Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Metabolic optimization of *Vibrio natriegens* based on metaheuristic algorithms and the genome-scale metabolic model

Yixin Wei[a,b], Tong Qiu[a,b,*], and Zhen Chen[a]

[a] Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
[b] State Key Laboratory of Chemical Engineering, Tsinghua University, Beijing 100084, China
* Corresponding Author: qiutong@tsinghua.edu.cn.

ABSTRACT

In recent years, burgeoning interest in products derived from microbial production across various sectors has significantly propelled the evolution of the field of metabolic engineering. As a Gram-negative bacterium, *Vibrio natriegens* is characterized by its fast growth, robust metabolic capabilities, and a broad substrate spectrum, making it a promising candidate as a standard biological host for the industrial bioproduction of metabolites. Genome-scale metabolic models (GSMMs) are mathematical representations constructed based on genome annotations and gene-protein-reaction (GPR) associations within a cell. These models enable the computational simulation of intracellular reaction flux distributions. In this study, we developed a hybrid method based on metaheuristic algorithms and the GSMM to optimize metabolism for the production of ethanol and 1,3-propanediol (1,3-PDO) as target products in *Vibrio natriegens*. The modified GSMM used in this study contains 1195 reactions, 1094 metabolites, and 880 genes, with the metaheuristic algorithms employed being Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The optimization results indicate that the method proposed in this study can achieve the production of ethanol and 1,3-PDO by adjusting the expression levels of just 2-4 genes, with the production of 1,3-PDO reaching its theoretical maximum. The strategies developed in this study can effectively increase the production capacity of specific target metabolites in *Vibrio natriegens*, serving as a starting point for metabolic engineering and providing guidance for metabolic engineering targets in practical experiments.

Keywords: *Vibrio natriegens*, Metabolic optimization, Genome-scale metabolic model, Metaheuristic algorithm

INTRODUCTION

In recent years, there has been a growing interest across various industries in products synthesized by microorganisms, which has spurred the development of metabolic engineering. The natural pathways of natural product hosts may face issues such as low efficiency, slow production rates, and susceptibility to interference. With the advancement of synthetic biology, scientists have begun to perform local or global metabolic engineering modifications on wild-type host chassis based on specific target products. This leads to the creation of chassis cells with improved traits and competitive advantages. Specifically, metabolic engineering aims to increase the yield of specific target compounds as much as possible by identifying the optimal design of microbial cell factories [1]. Metabolic engineering techniques allow for precise regulation of gene expression levels in engineered bacteria, with flux simulation, computation, and optimization driving strain modification towards greater efficiency and precision. *Escherichia coli*, *Saccharomyces cerevisiae*, and *Corynebacterium glutamicum* stand as the most prevalent biological hosts for constructing these cell factories. *Vibrio natriegens* is a Gram-negative bacterium known for its remarkable growth rate, robust metabolic capabilities, and a broad substrate spectrum, holding promise as a prospective standard biotechnological host for laboratory and industrial bio-production,

specifically tailored to produce target metabolites [2].

The advancement of genome sequencing technology and the gradual completion of biochemical databases have driven the development of GSMMs, establishing them as an important tool for studying cellular metabolic characteristics. GSMMs are mathematical representations of cellular models that originate from the cell's annotated genomic information. They construct a matrix of all the metabolic reactions in the cell based on GPR associations. GSMMs effectively integrate the intracellular metabolites, enzyme-catalyzed metabolic reactions, enzymes, and the genes responsible for enzyme expression, thus constructing complex models of metabolic networks. Currently, GSMMs are extensively utilized in the field of metabolic engineering for the computational tasks, simulations, and analyses of cells. They can predict cellular growth under various gene expression conditions and environmental changes, assessing the global distribution of metabolic fluxes, and determining the production flux of specific metabolites, making them vital tools for gene essentiality analysis and the prediction of cellular viability. GSMMs have been established for various model microorganisms, with continuous improvements being made to the number of genes, reactions, and metabolites included in the models, as well as the constraints applied, all of which enhance the accuracy of the models. In 2023, Coppens et al. [2] developed the first GSMM for *Vibrio natriegens*, which showed good consistency with experimental data. With the increasing convergence of computer science and bioinformatics, and considering the efficiency and effectiveness of metaheuristic algorithms in finding global optima, scientists have begun to apply metaheuristic algorithms to the analysis of GSMMs for hosts such as *Escherichia coli* [3].

