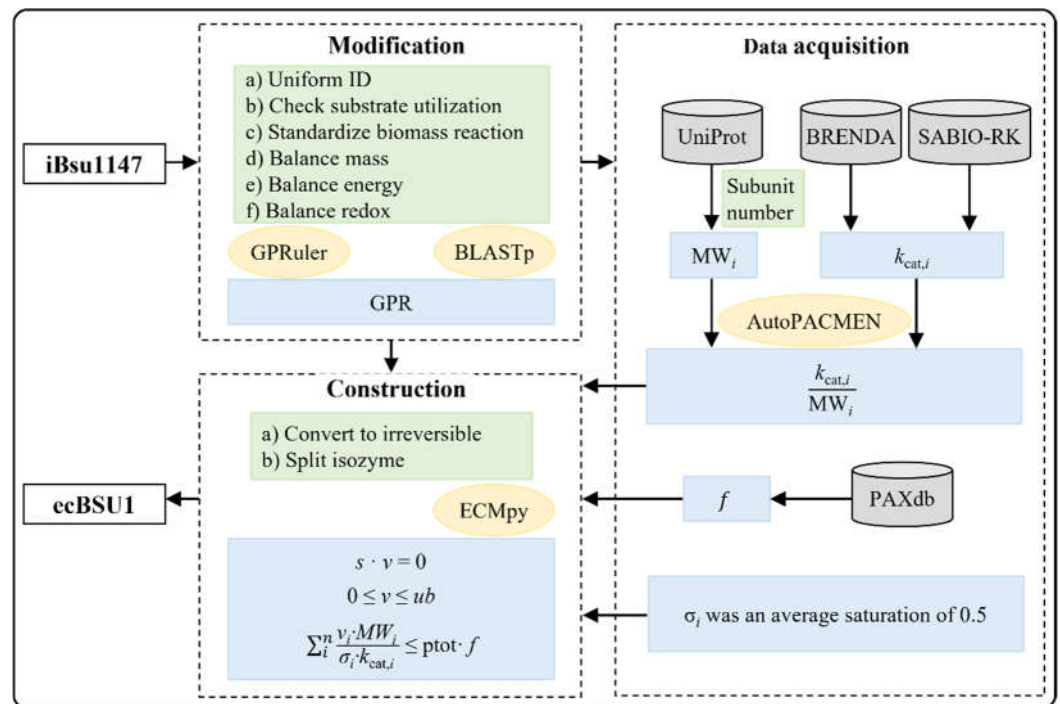


Figure 1. Workflow for the construction of ecBSU1.

2.4. Calibration of enzyme kinetic parameters

To improve the agreement of model predictions with experimental data, the enzyme-constrained model required further adjustment of the original k_{cat} values, analogous to GECKO, AutoPACMEN and ECMpy. In this work, we calculated the reaction enzyme cost (see [15] for details) to identify potential reactions that need correction, based on the largest enzyme cost. For enzymes that require calibration, we modified the reaction k_{cat} to the maximal corresponding k_{cat} in the BRENDA and SABIO-RK databases. We reiterated the above correction until the growth rate reached a reasonable steady state, as described in GECKO 2.0 [24].

2.5. Phenotype phase plane (PhPP) analysis

Different rates of substrate uptake and oxygen supply affect the cellular metabolic phenotype, leading to different maximal growth rates. We performed PhPP analysis on iBsu1147^R and ecBSU1 as described in the literature [16] to obtain a global view of how optimal growth rates are affected by varying glucose and oxygen uptake rates. To implement the PhPP analysis, the reaction fluxes of oxygen and glucose were respectively varied in the range of 0-50 mmol/gDW/h and 0-15 mmol/gDW/h, after which pFBA calculations were performed with biomass maximization as the objective.

2.6. Prediction of growth rates on different carbon sources

To evaluate the ability of ecBSU1 to predict phenotypes, we simulated the growth rate of *B. subtilis* on 8 different carbon sources, and compared the prediction results of iBsu1147^R and ecBSU1 with reported values [25]. Then, the model and experimental results were used to calculate the estimation error of the growth rate and normalized flux error (see [15] for detail).

2.7. Simulation of overflow metabolism

We explored the overflow metabolism of the *B. subtilis* using ecBSU1 by setting the substrate uptake rate on a gradient of 0 to 10 mmol/gDW/h and solving for pFBA to maximize the biomass. To further explain the metabolic overflow phenomenon, we analyzed the biomass yield, enzyme usage efficiency, reaction enzyme cost, energy synthesis enzyme cost, and oxidative phosphorylation ratio (proportion of glucose used for the oxidative phosphorylation pathway to total glucose) (see [15, 26] for details).

2.8. Prediction of metabolic engineering targets

Compared to GEMs, ecModels can calculate enzyme costs in addition to reaction fluxes, which is useful for identifying key enzymes in the pathway [9]. In this work, we analyzed the enzyme cost of each reaction to identify kinetic bottleneck reactions (the reactions with the largest enzyme cost) (Eq. 2) by setting glucose as the substrate, the product as the objective, and the low bound of biomass reaction as 10% of the maximal growth rate, as described in the literature [9]. Finally, we selected the Top 15 reactions with the highest enzyme cost as potential metabolic engineering targets.

$$\text{Enzyme cost}_i = \frac{v_i \cdot MW_i}{\sigma_i \cdot k_{cat,i}} \quad (2)$$

