Int. J. Advance Soft Compu. Appl, Vol. 10, No. 2, July 2018
ISSN 2074-8523

A hybrid of Differential Search Algorithm and Flux Balance Analysis to

Identify Knockout Strategies for in silico Optimization of Metabolites Production

Kauthar Mohd Daud[1], Zalmiyah Zakaria[1], Zuraini Ali Shah[1],Mohd Saberi Mohamad[2], Safaai Deris[2], Sigeru Omatu[3], Juan Manuel Corchado[4]

[1]Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai 81310, Malaysia.
e-mail: kautharmohdaud@yahoo.com, zalmiyah@utm.my, aszuraini@utm.my

[2]Faculty of Creative Technology & Heritage, Universiti Malaysia Kelantan, Locked Bag 01, 16300 Bachok, Kota Bharu, Kelantan, Malaysia.
e-mail: saberi@umk.edu.my, safaai@umk.edu.my

[3]Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan.

[4]Biomedical Research Institute of Salamanca/BISITE Research Group, University of Salamanca, Salamanca, Spain.

Abstract

An increasing demand of naturally producing metabolites has gained the attention of researchers to develop better algorithms for predicting the effects of reaction knockouts. With the success of genome sequencing, in silico metabolic engineering has aided the researchers in modifying the genome-scale metabolic network. However, the complexities of the metabolic networks, have led to difficulty in obtaining a set of knockout reactions, which eventually lead to increase in computational time. Hence, many computational algorithms have been developed. Nevertheless, most of these algorithms are hindered by the solution being trapped in the local optima. In this paper, we proposed a hybrid of Differential Search Algorithm (DSA) and Flux Balance Analysis (FBA), to identify knockout reactions for enhancing the production of desired metabolites. Two organisms namely Escherichia coli and Zymomonas

mobilis were tested by targeting the production rate of succinic acid, acetic acid, and ethanol. From this experiment, we obtained the list of knockout reactions and production rate. The results show that our proposed hybrid algorithm is capable of identifying knockout reactions with above 70% of production rate from the wild-type.

Keywords: *knockout strategy, flux balance analysis, optimization, genome-scale metabolic network, in silico metabolic engineering*

1 Introduction

Metabolic engineering is defined as the process of improving the cellular activities by manipulating enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology. The purpose of metabolic engineering is to improve the microbial strain in order to economically and industrially produces the desired metabolites. The retrofitting of a metabolism of a bacteria is done in the traditional techniques such as random mutagenesis and screening [1]. These traditional techniques, however, is irreversible. Undoubtedly, it is time-consuming, expensive and laborious. Another disadvantage of the traditional method is that it requires prior knowledge of the genetic manipulations, which eventually makes it tedious and challenging.

Thus, with the breakthrough of full genome sequencing, the construction of genome-scale model (GSMM) of a particular organism has become possible. Eventually, it enables the genetic modifications to be carried out easily, therefore the consequences of genetic modifications and phenotypic behavior can be predicted. These GSMMs have initiated a new interdisciplinary in metabolic engineering known as *in silico* metabolic engineering. The aforementioned limitations of traditional method can be significantly improved through the application of *in silico* metabolic engineering. The advantages of *in silico* compared to *in vitro* or *in vivo* experiments, the former method are easy, less equipment, and costs needed, to name a few.

However, the complexity of metabolic network has resulted in the dimensions of solutions space to be large and thus it is difficult to obtain a near-optimal set of reactions to be knockout in order to enhance the production of desired metabolites [2]. Due to the complexity of the dataset, the computational time increases exponentially. In this research, a hybrid of Differential Search algorithm (DSA) and Flux Balance Analysis (FBA) were proposed and developed to simulate and identify a set of reactions knockout towards the overproduction of desired metabolites. Three GSMMs, which are *Escherichia coli* (iAF1260 and core model) and *Zymomonas mobilis* (iEM439) were used to test the developed hybrid algorithm. It focuses on the organism's metabolism and phenotypic characteristics by optimizing the production rate of succinic acid, lactic acid, acetic acid and ethanol, while maintaining the growth rate of organisms after the perturbation.

The simulations results obtained by DSAFBA were compared with previous developed algorithms.

The paper is categorized as follows: Section 2 discusses the related works pertaining to the constraint based approaches. Section 3 describes the details of problem formulation. Section 4 describes the proposed method, while Section 5 presents the results and discussion. Lastly, Section 6 provides the conclusion and future work of the study.

2 Related Work

In previous years, many approaches have been developed to redesign and analyze the metabolic network for improving the production of desired metabolites [1,3]. One of the approaches is a constraint-based modeling that represents the metabolic network in mathematical format by imposing a set of constraints on phenotype. It captures the relationships of genotype-phenotype; therefore, it enables the researchers to predict the outcomes of organisms after perturbations such as products and growths [1]. The earliest algorithm developed in 2003, OptKnock, is based on the bi-level linear optimization, where mixed integer linear programming (MILP) is used to identifies knockout strategies that lead to the maximum production of metabolites and maximum growth rate [4].

Later, OptKnock derived algorithm, named RobustKnock is developed to predict the reactions deletion strategies for overproducing the production rate of target metabolites. This method applies max-min optimization framework with consideration of the presence of competing pathways in the network [5]. Another algorithm that extended from OptKnock is OptGene. It uses a Genetic Algorithm (GA) and Evolutionary Algorithm (EAs) for formulating *in silico* design problems with flexible objective and determining the possible best-set number of reactions knockout [1]. Another recently developed algorithm, named as IdealKnock, has been developed for identifying knockout strategies by searching the whole genome metabolic model without restricting the number of knockout [6].

The aforementioned methods are mainly centralize on flux balance analysis (FBA) for modeling and bi-level optimization for optimizing the knockout strategies. Biologically nature inspired evolutionary algorithms such as ant colony, bees algorithm, simulated annealing, and particle swarm optimization has been used in obtaining the near-optimal set of knockout [7–9]. The search strategy of these algorithms, use more than one particle for searching the solution space, thus it is well known among the researchers to apply them for finding the near-optimal reactions knockout strategies.