In this study, we propose a hybrid method for the metabolic optimization of *Vibrio natriegens*. We first modified the GSMM for *Vibrio natriegens* developed by Coppens et al., and then we combined it with different metaheuristic algorithms, including GA and PSO, to identify the optimal gene expression level optimization strategies encompassing both knockouts and underexpression, as well as overexpression, for maximizing production fluxes of ethanol and 1,3-PDO, our two target products. Integrating machine learning with traditional metabolic engineering can aid in a more efficient discovery of better metabolic optimization strategies and facilitate a more comprehensive understanding of the reasons why different metabolic optimization strategies are conducive to increasing the yield of target metabolites. The optimization results suggest that the hybrid method can effectively enhance the production capabilities of specific targeted metabolites in *Vibrio natriegens*, offering strategic guidance for target selection in metabolic engineering modifications for practical experiments.

PROBLEM STATEMENT

The mathematical expression of metabolic engineering strategy optimization

The optimization of metabolic engineering strategies based on GSMM is a bi-level optimization problem. The inner layer consists of a linear programming problem, also known as the simulation of the GSMM, which aims to optimize an objective function defined as the linear combination of the reaction fluxes within the network (for example, the flux of the biomass production reaction). This is achieved by employing equality constraints derived from the quasi-steady-state assumption in the cell, along with inequality constraints that impose direct limits on the maximum and minimum values of the reaction fluxes. The inner layer provides the distribution of reaction fluxes within the metabolic network, as well as the maximum value of the corresponding objective function. Building on the distribution, the outer optimization introduces various gene expression optimization strategies, with objective functions such as Biomass-Product Coupled Yield (BPCY) and Weighted Yield (WYIELD) that are related to the target product production flux. Gene expression strategies encompass gene knockout, underexpression, and overexpression, which impact the inequality constraints of the inner layer optimization. This, in turn, affects the distribution of network fluxes and, as a result, the optimization results as well as the value of the objective function in the outer optimization.

The bi-level optimization problem can be mathematically represented as equations 1-4.

$$F(\boldsymbol{strategy}) = \max\{BPCY(\boldsymbol{v}),\ WYIELD(\boldsymbol{v}),\ etc.\} \quad (1)$$

$$Z = \max\{\boldsymbol{C}^{\mathrm{T}} \cdot \boldsymbol{v}\} \quad (2)$$

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0 \quad (3)$$

$$\boldsymbol{lb}(\boldsymbol{strategy}) \le \boldsymbol{v} \le \boldsymbol{ub}(\boldsymbol{strategy}) \quad (4)$$

Where $\boldsymbol{v}$ is the flux vector composed of the fluxes of all reactions in the metabolic network, with a dimension of N, which corresponds to the number of reactions in the model; $\boldsymbol{C}^{\mathrm{T}}$ is the transposed vector of the integer coefficient of each flux in the objective function Z; $\boldsymbol{S}$ (M×N) is the stoichiometric matrix (S_{ij} corresponds to the stoichiometric coefficient of metabolite i in reaction j); $\boldsymbol{ub}$ and $\boldsymbol{lb}$ record the upper and lower bounds for the flux of each reaction in the model, respectively; $\boldsymbol{strategy}$ indicates the gene expression optimization strategy, where each dimension corresponds to a gene in the model, and each gene corresponds to an enzyme (A $strategy_i$ value of 0 signifies that the gene is knocked out, a value between 0 and 1 signifies underexpression, and a value greater than 1 signifies overexpression); $BPCY$ is the product of biomass and the yield of the target product, whereas $WYIELD$ is the weighted sum of the minimum and

maximum product fluxes under a fixed growth rate, with the weight for the maximum product flux set at 0.3.

Impact of gene expression optimization strategies on flux constraints

Changes in gene expression levels affect the concentration of enzymes in the system, thereby influencing the flux constraints of reactions. Figure 1 is a schematic diagram illustrating how gene expression optimization strategies affect reaction flux constraints.