3. Results

3.1. GPR correction of iBsu1147

EC numbers affect the extraction and assignment of k_{cat} data in the workflow, ultimately affecting the prediction accuracy of the ecModel. Consequently, we replaced the old EC numbers in the model based on BRENDA, updating a total of 38 reactions (Table S2). MW is also an important factor affecting the prediction accuracy of ecModels. Two major factors influence the final MW of the enzyme assigned to a specific reaction: whether the protein is composed of subunits (GPR relationship) and the number of each subunit. We systematically corrected the GPR relationships in the model by referring to the methods used in CGL1 (GPRuler tool and protein homology similarity) [17]. We first identified 146 reactions containing protein complex information using the GPRuler tool, 80 of which were consistent with the model. For the remaining 66 reactions in the model that contained “and” relationships, we performed a manual check using the UniProt, KEGG, and BioCyc databases, and found that 27 reactions were correct in the model, 6 reactions were correct in the GPRuler tool, and the remaining 33 reactions were incorrect in both (Table S3). For example, the GPR relationship of the ribose ABC transport system is ‘BSU35930 and BSU35940 and BSU35950 and BSU35960’ in the model, but the GPRuler tool did not include ‘BSU35930’. By searching UniProt, we found that ‘BSU35930’ encodes

D-ribose pyranase, which catalyzes the interconversion of beta-pyran and beta-furan forms of D-ribose, indicating that the result of GPRuler tool is correct. In addition, when verifying the results of the GPRuler tool, some reactions in the model were modified based on the annotation information of the proteins in the databases, including the deletion of 7 reactions (e.g., PDHbr and AKGDb), the addition of 4 new reactions (e.g., AKGDH and FCLT3), and the modification of GPR relationships for 5 reactions (e.g., RBFSb and 26DPAi).

We also observed that there were 58 reactions in the model for which the “and” relationships were not identified by the GPRuler tool, and only 9 reactions had more than 25% similarity (Table S4). By searching the database, we found that 7 reactions needed to be corrected, 3 of which needed to be changed from “or” to “and” relationships, and 4 in which the proteins needed to be replaced (Table S4). For example, the GPR relationship for NADH-dependent butanol dehydrogenase (BTS) is ‘BSU31360 and BSU31370’, which has 74% sequence similarity. By further searching the BioCyc database for verification, we found that the protein encoded by BSU31370 is a bifunctional enzyme that catalyzes two reactions (NADPH-dependent furan aldehyde reductase and NADPH-dependent butanol dehydrogenase), whereas the protein encoded by BSU31360 catalyzes only the NADPH-dependent butanol dehydrogenase reaction. They are more likely to be two iso-enzymes, and therefore the correct GPR relationship for BTS should be ‘BSU31360 or BSU31370’. Finally, 1736 reactions, 1459 metabolites, and 1155 genes were included in iBsu1147^R.

3.2. Other modifications of iBsu1147

After running quality control of the iBsu1147^R model in terms of substrate utilization, redox balance, energy balance, biomass reaction standardization, and mass balance, we found that all these aspects led to abnormalities in the metabolic pathways generated by the simulation. The boundary of 6 reactions was modified in terms of substrate utilization (e.g., EX_chor_e and MALt2r). For example, experiments have shown that *B. subtilis* can grow using malate [25], so the upper and lower boundaries of the malate transport reaction (MALt2r) in the model should not be 0. From a reducing power perspective, we modified the catalytic orientation of 4 reactions (e.g., NODOx and P5CR) based on BioCyc [27] to avoid pathway calculation errors. For example, P5CR (1pyr5c_c + h_c + nadph_c → nadp_c + pro__L_c) is reversible in the model, but a search by BioCyc revealed that the reaction acts as the final step of the L-proline synthesis pathway I, which is unidirectional. In addition, the catalytic orientations of 3 respiratory chain-related reactions were also modified (e.g., CYOR3m and CYTB_B2). In total, 13 reactions in the model were corrected for the boundaries (Table S5).

In addition, the molar mass of biomass and its components (e.g., proteins, nucleic acids, etc.) was 1 g/mmol, and deviations from this value will result in errors in the calculated specific growth rate. Using the BiomassMW algorithm [28], we examined the biomass equation of the iBsu1147^R model and found that the original molecular mass of the biomass was 1.025 g/mmol, and H⁺ was missing in the right side of the biomass equation (the coefficient was 105) produced by the hydrolysis of ATP for growth-associated maintenance energy. After correction, the molar mass of the biomass became 0.92 g/mmol. Subsequently, we examined the precursor metabolites and found that the molar masses of protein, DNA, and RNA were 0.86 g/mmol, 0.95 g/mmol, and 0.95 g/mmol, respectively. We normalized the coefficients so that the molar masses of all components were 1 g/mmol, and the molar masses of the biomass also became 1 g/mmol. The details of all the modified reactions are listed in Table S6.

Finally, we found that iBsu1147^R contains different IDs for metabolites and reactions, including KEGG IDs and ModelSEED IDs. To meet the input requirements of AutoPAC-MEN, we converted the KEGG IDs of 1007 metabolites and 785 reactions, as well as the ModelSEED IDs of 265 metabolites and 542 reactions into BiGG IDs. In addition, we kept the original IDs for reactions and metabolites that were not included in the BiGG database.

3.3. Basic information of ecBSU1

We used AutoPACMEN to match 2331 k_{cat} values (439 were obtained by filling) for 3307 reactions (splitting of reversible reactions and isozymes), accounting for 70.5% of the total reactions and 76.4% excluding exchange reactions. In total, 1892 reactions were catalyzed by enzymes with 549 different EC numbers, among which oxidoreductases and transferases accounted for the majority (Fig. 2A, inner ring). These k_{cat} values spanned 9 orders of magnitude, with a median value of 46.17 s⁻¹ (Fig. 2B). In total, the molecular weights for 1155 enzymes were obtained from UniProt based on the corresponding gene IDs, covering 3 orders of magnitude, with a median value of 50.41 kDa (Fig. 2C). The enzyme mass fraction f was calculated from the proteomic data in the PAXdb. For *B. subtilis*, we chose the dataset “*B. subtilis*-Whole organism (Integrated)” with the highest measurement coverage and evaluation score, and calculated $f=0.588$ g enzyme/g protein according to Eq. 1. Finally, the initial *B. subtilis* ecModel (ecBSU1) contained 1155 genes, 1459 metabolites, 3307 reactions, and 2331 enzyme kinetic parameters, with a total enzyme bound of 0.165 g enzymes/gDW.