However, these aforementioned methods are comparatively weak in local search, does not guarantee an optimal solution or represent sub optimal solution, and computationally expensive [1,10,11]. A detailed review focusing on

computational constraint based optimization methods together with the real applications has been discussed [12].

3 Problem Formulations

The problem of identifying a set of reactions knockout from the GSMM can be formulated mathematically by representing the GSMM in stoichiometry matrix (*S*) that consists of reactions and metabolites (size of $m \times n$). The stoichiometry matrix describes the dynamic mass balance equation of the GSMM by a differential equation between the flux rates of the reactions (*v*) and concentrations of metabolites (*c*). The mass balance must be at a steady-state level that signifies there is no internal and external change in metabolite concentration [13]. Assume that there are *m* metabolites and *n* reactions:

Flux vector, *V* of length *n*:

$$v = (v_1, v_2, \cdots v_n) \tag{1}$$

Concentration vector of reactions, *c* of length *m*:

$$c = (c_1, c_2, \cdots c_m) \tag{2}$$

From the above equations, the dynamic mass balance equation with respect to time can be mathematically represented as:

$$\frac{dc}{dt} = S.v \tag{3}$$

Where *t* represents the time, *v* is the flux rates of reaction, *m* is the number of reactions, *n* represents the number of metabolites, *c* is the concentration vector of reactions, and *S* is the stoichiometric matrix.

The dynamic mass balance equation describes the changes of concentration of each metabolite, *c*, over time. This study focuses on finding the optimal set of reactions knockout in order to enhance the production rate of desired metabolites while maintaining the growth rate of the model. In doing so, FBA seeks the optimal points for maximizing the objective function, $Z = c^{\mathrm{T}}v$ where *c* is a weight vector of reactions towards the objective function. Meanwhile, *v* is the flux vector and *S* is the stoichiometric matrix. This optimization is accomplished with the uses of linear programming and sets of lower and upper bounds as constraints on *v*, while maintaining the system of mass balance equation at steady state, which is $Sv = 0$.

4 Methodology

This section describes the proposed method, hybrid of Differential Search Algorithm and Flux Balance Analysis, termed as DSAFBA, followed by the datasets being used.

4.1 Differential Search Algorithm

DSA is one of the swarm-based metaheuristics algorithm inspired by the migration behavior of organisms towards the fruitful condition. The migration behavior mimics the *Brownian-like random walks* movement [14]. The migration process consists of a group of individuals with a superorganism. The superorganism moves from one habitat to another habitat in order to find a new habitat that is favorable to them. If the new habitat is able to satisfy the needs, the superorganism will reside at the new habitat named as a stopover site for the certain period of time. The migration and movement process continues from the previous site and moves towards more fruitful habitat. The exploration and exploitation of DSA are carried out by Brownian-like random walks. Previously, however, most optimization applications using DSA were focusing on solving continuous problem [14–16]. Thus, the conventional DSA has been modified to cater for binary problem as the domain problem focused in this research is binary.

4.2 The proposed hybrid of Differential Search Algorithm and flux balance analysis (DSAFBA)

We proposed DSAFBA in identifying the near-optimal set of reactions to be knocked out with the purpose of enhancing the production of desired metabolites. DSA algorithm uses multiple random numbers in the process of generating new artificial-organisms and selects random artificial-organisms, which lead the algorithm to diverge from local optimum. Meanwhile, FBA is used to calculate the fitness score of the solutions.

The proposed algorithm, DSAFBA is the combination of optimization algorithm and modeling method. The difference between proposed algorithm and previous algorithms is the search strategy of DSA. Since DSA simultaneously uses more than one individual in determining the global optimum; therefore, it has no preference to the best possible solution directly to the problem. As shown in Fig. 1, the hybrid method, DSAFBA is divided into two parts. First part is the initialization and fitness evaluation, while the second part is exploration and exploitation. There are two improvements being made: (1) the constraint based modeling, FBA is used as the fitness function and hybrid into DSA optimization algorithm as depicted by dashed box and (2) modified transfer *Tanh* function has been applied to generate binary numbers of 0 and 1 as standard DSA is only applicable to the continuous problem [17,18].