In a GSMM, each gene expresses a protein (enzyme) that catalyzes a reaction; however, enzymes and reactions do not correspond on a one-to-one basis. There may be multiple enzymes that work together to catalyze a single reaction (cooperative), or several enzymes that can each independently catalyze a particular reaction (competitive). Before determining the new constraints for each reaction, the following calculations need to be performed: (1) Calculate the reaction fluxes of the initial model (wild type) as a $\boldsymbol{reference}$; (2) Calculate the values of the optimization strategies for each reaction. For cooperative genes (G3 and G4), the corresponding reaction optimization strategy is taken as the minimum of the gene strategies, whereas for competitive genes (G2 and G3), the reaction optimization strategy is the maximum of the gene strategies, as shown in Figure 1. Based on the optimization strategy ($strategy_i$) and reference flux value ($reference_i$) for each reaction, updating the constraint of the reaction flux involves two steps. (1) Determine the direction of the reaction: if $reference_i > 0$ (forward reaction), set the lower bound of the reaction to 0; if $reference_i < 0$ (backward reaction), set the upper bound of the reaction to 0. (2) Update the constraint according to the direction of the reaction and the optimization strategy. For knockouts (strategy equals 0), set both the upper and lower bounds of the reaction to 0; for underexpression (strategy between 0 and 1), underexpressing a forward (backward) reaction is achieved by setting the upper (lower) bound to $strategy_i \times reference_i$; for overexpression, overexpressing a forward (backward) reaction involves setting the lower (upper) bound to $strategy_i \times reference_i$.

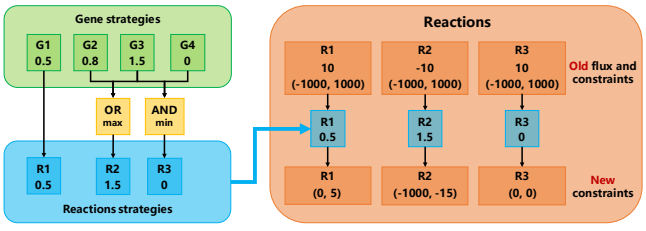

Figure 1. The relationship between reaction flux constraints and gene expression optimization strategies.

MATERIALS AND METHODS

Dataset and experiment setup

We utilized the only published GSMM of *Vibrio natriegens* in SBML format and made the following modifications to it. We adjusted the upper and lower bounds of the glycerol demand reaction in the model, enabling the model to simulate cell culture processes with glycerol as the sole carbon source. We controlled the oxygen input in the model to be greater than zero (aerobic cultivation). We also removed three reactions that were not directly adjacent to the main reaction network, along with the corresponding six metabolites and three genes. Additionally, since the metabolic network of the wild-type *Vibrio natriegens* does not contain 1,3-PDO, we introduced a two-step heterologous pathway for the conversion of glycerol to 1,3-PDO via 3-hydroxypropionic acid, ensuring that the model could be used for metabolic optimization of 1,3-PDO production. The modified GSMM comprises 1195 reactions, 1094 metabolites, and 880 genes. We employed parsimonious Flux Balance Analysis (pFBA) [4] to simulate each phenotype during the optimization process. The pFBA method can return a flux distribution result where the total flux of all metabolic reactions is minimized while still maximizing the objective function. GA and PSO were used to perform metabolic optimization with ethanol and 1,3-PDO as the target products of cellular production. All simulations and optimizations of the GSMMs were conducted in Python, utilizing the Cobrapy [5] and Mewpy [6]. The hyperparameter settings for the GA and PSO are shown in Table 1. In both algorithms, we limited the maximum number of genes with adjusted expression levels to 6.

Table 1. Hyperparameter settings for GA and PSO.

Algorithm	Hyperparameter	Value
GA	Number of generations	100
	Others	Default
PSO	Number of particles	2500
	Number of iterations	100
	Inertia weight	1.5
	Personal learning rate	2.1
	Global learning rate	1.6
Both	Maximum number of genes with adjusted expression levels	6
	Fold of expression levels	{0.125,0.25, 0.5,1,2,4,8}

Description of the hybrid metabolic optimization method

Simulation methods such as pFBA can determine the global distribution of reaction fluxes under various gene expression strategies, but they do not provide guidance on how to optimize these strategies. Given that the relationship between the production flux of biomass and