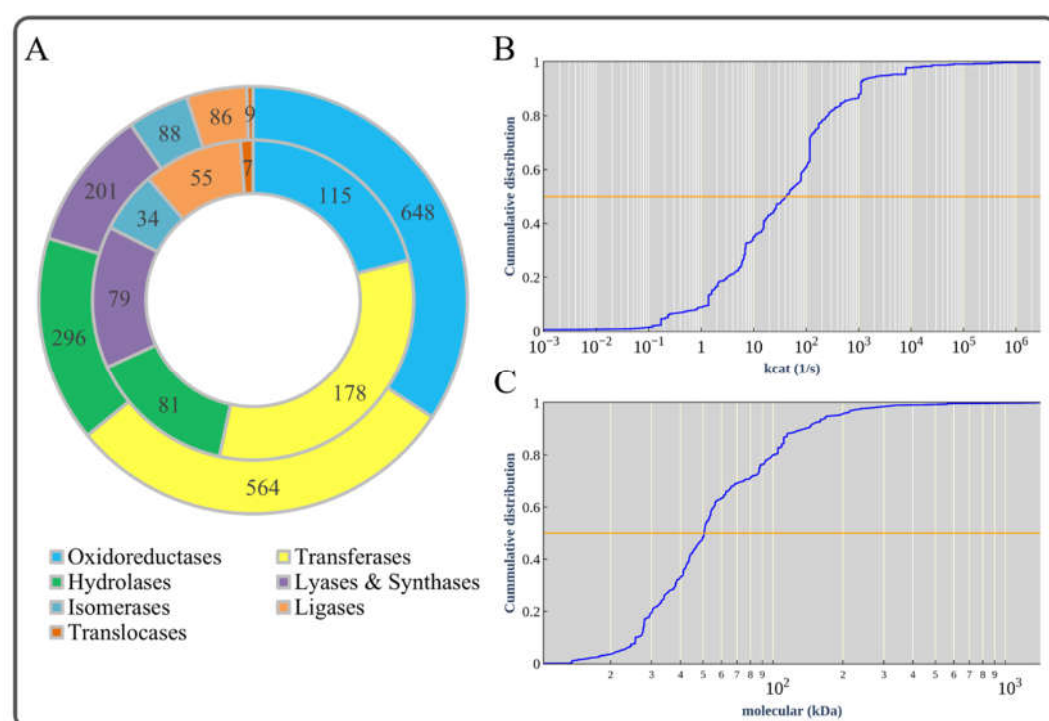


Figure 2. Basic information of ecBSU1. (A) Enzyme classification. The outer ring indicates that there were 1892 reactions with enzyme kinetic data, which can be divided into seven categories. The inner ring indicates that these enzymes included 549 kinds of enzymes according to the corresponding EC numbers, which can also be divided into seven categories. (B) Cumulative distribution of k_{cat} values. (C) Cumulative distribution of molecular weights.

3.4. Correction of enzyme kinetic parameters to overcome model over-constraint

Over-constraint in the initial ecModel is normal due to some reactions with abnormal k_{cat} values (usually too low), as was reported for ecYeast7 [13], eciML1515 [9], eciJB1325 [16] and eciJO1366 [29], which all needed k_{cat} correction. We found that the maximal growth rate calculated by ecBSU1 with glucose as substrate was 0.092 h⁻¹ (Fig. 3A), which was significantly lower than the experimental value of 0.59 h⁻¹ [25]. To overcome this over-constraint, we modified the k_{cat} values of the reactions with the largest enzyme cost to the corresponding maximal k_{cat} values in BRENDA and SABIO-RK. After modifying 28 reactions (Table S7), the maximal growth rate on glucose reached 0.612 h⁻¹ (Fig. 3B), which was consistent with the experimental observations. However, the growth rate predicted by iBsu1147^R increased linearly with increasing carbon source consumption (Fig. 3C), which

is unrealistic. In addition, PhPP analysis showed that the solution space of ecBSU1 was significantly reduced compared with the metabolic network model (Figs. 3B and C). These results demonstrated that incorporating enzymological constraints into a GEM can reduce the flux solution space and enable the model to simulate a more realistic cellular phenotype.