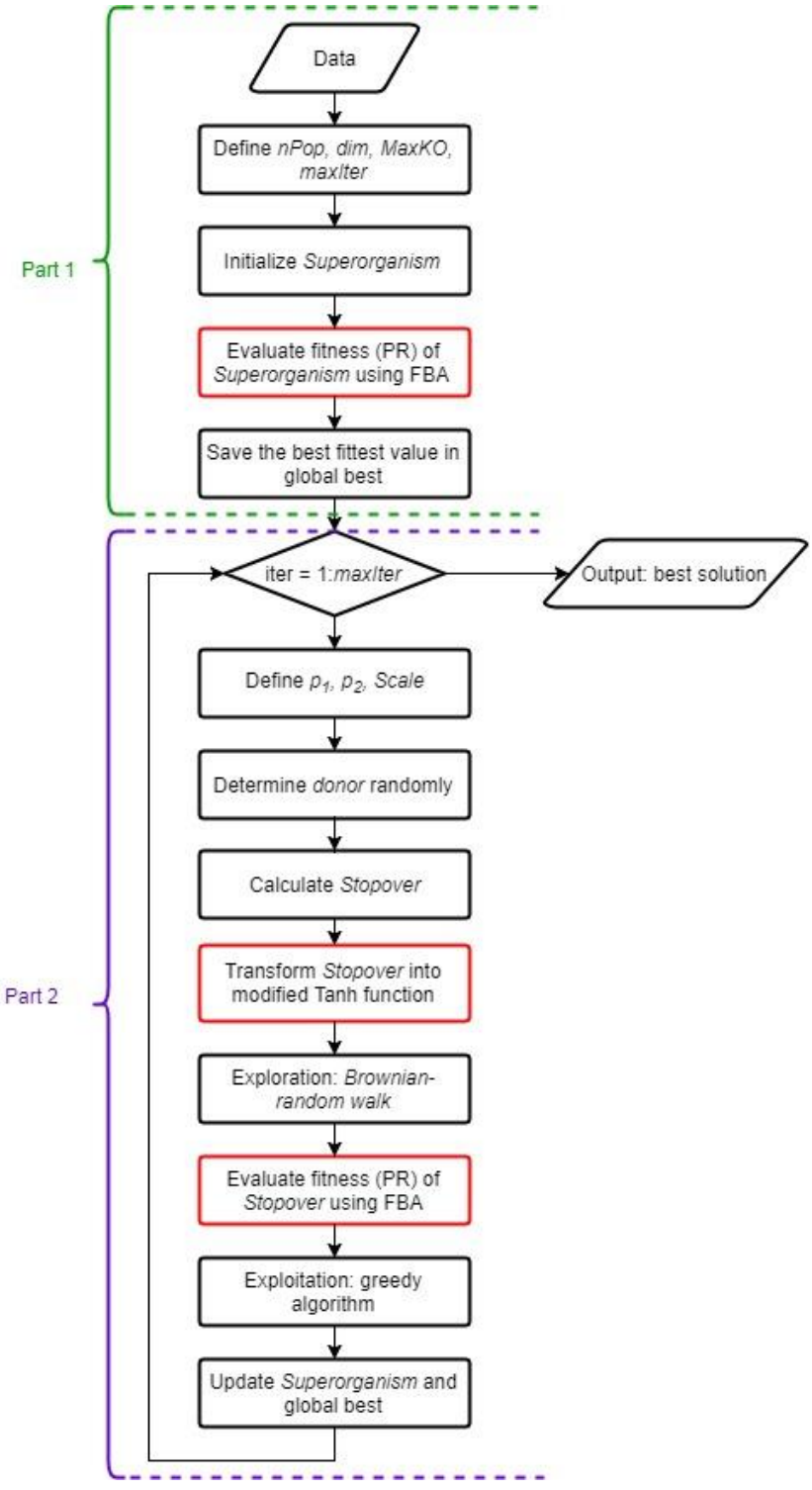

Fig. 1: The flowchart of DSA and FBA hybrid algorithm

4.2.1 Initialization of population

The initialization of population is a random binary matrix of size $N \times D$, whereby N is the number of artificial organisms and D is the dimension. In the metabolic model, the abundance of thousands of genes that represent more than one reactions can be described as the combinatorial problem due to the process of selecting best possible reactions for improving the production rate of metabolite are hindered. Therefore, the genes in metabolic network were assigned with binary representation, "0" and "1" which indicates maintained the genes and knockout, respectively. By randomly assigning the binary value, we can later deduce what reactions being knockout and the effects of knockouts towards the metabolite production. Furthermore, as in the case of multiple genes for one reaction, the effects of knockout is varied for "AND" relationship and "OR" relationship (gene-protein-reaction). The former will effect on all genes being inactivated, whereas the latter, adequate with only one gene being inactivated (Fig. 2).

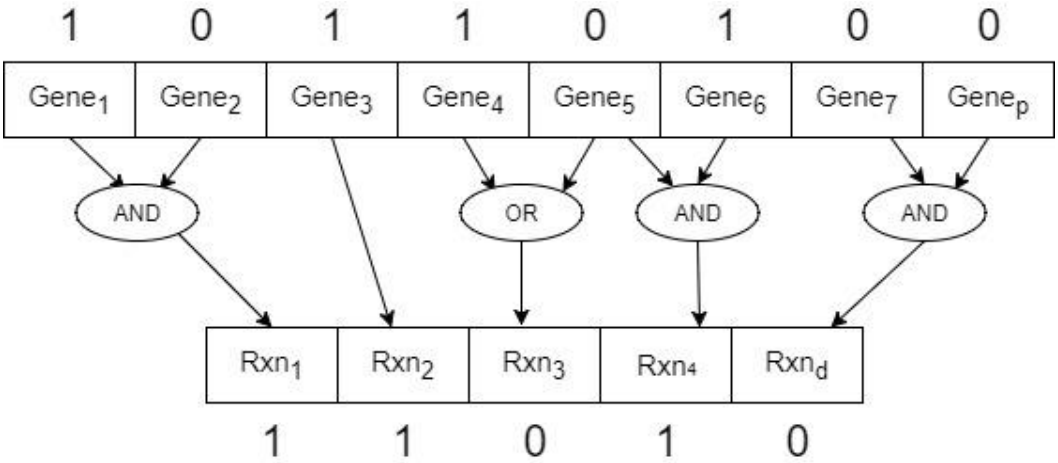

Fig. 2: Representation of binary 0 and 1 value to genes-reactions.

4.2.2 Product rate evaluation using Flux Balance Analysis

This fitness score is a marker as for whether the artificial-organism is suitable for the next generation. In order to calculate the fitness score, FBA is applied whereby the objective function, Z, is maximized according to this function,

$$Z = \sum_i^m c_i^T v_i \quad (4)$$

Where c is the weight vectors of reactions towards the objective function, v is flux vector, and i is the index variable of size m (m is the size of reactions).

Each reaction that has given knockout status, the upper and lower bounds of the respective metabolic flux were assigned to zero and consequently, a new modified metabolic model is derived. From the modified model, the fitness score of the objective function is calculated and the maximum score is selected for the next generation.

The stoichiometry-based models cannot predict rates without a priori knowledge or assumption of the substrate uptake rate [2]. Since FBA requires substrate uptake rate and oxygen consumption, therefore, production rate has been used as the objective function in this research. Fig. 3 shows the flow of calculation of fitness score, production rate. First, the growth rate of modified model were obtained and only mutant having growth rate of more than 0.1 will be selected as a candidate for optimizing the production of desired metabolite. However, the maximum production of desired metabolites (maximum theoretical yield) can only be achieved when the growth rate is at 0.