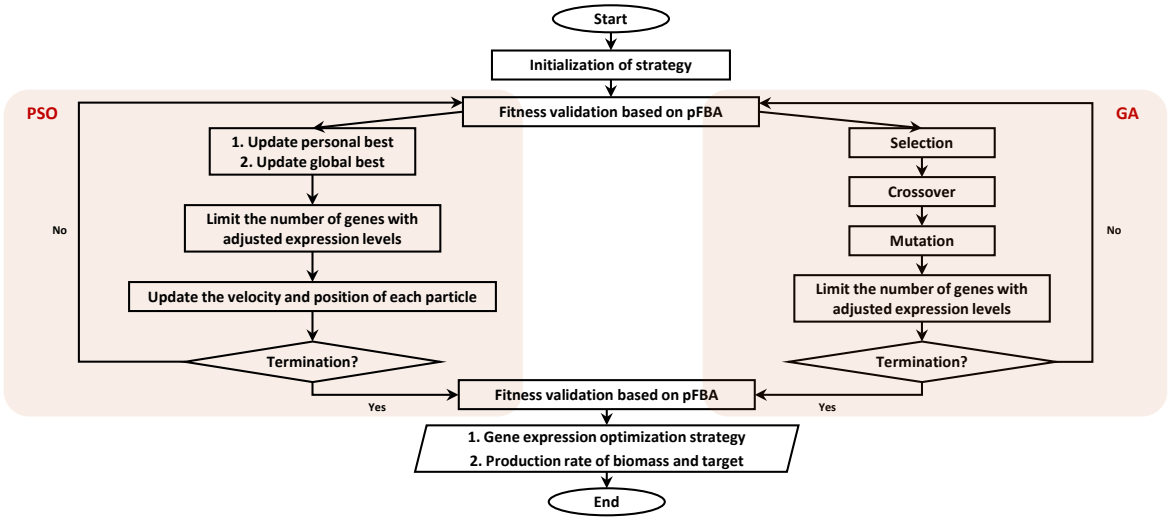

Figure 2. Flowchart of the hybrid optimization method.

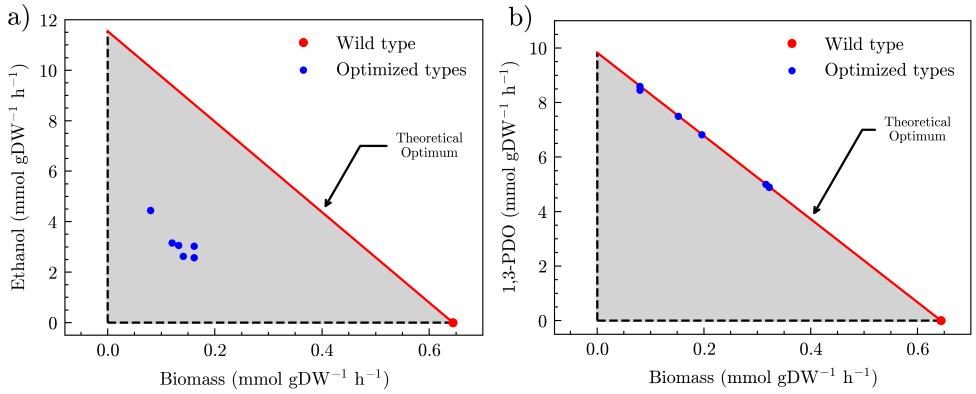

Figure 3. Visualization of metabolic optimization for a) ethanol production, and b) 1,3-PDO production.

the target product is not a simple function of the strategy vector, and that calculating the gradient is challenging due to the complexity of the system, coupled with the extensive search space for the strategy vector, it is particularly appropriate to employ metaheuristic algorithms to find more optimal values of the strategy vector. The metabolic optimization method proposed in this paper has explored the use of different metaheuristic algorithms to control the update of the strategy vector. The pFBA method is employed to validate whether this strategy vector can indeed optimize the production rate of the target products while ensuring cell viability. Figure 2 is a flowchart of the hybrid optimization method. The method ultimately outputs both the optimal strategy vector and the corresponding production rate of biomass and target product.

RESULTS AND DISCUSSION

Metabolic optimization was carried out for ethanol and 1,3-PDO as separate target products, with glycerol being used as the sole carbon source for cell growth and its uptake rate controlled at 10 mmol gDW^{-1} hr^{-1}. The experiment found that it was not possible to increase the production rate of the product solely by gene knockout, likely because the competing pathways for intermediates are essential for cell growth. However, through optimization, some feasible sites for gene overexpression and underexpression were identified. These strategies demonstrated effective increases in the production rate of the target product. Table 2 shows the results of the optimized gene expression strategies, where each row corresponds to a strategy, including genes that are knocked out, overexpressed, or underexpressed, along with the associated strategy values. In addition, the heuristic algorithm corresponding to each strategy, the biomass, target production rate, and the values of the objective functions were also calculated and recorded.