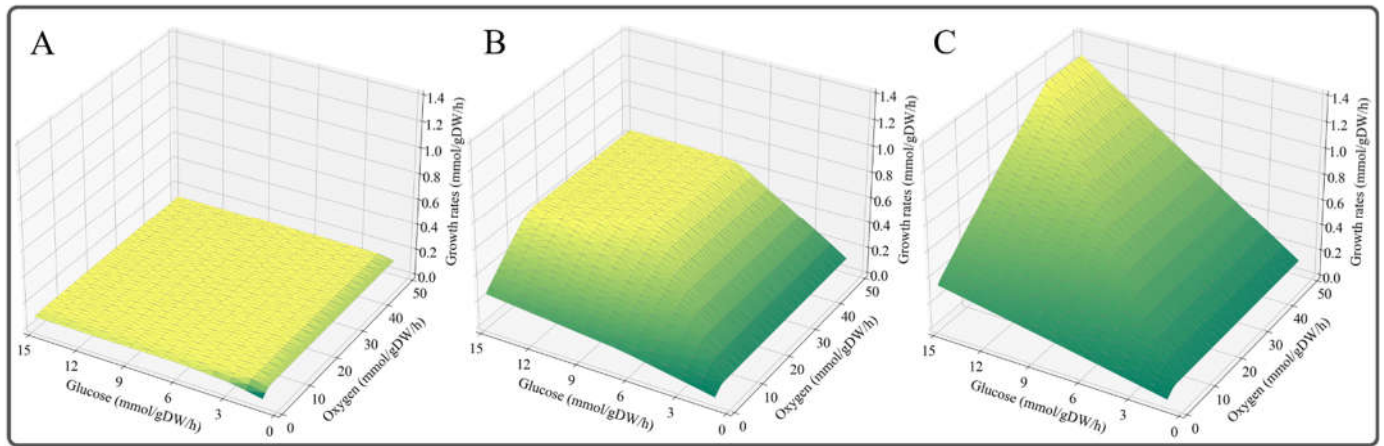


Figure 3. The solution space of iBsu1147^R and ecBSU1. Changes in the maximal growth rate with the increase of glucose and oxygen uptake rates in ecBSU1 before k_{cat} correction (A), ecBSU1 after k_{cat} correction (B), and iBsu1147^R (C).

3.5. Simulation of overflow metabolism

Enzyme-constrained models have been used to simulate overflow metabolism in *S. cerevisiae* [13, 30], *C. glutamicum* [17], and *E. coli* [15]. It has been reported in the literature that *B. subtilis* exhibits overflow metabolism in the presence of excess glucose [31]. We explored the phenomenon of overflow metabolism in *B. subtilis* using ecBSU1, and the simulation results showed that at a high glucose uptake rate of 8 mmol/gDW/h, *B. subtilis* was indeed able to engage metabolic overflow into acetate at high glucose uptake rate (Fig. 4A). By contrast, in iBsu1147^R, glucose increased linearly with the growth rate and could not simulate the phenomenon of overflow metabolism (Fig. 4A). We then calculated the energy synthesis enzyme cost and oxidative phosphorylation ratio to explore the pathway adjustment strategy of overflow metabolism in *B. subtilis*. The simulation results indicated that at high growth rates, the acetate-producing fermentation pathway was activated to maintain growth due to its low enzyme cost in comparison with the energetically efficient respiratory oxidative phosphorylation pathway (Table S8 and Fig. 4B).

To further explain the metabolic overflow phenomenon, we analyzed the biomass yield and enzyme usage efficiency at different glucose uptake rates. As shown in Fig. 4C, there was a clear trade-off between yield and enzyme usage efficiency, so that the metabolic processes could be divided into a substrate-limited stage, overflow switching stage, and overflow stage. At the substrate-limited stage, the glucose uptake rate was low (less than 2.5 mmol/gDW/h) and did not reach the constraint of protein resources, resulting in a substrate uptake rate proportional to the growth rate. At the overflow switching phase (between 2.5 and 8 mmol/gDW/h), the cells redistributed the metabolic fluxes toward pathways with high enzyme usage efficiency but low biomass yield. Finally, overflow metabolism occurred in the overflow stage (greater than 8 mmol/gDW/h). That means at a high glucose uptake rate (8 mmol/gDW/h), *B. subtilis* needs to activate a fermentation pathway with low energy production efficiency but high enzyme efficiency to maintain growth, resulting in the overflow to acetate (Figs. 4A-C).

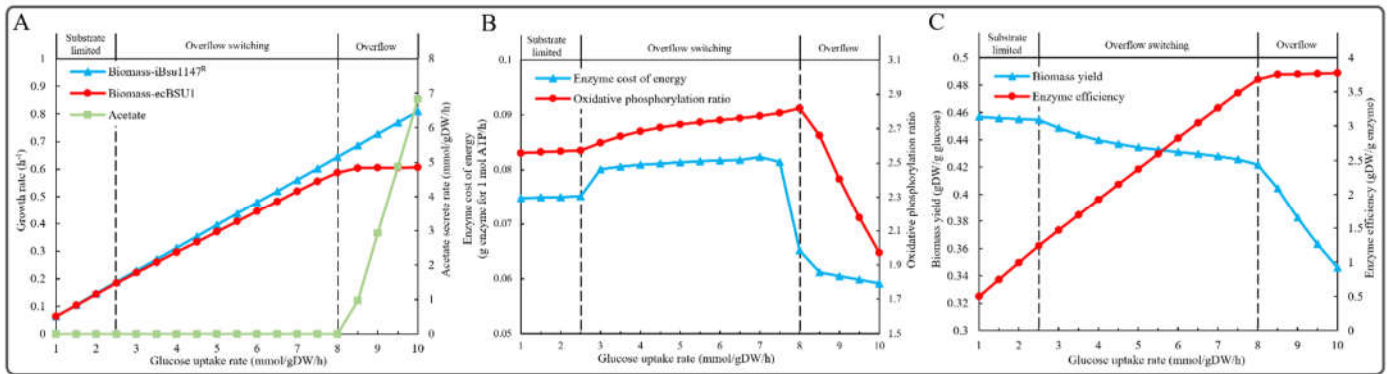


Figure 4. Simulation of overflow metabolism. (A) Comparison of in silico overflow between iBsu1147^R and ecBSU1. (B) Enzyme cost of energy metabolism and oxidative phosphorylation ratio. (C) Trade-off phenomenon simulated by ecBSU1.