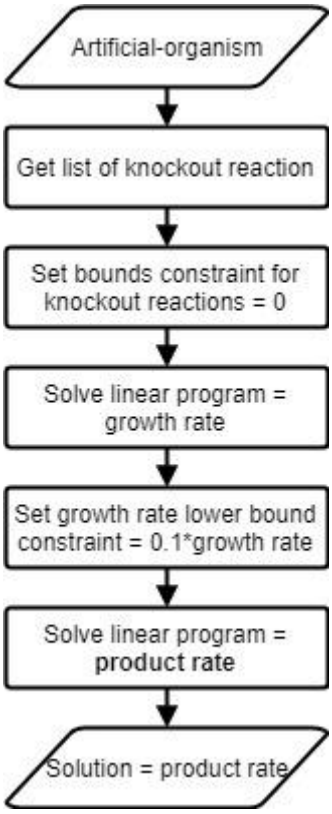

Fig. 3: The flow of calculation of fitness score which is production rate.

This research used production rate (PR) as the fitness score. Thus, each artificial-organism is assigned a fitness score, PR, as an indicator whether to select the artificial-organism for next generation. The fitness function can be formulated as below:

$$Product\ rate = \frac{amount\ of\ product\ produced}{time} \left(\frac{mmol}{gDW \cdot hr}\right) \quad (5)$$

4.2.3 New artificial-organism and termination

After the fitness score of each artificial-organisms is calculated, the best artificial-organism is selected as global fitness. In DSA, the exploration is done by *Brownian-random* walk motion whereby, individuals of the artificial-organisms,

known as *donor*; are selected randomly and move towards more fertile position, known as stopover, *So*. *Donor* is used to search for potential fruitful area (in this research is the combinations of reactions knockout giving higher production rate) is generated by uniform distributed random numbers of dimension. However, in conventional DSA, *donor* is a real number vector of dimension size. Thus, in this research, *donor* is transformed into binary number by modified *Tanh* transfer function as follows.

$$f(x) = \left|\frac{e^{x\tau} - 1}{e^{x\tau} + 1}\right| \tag{6}$$

Meanwhile, greedy algorithm is utilized during the exploitation process, whereby the fitness of the stopover sites was calculated and compared with the global fitness score. If the fitness of stopover site is better than the previous one, thus the artificial-organism can immediately settle at the discovered position and continue their migration from this position. These steps are repeated until it reaches maximum iteration (condition of stopping criterion). Algorithm 1 shows the pseudocode of *Brownian-random* walk motion, where there are five uniform distributed random numbers and four counters involved in the searching process.

Algorithm 1: Pseudocode for Brownian-random walk motion

```
if rand1 < rand2
    if rand3 < p1
        r = rand(d,n)
    for Counter1 = 1:d do
        r(Counter1,:) = r(Counter1,:) < rand4
    endfor
else
    r = ones(d,n)
        for Counter2 = 1:d do
            r(Counter2,randi(n)) = r(Counter2,randi(n)) < rand5
        endfor
    endif
else
    r = ones(d,n)
        for Counter3 = 1:d
            temp = randi(n,1,ceil(p2×n))
            for Counter4 = 1:size(temp)
                r(Counter3,temp(Counter4)) = 0
            endfor
        endfor
endif
```

In order to show the differences between standard DSA and our proposed modified DSA, we have carried out the pre-experiment between standard DSAFBA and proposed method. The difference between standard DSAFBA and

modified DSAFBA is modified *Tanh* transfer function has been applied to the latter method to generate binary number 0 and 1 from a matrix of stopover. In the case of *E.coli* core model for succinic acid production between standard DSAFBA and modified DSAFBA, the former method is able to achieve 15.3756 (in unit mmol/g[DW]×hr) while our proposed method able to obtain mutant producing succinic acid of 15.50 (in unit mmol/g[DW]×hr).

4.3 Experimental Setup and Dataset

The effectiveness of DSAFBA algorithm is tested by experimenting on three genom-scale metabolic model of two organisms, Escherichia coli, and Zymomonas mobilis. The properties of the organisms used throughout the work described herein is shown in Table 1. The genome-scale model of the three organisms is publicly available in System Biology Markup Language (SBML) format from BiGG database.

Table 1: Properties of organisms used

Model	Reactions	Genes	Metabolites	Ref
***E.coli* core model**	95	137	72	[19]
iAF1260	2162	1261	1461	[20]
iEM439	767	794	705	[21]

All models used were undergo preprocessing steps. Fig. 5 shows the steps involved in preprocessing. Since metabolic models of organisms are large, thus it is crucial in removing certain reactions that are unnecessary. The preprocessing steps are based on computational approaches and biological assumptions with the intention of eliminating unrequired reactions from the models. Also, the preprocessed steps were done in order to minimize the solutions space of the model, thus the candidate reactions for knockout could serve as valid knockout reactions.

Firstly, the essential reactions were removed from the list of potential knockout reactions. The essential reactions are the reactions that are necessary for the growth of the organisms [22,23]. Next step is to exclude transport reactions and reactions that are not associated with the production of desired metabolites which were the reactions from certain subsystems, including cell envelope biosynthesis, membrane lipid metabolism, glycerophospholipid metabolism, and inner and outer membrane transport. Lastly, the reactions that act on high-carbons were removed from the candidates of reactions knockout due to the difficulties in carrying high flux towards the production of desired metabolites. Therefore, the outcome of removing these reactions has greatly reduced the solution space and consequently, the computation time as well. Table 2 shows the preprocessed model and number of candidate reactions for knockout.