Ethanol is one of the typical fermentation products, generally associated with acetyl-CoA in the metabolic network. Figure 3a illustrates the cell biomass and ethanol production corresponding to strategies obtained by metabolic optimization. The gray triangular area in Figure 3 represents the theoretically feasible combinations of biomass and target production, while the red line indicates the theoretical maximum target production for

Table 2. Metabolic optimization results.

Target product	Algorithm	Overexpression		Underexpression and knock out		BPCY[c]	WYIELD[c]	Biomass[c]	Production rate[c]
		Id[a,b]	Strategy	Id[a,b]	Strategy				
Enthnol	/	/	/	/	/	0	0	0.644	0
	GA	14435	2	02835	0.125	0.358	2.908	0.080	4.446
		16675	8	10485	0.25	0.378	2.366	0.120	3.152
		01145	8						
		03645	8						
		01145	8	10810	0.25	0.404	2.347	0.132	3.058
		03645	8						
		16675	4	10485	0.25	0.370	2.477	0.141	2.624
		01145	8						
		23080	4						
		01145	8	10485	0.25	0.414	2.207	0.161	2.574
		03645	8						
	PSO	00270	2	10255	0.25	0.488	1.567	0.161	3.032
		02935	8	01945	0				
1,3-PDO	/	/	/	/	/	0	0	0.644	0
	GA	11695	2	00900	0.125	0.691	2.584	0.080	8.594
		/	/	00900	0.125	0.680	2.584	0.080	8.450
		21510	2	01955	0.5	1.338	5.553	0.196	6.823
		11695	2	08360	0.25				
		/	/	01955	0.25	1.144	7.507	0.152	7.500
		/	/	01955	0.5	1.581	5.014	0.316	5.000
		11695	4	09270	0	1.578	1.489	0.322	4.904
				09740	0.5				
	PSO			00175	0.5	1.574	1.489	0.322	4.891
				01340	0.25				

a: In *Vibrio natriegens*, the prefix for all gene ids is "PN96_". For example, the gene corresponding to id: 14435 would be "PN96_14435".
b: In the table, each grid may contain multiple Ids, indicating that the expression levels of genes corresponding to each Id will be controlled in that strategy.
c: BPCY is the product of biomass and the production rate of the target product. The units for WYIELD, biomass, and production rate are all mmol $gDW^{-1}\,hr^{-1}$.

each biomass level. The red dot represents the wild type corresponding to the original GSMM, and each blue dot corresponds to a strategy listed in Table 2. Figure 3a shows that the optimized GSMM can demonstrate ethanol production, and there is a negative correlation between ethanol production and biomass. However, there is still a certain distance from the theoretical maximum production of ethanol, which may be due to competition from other fermentation products such as acetate.

The maximum production strategy for ethanol from GA involves the overexpression of the gene PN96_14435 (gpmM), which increases the flux from glycerate-3-phosphate to glycerate-2-phosphate, thereby enhancing the production of pyruvate and acetyl-CoA, consequently benefiting the production of ethanol. The strategy also involves the underexpression of the gene PN96_02835 (aroC), which leads to a reduction in the production of branch acids, attenuating the biosynthesis of phenylalanine, tyrosine, and tryptophan, thereby diverting flux into glycolysis, which in turn facilitates the production of ethanol. The strategy provided by PSO involves the overexpression of genes PN96_00270 (lysC) and PN96_14435 (gpmM). The former enhances the phosphorylation of aspartate, which is beneficial for the generation of fermentation products, while the latter is the same as in the GA strategy. The strategy also underexpresses gene PN96_10255 (ppnK) and knocks out gene PN96_01945 (glpD). The former regulates the intracellular redox balance, while the latter ensures that glycerol-3-phosphate is converted into glycerone-phosphate only through interactions with NADP+ or NAD+, also regulating the intracellular redox balance, thus promoting the production of ethanol.