3.6. Enzyme-constrained integration improved the phenotype prediction

To further test the enzyme-constrained model, we simulated the growth rates of *B. subtilis* on 8 different carbon sources reported in the literature [25], and compared the prediction results of iBsu1147^R and ecBSU1 (Table S9). As shown in Figs. 5A and B, the predicted growth rates of ecBSU1 were lower than those of iBsu1147^R due to the introduction of enzymatic constraints. Especially for the two experiments using malate alone or malate and glucose as substrates, the prediction results of iBsu1147^R were unreasonably higher than the experimental results (Fig. 5A). By contrast, the predicted rates of ecBSU1 were closer to the experimental values (Figs. 5B and C). For all other carbon sources, the prediction results of ecBSU1 were also similar to or better than those from iBsu1147^R. Moreover, we identified errors in some reactions of ec_iYO844 (the first ecModel for *B. subtilis*), as 17 reactions with k_{cat} values were unidirectional, resulting in its ability to simulate growth using only glucose as a substrate (Fig. S2).

Further in-depth analysis using the ecModel revealed that the addition of the enzyme kinetic constraint information allowed ecModel to simulate the overflow of by-products from cells at high substrate concentrations. For example, the ecBSU1 results showed that it severe overflow metabolism would occur when utilizing malate at 26.51 mmol gDW⁻¹ h⁻¹, producing 16.39 mmol gDW⁻¹ h⁻¹ of acetate, thus predicting a biomass growth rate (0.618 h⁻¹) close to the experimental value [32]. However, although the growth rate was predicted accurately, the overflow products were different from the experimental results (9.50 mmol gDW⁻¹ h⁻¹ acetate and 3.93 mmol gDW⁻¹ h⁻¹ pyruvate). The difference in overflow products is mainly due to the limitations of the current ecModel, as the optimization process of the model only supports the pathway that generates the lowest enzyme cost, so that the current ecModel will only generate one overflow product.

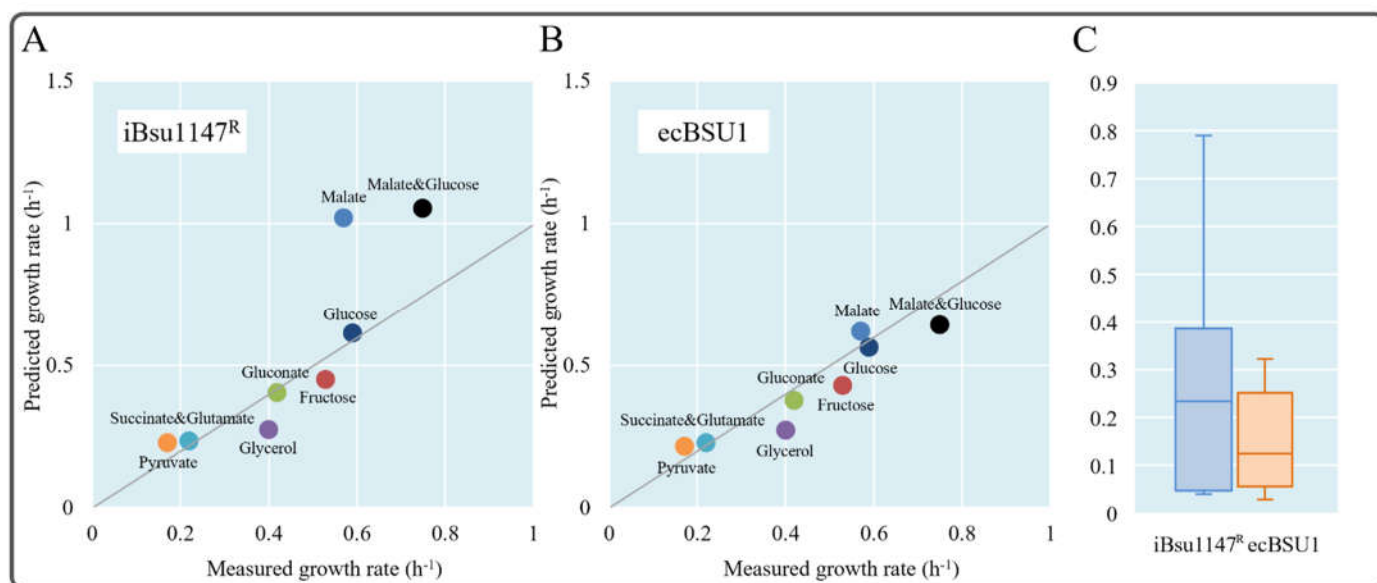


Figure 5. Predicted *B. subtilis* growth rates on different carbon sources using iBsu1147^R (A) and ecBSU1 (B). C. Distribution of prediction errors of internal fluxes from iBsu1147^R and ecBSU1.

3.7. The enzyme-constrained model predicted target genes for improving the production of chemicals

GEMs play a guiding role in predicting metabolic engineering targets. In this work, we predicted potential target genes for the synthesis of several important chemicals (e.g., riboflavin, menaquinone 7, acetoin, etc.) in *B. subtilis* based on the enzyme cost of reactions. We selected 10 products for analysis from 51 commercial chemicals produced using the *B. subtilis* platform summarized in the literature [33]. The products were classified according to the need to introduce exogenous reactions, and the location of the precursor in the central metabolic pathway (Fig. 6A). We set glucose as the substrate and fixed the growth rate at 0.06 h⁻¹ (10% of the maximum) and performed reaction enzyme cost calculations with each of these 10 products as the objective function, respectively. Subsequently, a literature search was performed to validate the top 15 reactions in terms of enzyme cost in each pathway.