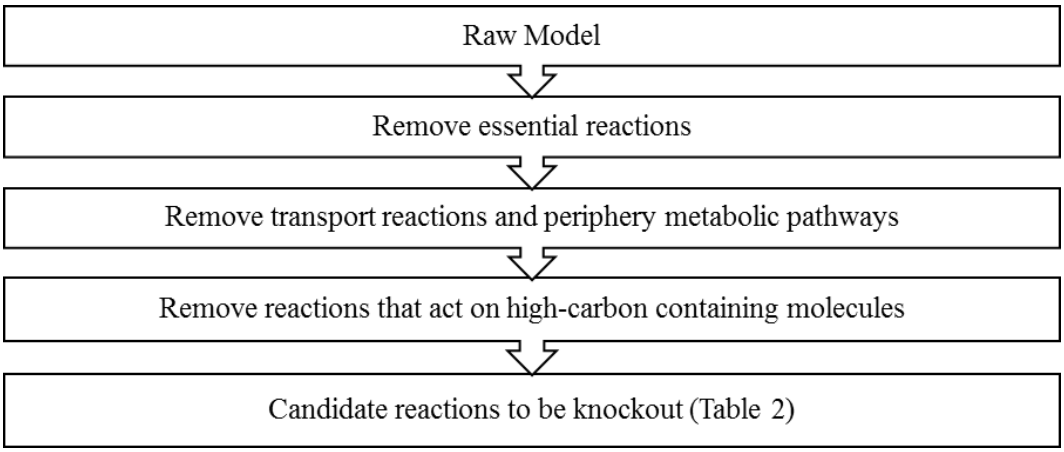

Fig. 5: Data preprocessing and candidate selection of reactions for knockout

Table 2: Preprocessed model and candidate reactions for knockout

Model	**Raw Model**		**Preprocess Model**	**Candidate reactions for knockout**
	Reactions	Metabolites	Reactions	
E.coli core model	95	75	48	48
iAF1260	2162	1461	461	461
iEM439	767	705	684	684

MATLAB and Constraints Based Reconstruction and Analysis toolbox (COBRA) are used for simulation of reactions knockout in the three datasets. The results obtained were assessed according to their production rate and growth-rate after the knockout. In this paper, all simulations were performed according to the conditions (substrate uptake rate, desired metabolite to enhance, and oxygen uptake rate) as shown in Table 3. The conditions of simulations were obtained from previous researches [9,21]. The experiments mentioned here were done using a 3.6GHz Intel Core i7 processor with 16 GB RAM workstation. For the ease of comparison, the number of knockouts for each case study will be the same. However, we have tested for different number of knockouts and evidently, the results shown in Section 3 are the best among the different number of knockouts.

Table 3: Conditions of simulation experiment used throughout this study and the theoretical yield of each metabolite

Organism	Oxygen	Substrate uptake rate	Desired Metabolite	Growth rate	Maximum theoretical yield	Ref
iAF1260	18.5	10	Succinic acid	0.8856	16.722	[9]
			Acetic acid		28.467	
iEM439	0	10	Ethanol	0.147	20	[21]
E.coli core model	10	10	Succinic acid	0.874	16.38	[19]

5 Results, Analysis and Discussions

5.1 Case Study 1: Succinic acid production in E.coli core model

Our proposed hybrid algorithm was compared with OptKnock and IdealKnock for production of succinic acid in *E.coli* core model. Table 4 shows the result of three different methods, together with their suggested knockouts, time taken, maximum production rate, growth rate and number of knockout. From this table, the suggested knockouts by our proposed method, DSAFBA, is able to achieve the highest production rate of succinic acid with high growth rate compared to OptKnock and IdealKnock with OptKnock obtained by Gu et. al [6]. Furthermore, we successfully obtained nearly the similar suggested knockouts as the other two methods, which are G6PDH2r, ME1, ME2, PYK, and SUCOAS.

Table 4: Suggested knockout strategies obtained by different methods for succinic acid in *E.coli* core model

Max. theoretical yield: 16.38 Wild-type growth rate: 0.874 (hr^{-1})			
Method	Production Rate (mmol/g[DW]×hr)	Growth rate (hr^{-1})	Suggested knockouts
DSAFBA	15.50	0.4836	ACALD, G6PDH2r*, GND, ME1*, ME2*, PFL, PYK*, SUCOAS*
OptKnock [6]	13.80	0.059	AKGDH, AKGt2r, GLUt2r, ME1*, NADTRHD, PDH, PTAr, PYK*

IdealKnock + OptKnock [6]	12.73	0.2	ACKr, AKGt2r, G6PDH2r*, GLUt2r, ME1*, ME2*, PYK*, SUCOAS*

* Denotes the same suggested knockouts obtained by DSAFBA and other methods.

In aerobic wild-type *E.coli*, the production of succinic acid is only occur as final product of citric acid cycle with the presence of glyoxylate bypass [24]. One way of optimizing the production of succinate is to manipulate the pyruvate metabolizing system. From the list of suggested knockouts, three reactions are involved in pyruvate metabolism; acetaldehyde dehydrogenase (ACALD), pyruvate formate lyase (PFL), and pyruvate kinase (PYK). The genes associated with these reactions are *adhE*, *pflB*, and *pykA*, respectively. These reactions are responsible for the conversion of phosphoenolpyruvate into other byproducts formation including lactate, formate, acetate, acetaldehyde, and ethanol. The knocking out of these reactions and their associated genes, will redirect the flux towards citric cycle. The first *E.coli* mutant strain, NZN111, was developed by inactivating the pflB and *ldhA* genes, however, it resulted in low biomass flux and succinic acid production [24–26]. Further improvement of mutant strain with improved growth rate and succinic acid production has been shown by strain KJ060 which made inactivation of *adhE*, *ldhA*, *ackA*, *focA*, and *pflB* genes showed improvement in succinate yields of 1.2 – 1.6 mol/mol glucose. However, there were traces of byproducts acetate and malate being produce as well [27].