Unlike ethanol, the production of 1,3-PDO has been introduced into the metabolic network as an exogenous pathway, competing with the accumulation of cell biomass directly for the carbon source glycerol. Figure 3b visualizes the biomass and 1,3-PDO production

corresponding to the optimized strategies, demonstrating that all strategies in Table 2 can achieve (or nearly achieve) the theoretical maximum production of 1,3-PDO while sacrificing a certain amount of biomass.

Compared to ethanol, the strategy for targeting 1,3-PDO as the product is simpler, requiring the underexpression of as few as one gene. For example, by reducing the expression of the gene PN96_00900 (pgk), the flux from 1,3-bisphosphoglycerate to 3-phosphoglycerate can be decreased, thereby weakening the glycolytic process and the accumulation of biomass, ultimately favoring the diversion of the carbon source towards 1,3-PDO. Building on this, further overexpression of gene PN96_11695 (mdh) can enhance the conversion of L-glutamate to L-aspartate, increasing the production of 1,3-PDO without reducing biomass.

In fact, due to the complexity of the GSMM and the metabolic network, there are multiple different gene strategies that can exhibit the same optimization effect. When selecting a strategy, one should comprehensively consider its biomass, stability (WYIELD), and cost (number of gene edits), which is also the reason why multiple strategies are listed in Table 2. Although the actual system may not conform to the quasi-steady-state assumption of the current simulation and optimization, with various intermediates possibly accumulating within the cells and affecting viability and target production, the existing strategies are valuable references for actual experiments and require subsequent experimental verification to confirm their practical effects.

To further enhance the precision of GSMM simulations and the reliability of the optimization strategy results of the hybrid method, future research may include: (1) further optimization of the GSMM, including more GPR associations and improved data annotation; (2) adding enzyme kinetics and thermodynamic constraints to the optimized GSMM to refine the feasible flux space and increase the accuracy of the GSMM simulations; (3) trying more algorithms, such as combining GA and PSO, to leverage PSO's strengths in locating and exploring local optima and GA's enhanced capability to search the global space; (4) further incorporating knowledge into the optimization process to avoid completely random searches and improve the interpretability of the strategies.

CONCLUSIONS

In this study, we combined machine learning with traditional metabolic engineering, and developed a hybrid method for metabolic optimization of *Vibrio natriegens*. We utilized the modified GSMM of the *Vibrio natriegens*, integrated with two heuristic algorithms, GA and PSO, to perform metabolic optimization with the endogenous metabolite ethanol and the exogenous metabolite 1,3-PDO as target products, respectively. The optimization process used glycerol as the sole carbon source, with its uptake flux kept constant, and the gene expression level optimization strategies included knockouts, underexpression, and overexpression. The optimization results indicate that the GSMM, optimized with the hybrid method developed in this study, can exhibit production of either ethanol or 1,3-PDO by adjusting the expression levels of only 2-4 genes, with a negative correlation between the production of the target products and biomass. The production of ethanol has not reached the theoretical optimum due to the competitive fermentation products such as acetate, while the production of 1,3-PDO can reach the theoretical maximum due to its direct competition for the carbon source. The strategies obtained in this study can be served as a starting point for metabolic engineering, providing guidance for metabolic engineering targets in practical experiments.

REFERENCES

1. Bai L, You Q, Zhang C, et al. Advances and applications of machine learning and intelligent optimization algorithms in genome-scale metabolic network models. *Syst. Microbiol. Biomanuf.* 3:193-206 (2023)
2. Coppens L, Tschirhart T, Leary DH, et al. *Vibrio natriegens* genome-scale modeling reveals insights into halophilic adaptations and resource allocation. *Mol. Syst. Biol.* 19:e10523 (2023)
3. Lee MK, Mohamad MS, Choon YW, et al. A Hybrid of Particle Swarm Optimization and Minimization of Metabolic Adjustment for Ethanol Production of *Escherichia Coli*. In: Practical Applications of Computational Biology and Bioinformatics, 13th International Conference. Ed: Fdez-Riverola, F., Rocha, M., Mohamad, M., Zaki, N., Castellanos-Garzón, J. Springer (2019)
4. Lewis NE, Hixson KK, Conrad TM, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6:390 (2010)
5. Ebrahim A, Lerman JA, Palsson BO, et al. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* 7:74 (2013)
6. Pereira V, Cruz F, Rocha M. MEWpy: a computational strain optimization workbench in Python. *Bioinformatics* 37:2494-2496 (2021)