We found that most of the predicted potential targets for the 10 products have been reported in the literature (Fig. 6B, Table S10). Among them, riboflavin and menaquinone 7 covered the most targets, with more than half of the predicted 15 potential targets (9 and 8, respectively) having been reported in the literature (Fig. 6B). Notably, enzymes with the highest enzyme cost in the synthetic pathways of riboflavin and uridine are GTP cyclohydrolase II (encoded by *ribA*, $\text{ru5p_D_c} \rightarrow \text{db4p_c} + \text{for_c} + \text{h_c}$) and carbamoylphosphate synthetase (encoded by *pyrAA* and *pyrAB*, $2.0 \text{ atp_c} + \text{gln_L_c} + \text{h2o_c} + \text{hco3_c} \rightarrow 2.0 \text{ adp_c} + \text{cbp_c} + \text{glu_L_c} + \text{h_c} + \text{pi_c}$), respectively (Table S9). Experiments have been performed to show that both enzymes are rate-limiting enzymes for their respective products [34-36]. For example, studies on a riboflavin production strain of *B. subtilis* showed that the insertion of an additional copy of *ribA* led to improved riboflavin titers and yields on glucose of up to 25% [34]. In addition, Wang et al. released the feedback inhibition of carbamoylphosphate synthetase encoded by *pyrAB*, leading to a 245% increase of uridine production, whereby the conversion of glucose to uridine increased by 10.5%, while overexpression of the *pyr* operon increased the production of uridine by a further 31% [35]. For 5-methyltetrahydrofolate, GTP cyclohydrolase 1 (encoded by *folE*, $\text{gtp_c} + \text{h2o_c} \rightarrow \text{ahdt_c} + \text{for_c} + \text{h_c}$) catalyzes a limiting step for the synthesis of the important precursor dihydrofolate (DHF), and co-overexpression of *folC*, *pabB*, *folE*, and *yciA* resulted in an additional 66.8-fold improvement of the 5-methyltetrahydrofolate titer, which reached 960.27 µg/L [37]. Therefore, we can speculate that the reactions with the

A

B

Production	Substrate	Enzyme cost															
		Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10	Rank11	Rank12	Rank13	Rank14	Rank15	
Riboflavin	Glucose	RibA	RibA	RibH	PurB	GlpD	RibD	RibD	—	RibE	Eno	SucC SucD	AroC	PfkA	PyrAA; PyrAB	PurH	
5-Methyltetrahydrofolate	Glucose	FolE	FabA; FabB	GlpD	—	AroC	Eno	SucC; SucD	PurB	PfkA	PyrAA; PyrAB	PurB	FabHIA; FabI	cysH	GlmM	Pyc	
Uridine	Glucose	PyrAA; PyrAB	Pyc	GlpD	—	Eno	SucC; SucD	PfkA	YnfN	YnfR	PyrC	AroC	PyrB	FbaA	PyrD; PyrK	SpoVN	
Poly-γ-glutamic	Glucose	Pyc	Eno	PfkA	GlpD	ChA	SpoVN	YnfR	FbaA	GapA	GlcK	Dut	Pyk	FumC	AroC	Pgk	
Menaquinone 7	Glycerol	MenD	MenB	AroC	MetK	IspF	IspH	Dxr	GlpD	—	Eno	Dxa	IspH	Pyc	AroA	Apt	
Thymidine	Glucose	PyrAA; PyrAB	Ndk	GlpD	Pyc	—	Eno	SucC; SucD	TrxB	PfkA	Dut	YnfB	AroC	YnfR	PyrC	FbaA	
Surfactin	Sucrose	PanB	PurB	PurB	CysH	IivC	GlpD	—	Eno	—	AccA; AccD	PanC	BfmBAA; BfmBAB	BfmBAA; BfmBAB	FabD	LeuC	
Acetoin	Glucose	AlsD	Eno	PfkA	YnfR	GlpD	PdhA; PdhB	AlsS	FbaA	GapA	GlcK	FumC	Pyk	AroC	—	Pgk	
2,3-Butanediol	Glucose	AlsD	Eno	BdhA	PfkA	PdhA; PdhB	AlsS	FbaA	YnfR	GlpD	GapA	GlcK	AroC	Pyk	FumC	Pgk	
N-acetylglucosamine	Glucose	GlnS	Eno	GlpD	PfkA	—	GlnA	GlcK	FbaA	GapA	AroC	YnfR	Pyk	Pta	Pgk	PyrAA; PyrAB	

4. Discussion

Using iBsu1147^R, we constructed ecBSU1 based on the ECMpy approach, in which 2331 of 3307 reactions were integrated with enzyme constraints according to

AutoPACMEN, and the k_{cat} coverage reached 76.4% after excluding 254 exchange reactions. The coverage of enzyme parameters was much higher than in ec_iYO844, which integrated only 17 reactions (1.67% of the total number of reactions) located in the central carbon metabolism with the addition of k_{cat} values. However, enzyme kinetic data is sparse, and measured turnover rates are normally available for only a small fraction of metabolic reactions even in well-studied organisms [38]. Therefore, even though we covered the k_{cat} data for most of the reactions in the model, only 163 reactions had k_{cat} values derived from *B. subtilis*.

Due to the diversity and incomplete coverage of enzyme parameters in the database, the initial ecModel was over-constrained, but it was able to accurately predict cellular phenotypes after k_{cat} correction for 28 reactions. The ecBSU1 corrects the problem that the growth rate of GEMs increases indefinitely with the increase of the carbon source utilization rate. Then, ecBSU1 and iBsu1147^R were respectively used to simulate the growth rate of *B. subtilis* on different substrates, and the results showed that the accuracy of ecBSU1 was much better than that of iBsu1147^R at the high growth rate stage with enzyme constraints. In addition, the overflow metabolism of *B. subtilis* was explored using ecBSU1, which showed a physiological state of overflow metabolism in the presence of excess substrates, and demonstrated a trade-off between biomass yield and enzyme usage efficiency.