Moreover, the intermediate metabolites, phosphoenolpyruvate (PEP) and malate are the precursor for formation of pyruvate, succinate and other byproducts. The conversion of PEP to pyruvate is done by *pykA* gene, while *maeA* and *maeB* genes do the conversion of malate to pyruvate. Thus, these genes encoded for PYK and malic enzymes (ME1 and ME2) are suggested for knockout as well [28,29]. Another approach of optimizing the succinate yield is via pentose-phosphate (PP) pathway by regulating the NADH supply. PP pathway generates NADPH via oxidation of glucose-6-phosphate by *zwf* gene (glucose 6-phosphate dehydrogenase reaction, G6PDH2r). According to [30,31], the knocking out of *zwf* and *gnd* genes will slightly affect the growth rate of organism and resulted in non-operational PP pathway. Eventually, flux towards the citric acid cycle will be increased as well. In activation of latter gene, which encodes for phosphogluconate dehydrogenase (GND) reaction, will reroute the flux towards Entner-Doudoroff pathway [31].

In the case of succinyl-coA synthetase (SUCOAS) reaction, encoded by gene *sucC* and *sucD* genes, it is responsible for the reversible conversion of succinate to succinyl-coA. Previous research has shown that this gene is important for the cell viability; and removal of these genes may resulted in reduced growth rate of the organism [32]. Considering that succinic acid is an intermediate product of TCA cycle, thus it is important to block the flux from divert from the succinic acid production. Therefore, the combination of suggested knockouts by DSAFBA

are responsible for growth rate and optimizing the production of succinic acid in the organism.

5.2 Case Study 2: Succinic acid and acetic acid production in E.coli iAF1260

The *E.coli* has been used in producing industrially important metabolites in bulk forms such acetic acid, succinic acid, and others [3,27,33]. Using *E.coli*, by removing certain reactions, the production of succinic acid and lactic acid could be improved. Succinic acid has been useful in wide range of applications and industries including food processing, pharmaceuticals, and others [34,35]. Regardless of the usefulness of succinic acid, however, the production of succinic acid as an intermediate minute fermentation product, has urged the scientists and researchers to reengineer and design the organism for increasing the production of succinic acid [35]. Meanwhile, acetic acid is one of the major fermentation products produced by fermentative bacteria. However, in high cell density of *E.coli*, the production of acetic acid may inhibits the cell growth [36].

Therefore, in second case study, *Escherichia coli* of strain iAF1260 has been used for our proposed algorithm for optimizing the production of succinic acid. While, for acetic acid case study, the knockout strategies are carried out to optimize the production rate, together with the growth rate. The simulation results obtained are compared with previous methods and validate biologically by referring to biological database and journals.

Table 5 shows the suggested knockout strategies obtained by different methods for succinic acid production in *E.coli* iAF1260. From Table 3, it shows that DSAFBA suggests knockout reactions that contain competing reactions either consuming succinate or precursor for production of succinic acid. The reactions involved are succinate dehydrogenase reaction (SUCDi) and phosphotransacetylase (PTAr), encoded by *pta* or *eutD* genes and *sdhA*, *sdhB*, *sdhD*, or *sdhC* genes, respectively. Considered that succinic acid is an intermediate product of fermentation, the knockout of the former reaction will hinder the production of fumaric acid [37]. PTAr catalyzes the reaction of coenzyme-A from acetyl-coA. However, this reaction competes with succinic acid production for its precursor [10]. Consequently, many other strain design algorithms suggested knockout combinations of *pta* and *sdhABCD* [38,39].

Table 5: Suggested knockout strategies obtained by different methods for succinic acid production in *E.coli*

Max. theoretical yield: 16.72 Wild-type growth rate: 0.8856 (hr^{-1})			
Method	Production rate (mmol/g[DW]×hr)	Growth rate (hr^{-1})	Suggested knockouts
DSAFBA	12.447	0.5749	DHORD2, IDOND *, PSERT, PTAr, TALA, SUCDi *

ReacKnock [40]	9.130	0.129	ACGAMK, ACt2rpp, ATPS4rpp, IDOND *, PSP_L, SUCDi *
IdealKnock [6]	9.254	0.0973	ALCD2x, CYTBD2pp, CYTBDpp, CYTBO3_4pp, D_LACt2pp, PYAM5PO

* Denotes the same suggested knockouts obtained by DSAFBA and other methods.

DSAFBA also suggests some other knockout reactions, including dihydroorotic acid dehydrogenase (DHORD2), L-idonate 5-dehydrogenase (IDOND), and phosphoserine transaminase (PSERT). These reactions are encoded by genes *pyrD*, *idnD*, and *serC*, respectively. Generally, after the perturbations, the organisms will try to maintain the homeostasis of the cell and balance between production of succinic acid and growth rate. DHORD2 and IDOND are responsible for maintaining the balance of H^+ in *E.coli*, while PSERT is responsible for the growth of cell [41]. Not only in *E.coli*, the suggested knockouts for improving the production of succinic acid can be seen in other organism as well such as *Mannheimia succiniproducens* [42].

In Table 6, although the difference of acetic acid production rate between our proposed algorithm and other methods are low (approximately 1.88), nevertheless, DSAFBA is able to achieve highest mutant growth rate compared to OptKnock and IdealKnock in *E.coli* iAF1260. Furthermore, DSAFBA is capable in suggesting four knockouts similar to OptKnock and IdealKnock, which are 6-phosphogluconate dehydratase (EDD), 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA), fructose 6-phosphate aldolase (F6PA), and fructose-bisphosphate aldolase (FBA). The two former reaction is associated to pentose phosphate pathway and encoded by *edd* and *eda* genes, while the latter two reactions are associated with glycolysis and encoded by genes *fsaA* and *fsaB*; *fbaA*, *ydjI*, and *fbaB*, respectively. The other suggested knockout is xylose isomerase (XYL12) encoded by gene *xylA*.

Table 6**:** Suggested knockout strategies obtained by different methods for acetic acid production in *E.coli*

Max. theoretical yield: 28.467
Wild-type growth rate: 0.8856 (hr^{-1})

Method	Production rate (mmol/g[DW]×hr)	Growth rate (hr^{-1})	Suggested knockouts
DSAFBA	21.2686	0.746	EDA *, EDD *, F6PA *, FBA *, XYL12
OptKnock [40]	23.150	0.119	ATPS4rpp, ECOAH5, EDA *, LYSt3pp, TPI
IdealKnock [6]	23.147	0.115	ATPS4rpp, EDD *, F6PA *, FBA *, PGCD

* Denotes the same suggested knockouts obtained by DSAFBA and other methods.