Thus, ecBSU1 can be used to guide the rational design of microbial cell factories from a new perspective. The simulation results of GEMs usually only contain reaction fluxes, which cannot distinguish the physiological characteristics of each reaction. By contrast, ecBSU1 combines the kinetic characteristics of each reaction based on fluxes, thus demonstrating the enzyme consumption of each reaction, which can assist us to locate the kinetic bottlenecks of different metabolic states. We simulated the enzyme consumption of *B. subtilis* for the synthesis of several commercial chemicals, and the reactions with the highest enzyme consumption were identified as metabolic engineering targets, which was in good agreement with the literature. This provides a new strategy and theoretical basis for metabolic engineering.

Currently, ecBSU1 has a typical limitation also found in other ecModels, as the implementation of the enzyme abundance constraint is highly dependent on precise kinetic parameters and abundance data for each enzyme [14], both of which are often inadequate. Although AutoPACMEN and GECKO adopt automated strategies to supplement the missing data by fuzzy matching to similar reactions or organisms (based on the EC number and substrate), this can cause model predictions to deviate significantly from experimental observations [39]. Machine learning or deep learning tools are valuable for uncovering global trends of enzyme kinetics and physiological diversity, which can further elucidate the details of a large-scale ecModel [40, 41]. In addition, the integration of enzyme constraints greatly improves the predictive power of GEMs and brings the model predictions closer to the experimental measurements, but the biological system is too complex and to be fully described by enzymatic constraints alone. Therefore, it is necessary to integrate more biological data into novel composite constraints, which can include data on thermodynamics [42, 43] and regulatory networks [44].

5. Conclusions

This work integrated enzymatic constraints into the GEM of *B. subtilis* on a genome-wide scale, which significantly improved its metabolic phenotype prediction ability. The resulting model can be used to explain metabolic overflow phenomena and predict metabolic engineering targets for the synthesis of commercial chemicals in *B. subtilis*. This study has guiding significance for the rational design of microbial cell factories and provides an important integrated metabolic network model of *B. subtilis*. Finally, it also offers insights for the improvement of GEMs of other strains, so that the role of such models in the development of synthetic biology can be broadened in the future.

Supplementary Materials: Figure S1: Model development of *B. subtilis*; Figure S2: Growth rates of *B. subtilis* on different carbon sources predicted using ec_iYO844; Table S1: Correspondence

between subunit descriptions and subunit numbers in UniProt; Table S2: EC number modification; Table S3: GPR modifications based on the GPRuler tool; Table S4: GPR modification based on homology; Table S5: Modification of reaction boundaries; Table S6: Biomass reaction standardization and mass balance using BiomassMW; Table S7: Reactions for k_{cat} calibration; Table S8: Enzyme cost of energy metabolism; Table S9: Growth rates on different carbon sources predicted using iBsu1147^R and ecBSU1; Table S10: Target prediction based on enzyme cost.

Author Contributions: Conceptualization, K.W., Z.M. and H.M.; data curation, K.W. and Z.M.; formal analysis, K.W. and J.N.; funding acquisition, Z.W. and H.M.; methodology, K.W., Z.M. and H.M.; project administration, Z.W. and H.M.; software, K.W., Z.M. and J.N.; validation, J.N., J.C., Q.Y. and L.Y.; writing—original draft, K.W., Z.M., Z.W., X.L. and H.M.; writing—review and editing, K.W., Z.M., Y.M., Z.W., X.L. and H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2021YFC2100700), National Natural Science Foundation of China (NSFC-21621004, 32101186, 21908239), Tianjin Synthetic Biotechnology Innovation Capacity Improvement Projects (TSBICIP-PTJS-001), and the International Partnership Program of Chinese Academy of Sciences (153D31KYSB20170121).

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/ecBSU1>.

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

References

- Gu Y, Xu XH, Wu YK, Niu TF, Liu YF, Li JH, Du GC, Liu L: Advances and prospects of *Bacillus subtilis* cellular factories: From rational design to industrial applications. *Metabolic Engineering* 2018, 50:109-121.
- Hao T, Han B, Ma H, Fu J, Wang H, Wang Z, Tang B, Chen T, Zhao X: In silico metabolic engineering of *Bacillus subtilis* for improved production of riboflavin, Egl-237, (R,R)-2,3-butanediol and isobutanol. *Molecular Biosystems* 2013, 9:2034-2044.
- Vikromvarasiri N, Shirai T, Kondo A: Metabolic engineering design to enhance (R,R)-2,3-butanediol production from glycerol in *Bacillus subtilis* based on flux balance analysis. *Microbial Cell Factories* 2021, 20.
- Kalantari A, Chen T, Ji B, Stancik IA, Ravikumar V, Franjevic D, Saulou-Berion C, Goelzer A, Mijakovic I: Conversion of Glycerol to 3-Hydroxypropanoic Acid by Genetically Engineered *Bacillus subtilis*. *Frontiers in Microbiology* 2017, 8.
- Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R: Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry* 2007, 282:28791-28799.
- Henry CS, Zinner JF, Cohoon MP, Stevens RL: iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biology* 2009, 10.
- Tanaka K, Henry CS, Zinner JF, Jolivet E, Cohoon MP, Xia F, Bidnenko V, Ehrlich SD, Stevens RL, Noirot P: Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Research* 2013, 41:687-699.
- Kocabas P, Calik P, Calik G, Ozdamar TH: Analyses of extracellular protein production in *Bacillus subtilis* - I: Genome-scale metabolic model reconstruction based on updated gene-enzyme-reaction data. *Biochemical Engineering Journal* 2017, 127:229-241.
- 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, 117:3533-3544.
- 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--glutamic acid production strains. *Microbial Cell Factories* 2019, 18.
- Ishchuk OP, Domenzain I, Sánchez BJ, Muñoz-Paredes F, Martínez JL, Nielsen J, Petranovic D: Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* 2022, 119:e2108245119.
- Zhang Z-X, Wang Y-Z, Nong F-T, Xu Y, Ye C, Gu Y, Sun X-M, Huang H: Developing a dynamic equilibrium system in *Escherichia coli* to improve the production of recombinant proteins. *Applied Microbiology and Biotechnology* 2022, 106:6125-6137.
- Sanchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J: Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology* 2017, 13.
- Bekiaris PS, Klamt S: Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics* 2020, 21.