Starting with the knockout of F6PA and FBA, the sequence of gluconeogenesis steps can be avoided. Therefore, there will be surplus of fluxes directed towards the production of pyruvate, which is the precursor for acetic acid production. Moreover, EDA reaction that catalyzes 2-dehydro-3-deoxy-D-gluconate-6-phosphate to pyruvates and D-glyceraldehyde 3-phosphate is suggested for knockout as well. Meanwhile, EDD reaction which responsible for production of 2-dehydro-3-deoxy-D-gluconate 6-phosphate, is suggested for knockout as well. Thus, the combination of these four suggested knockouts, we hypothesized that the Embden-Meyerhof-Parnas (EMP) pathway is being deactivated, and instead activate the ED pathway, which produced pyruvate from glucose-6-phosphate [43]. In the case of xylose isomerase reaction, it is responsible for ethanol formation [44]. Since pyruvate is precursor for production of ethanol and acetic acid, thus, by removing *xylA* gene associated XYL12 reaction, the flux towards ethanol production is forward to the production of acetic acid.

5.3 Case study 3: Ethanol production in Z.mobilis

Z.mobilis is an anaerobic ethanologenic bacterium that uses the Entner-Doudoroff (ED) pathway for glycolysis, relative to the other engineered bacteria such as *S.cerevisiae* and *E.coli* which use Embden-Meyerhof-Parnas (EMP) [21,45]. The advantages of *Z.mobilis* including higher substrate uptake rates, high tolerance towards alcohol, and lower biomass rate, has let *Z.mobilis* become a promising host for ethanol productions [46]. Furthermore, different model strains of *Z.mobilis* has been developed and curated, with the latest account for 692 reactions [21]. In addition, detailed reviews of this promising bacteria for ethanol production has been thoroughly reported [46–48]. In this research, we applied GSMM of *Z.mobilis* for identifying set of reactions knockout for optimizing the production of ethanol.

For the production of ethanol in this organism, however, we could not manage to get comparison from other methods, as *Z.mobilis* is still new for fuel-bio refineries in the industrial scale. From Table 7, DSAFBA suggested three knockouts for optimizing the production rate of ethanol; glutamate dehydrogenase (GLUDy), lactate dehydrogenase (LDH_D), and phosphoglucose isomerase (PGI).

Table 7: Suggested knockout strategies obtained by different methods for ethanol production in *Z.mobilis*

Max. theoretical yield: 20
Wild-type growth rate: 0.147 (hr^{-1})

Method	Production Rate (mmol/g[DW]×hr)	Growth rate (hr^{-1})	Suggested knockouts
DSAFBA	19.80	0.130	GLUDy, LDH_D, PGI
OptKnock	17.90	0.134	CYSDS, SGDS, PPKr

According to Lee *et al.* [49], pyruvate decarboxylase is the sole source for ethanol production in *Z. mobilis*. This is due to the production rate of ethanol dropped

when pyruvate decarboxylase is removed from *Z.mobilis*. However, the knockout of glutamate dehydrogenase which encoded by *gltB* gene, did not affect the maximum production rate of ethanol. Furthermore, work done by Nissen *et al.* [50] for overproducing ethanol production in *S.cerevisiae*, had shown the deletion of the NADPH-dependent glutamate dehydrogenase has resulted in increase of ethanol production and lower the byproduct formation. Nevertheless, we conclude that this gene is responsible for the growth rate of *Z.mobilis,* as it has slightly minimized the growth rate of mutant.

Pyruvate is the precursor for ethanol production. However, other byproducts such as lactate and succinate compete with ethanol. In the case of *ldhA* gene associated LDH_D reaction, which transform pyruvate to lactate, it has been suggested for knockout as well. Referring to the work done by Seo *et al.* [51], the deletion of *ldhA* gene has increased the ethanol production by 0.25 fold. Furthermore, the deletion of this reaction for optimizing the ethanol production is observable in other organisms as well, such as *Clostridium thermocellum*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* [52–54]. Therefore, knocking out the LDH_D reaction may increase the production of ethanol and succinic acid as it blocks the production lactate and redirect the flux towards other productions.

Meanwhile, phosphoglucose isomerase encoded by *pgi* gene that convert glucose-6-phosphate into fructose-6-phosphate has been suggested for knockout as well. Considering that *Z.mobilis* used ED pathway for fermenting glucose, thus, knocking out *pgi* gene eventually will divert the flux towards gluconate and ED pathway. Furthermore, the removal of this gene, will hinder the byproducts production including xylose and ribulose [49]. The same observation of ethanol production in *E.coli* has been made by [55]. Thus, the combinations of reactions knockout suggested by DSAFBA may represent a guideline for biologists to design and reengineer the organism in biological experiment.

In the meantime, Table 8 shows the average time taken of different methods using two *E.coli* models. Considering that *E.coli* core model is a condensed version and contains central metabolism reactions, the computational time taken of DSAFBA is the fastest compared to other methods. However, for *E.coli* strain iAF1260, the computational time taken of DSAFBA to obtain the near optimal result was slightly higher than IdealKnock. This is probably due to the searching strategy of DSA that employs multiple artificial-organisms. Nevertheless, regardless of the longer computation time for DSAFBA, it successfully able to identify knockout strategies and surpasses the knockout strategies obtained by the other two methods.

Table 8: Comparison of computational time (in sec) for different methods and organisms

Model	Method	Computation time (sec)
E.coli core model	DSAFBA	10.8
	OptKnock	270

	IdealKnock	111.6
Escherichia coli iAF1260	DSAFBA	1296
	ReacKnock	201600
	IdealKnock	1080

From Table 8, it can be concluded that DSAFBA perform better than OptKnock and ReacKnock. Moreover, considering the size of model used, DSAFBA is applicable for small model, however for large dataset; DSAFBA requires slightly longer computational time. According to [6], although their proposed algorithm was able to identify knockout strategies with less computation time, however, the results obtained does not ensure for global optima.