15. Mao Z, Zhao X, Yang X, Zhang P, Du J, Yuan Q, Ma H: ECMpy, a Simplified Workflow for Constructing Enzymatic Constrained Metabolic Network Model. *Biomolecules* 2022, 12.
16. Zhou J, Zhuang Y, Xia J: Integration of enzyme constraints in a genome-scale metabolic model of *Aspergillus niger* improves phenotype predictions. *Microbial Cell Factories* 2021, 20.
17. Niu J, Mao Z, Mao Y, Wu K, Shi Z, Yuan Q, Cai J, Ma H: Construction and Analysis of an Enzyme-Constrained Metabolic Model of *Corynebacterium glutamicum*. *Biomolecules* 2022, 12:1499.
18. Di Filippo M, Damiani C, Pescini D: GPRuler: Metabolic gene-protein-reaction rules automatic reconstruction. *PLoS Comput Biol* 2021, 17:e1009550.
19. Seaver SMD, Liu F, Zhang Q, Jeffries J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber Moritz E, et al: The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research* 2020, 49:D575-D588.
20. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE: BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* 2015, 44:D515-D522.
21. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblit J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D: BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research* 2020, 49:D498-D508.
22. Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Algaa E, Weidemann A, Sauer-Danzwith H, Mir S, et al: SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Research* 2011, 40:D790-D796.
23. Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C: Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *PROTEOMICS* 2015, 15:3163-3168.
24. 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. *Nature Communications* 2022, 13:3766.
25. Chubukov V, Uhr M, Le Chat L, Kleijn RJ, Jules M, Link H, Aymerich S, Stelling J, Sauer U: Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Molecular Systems Biology* 2013, 9.
26. van Hoek MJA, Merks RMH: Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Systems Biology* 2012, 6:22.
27. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al: The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics* 2017, 20:1085-1093.
28. Chan SHJ, Cai J, Wang L, Simons-Senftle MN, Maranas CD: Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics* 2017, 33:3603-3609.
29. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO: A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Molecular Systems Biology* 2011, 7.
30. Lu X, Liu Y, Yang Y, Wang S, Wang Q, Wang X, Yan Z, Cheng J, Liu C, Yang X, et al: Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. *Nature Communications* 2019, 10:1378.
31. Dauner M, Storni T, Sauer U: *Bacillus subtilis* Metabolism and Energetics in Carbon-Limited and Excess-Carbon Chemostat Culture. *Journal of Bacteriology* 2001, 183:7308-7317.
32. Kleijn RJ, Buescher JM, Le Chat L, Jules M, Aymerich S, Sauer U: Metabolic Fluxes during Strong Carbon Catabolite Repression by Malate in *Bacillus subtilis*. *Journal of Biological Chemistry* 2010, 285:1587-1596.
33. Zhang Q, Wu Y, Gong M, Zhang H, Liu Y, Lv X, Li J, Du G, Liu L: Production of proteins and commodity chemicals using engineered *Bacillus subtilis* platform strain. *Essays in Biochemistry* 2021, 65:173-185.
34. Hümbelin M, Griesser V, Keller T, Schurter W, Haiker M, Hohmann HP, Ritz H, Richter G, Bacher A, van Loon APGM: GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone 4-phosphate synthase are rate-limiting enzymes in riboflavin synthesis of an industrial *Bacillus subtilis* strain used for riboflavin production. *Journal of Industrial Microbiology and Biotechnology* 1999, 22:1-7.
35. Wang Y, Ma R, Liu L, He L, Ban R: Improvement of uridine production in *Bacillus subtilis* by metabolic engineering. *Biotechnology Letters* 2018, 40:151-155.
36. Lehmann M, Degen S, Hohmann H-P, Wyss M, Bacher A, Schramek N: Biosynthesis of riboflavin. *The FEBS Journal* 2009, 276:4119-4129.
37. Yang H, Liu Y, Li J, Liu L, Du G, Chen J: Systems metabolic engineering of *Bacillus subtilis* for efficient biosynthesis of 5-methyltetrahydrofolate. *Biotechnology and Bioengineering* 2020, 117:2116-2130.
38. Davidi D, Milo R: Lessons on enzyme kinetics from quantitative proteomics. *Current Opinion in Biotechnology* 2017, 46:81-89.
39. Kerkhoven EJ: Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints. *Current Opinion in Microbiology* 2022, 68:102168.
40. 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. *Nature Communications* 2018, 9.
41. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J: Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis* 2022.
42. 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.
43. Salvay P, Hatzimanikatis V: The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nature Communications* 2020, 11.

-
44. Shen F, Sun R, Yao J, Li J, Liu Q, Price ND, Liu C, Wang Z: OptRAM: In-silico strain design via integrative regulatory-metabolic network modeling. *PLoS Computational Biology* 2019, 15:e1006835.