Although DSA employs multiple artificial-organisms, it is capable to outperform other methods in finding the best knockout strategies with highest production rate and growth rate. Furthermore, DSA is a new population-based metaheuristic optimization algorithm mainly having three advantages; using less parameters, able to find global near-optimal solutions, and less computational time. Fig. 6 shows the convergence graph of objective function for this study.

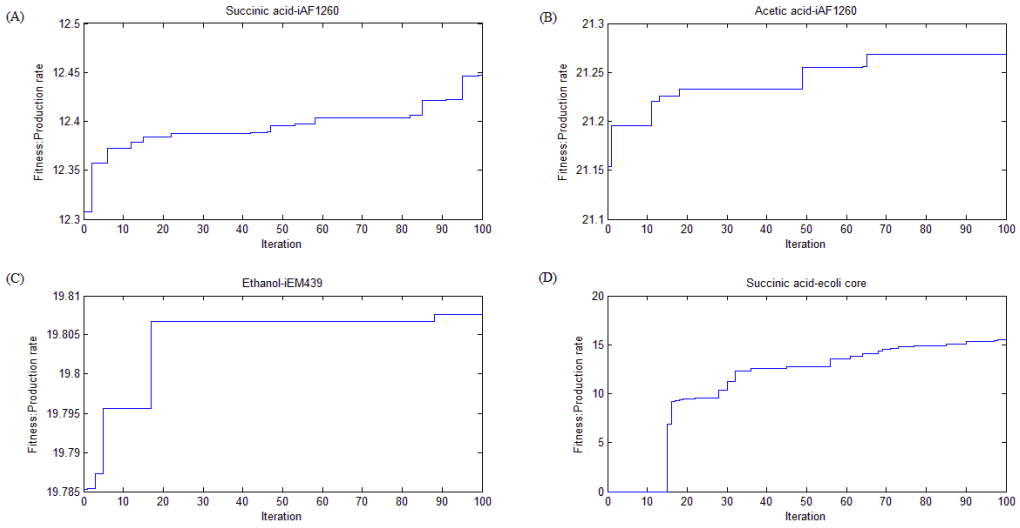

Fig. 6: Convergence graph of the DSAFBA of different production of desired metabolites. The x-axis represents iterations whereas the y-axis represents the production rates.

6 Conclusion

The success of developing the precise and effective hybrid modeling and optimization methods for simulation of genome-scale metabolic networks has significantly contributed to the field of *in silico* metabolic engineering. Eventually, it leads to the time improvement in the production of desired metabolites.

In this study, we proposed and developed a hybrid of modeling; Flux Balance Analysis, and optimization algorithm; Differential Search Algorithm, to identify

knockout strategies for improving the production rate of desired metabolites. Three genome-scale metabolic models of two organisms, *Escherichia coli* and *Zymomonas mobilis,* were used along the study. DSAFBA is able to identify near-optimal sets of reactions knockout for enhancing the production of the desired metabolites while maintaining the viability of organisms. According to the results, it shows a significant improvement in the productions of desired metabolites compared to the previous methods. DSA effectiveness has been shown to identify the best set of reactions knockout. Based on the predicted set of knockout done by DSA, FBA is applied as the fitness function to calculate the optimum production of desired metabolites.

However, due to the large complexity of genome scale metabolic model, some suggested knockouts are non-intuitive, therefore, it is important to have biological validation by means of *in vivo* experiment. Despite the irrelative suggested knockout directly towards the production of desired metabolites, we assumed that our proposed algorithm assessed the metabolic network as highly complex reactions – interactions, thus incorporates reactions from different pathway as well.

The results presented here are limited to computational validation and could be applied as a prior knowledge during *in vivo* implementation. However, in the real situation, the media condition for cultivating organisms, and other factors should be considered as well. Nevertheless, the *in silico* results presented here can be used as a reference in aiding the scientists and biologists to assess the validity of the results in the wet lab experiment. For future work, consideration of simultaneous multi-objective optimization between growth rate and production rate is proposed for improving the previous algorithm.

ACKNOWLEDGEMENTS
This research is supported by Universiti Teknologi Malaysia through Research University Grant (GUP) Tier 1 Research Grants: (Grant number: Q.J130000.2528.11H45 and Q.J130000.2528.11H11).

References

[1] Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., & Palsson, B. Ø. (2010). Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metabolic engineering, 12(3), 173-186.

[2] Patil, K. R., Rocha, I., Förster, J., & Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. BMC bioinformatics, 6(1), 308.

[3] Lee, S. Y., Hong, S. H., & Moon, S. Y. (2002). In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered

Escherichia coli as an example. Genome Informatics, 13, 214-223.

[4] Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering, 84(6), 647-657.

[5] Tepper, N., & Shlomi, T. (2009). Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics, 26(4), 536-543.

[6] Gu, D., Zhang, C., Zhou, S., Wei, L., & Hua, Q. (2016). IdealKnock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. Computational biology and chemistry, 61, 229-237.

[7] Salleh, A. H. M., Mohamad, M. S., Deris, S., Omatu, S., Fdez-Riverola, F., & Corchado, J. M. (2015). Gene knockout identification for metabolite production improvement using a hybrid of genetic ant colony optimization and flux balance analysis. Biotechnology and bioprocess engineering, 20(4), 685-693.

[8] Choon, Y. W., Mohamad, M. S., Deris, S., Illias, R. M., Chong, C. K., Chai, L. E., Omatu, S. & Corchado, J. M. (2014). Differential bees flux balance analysis with optknock for in silico microbial strains optimization. PloS one, 9(7), e102744.

[9] Rocha, M., et al. (2008). Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC bioinformatics, 9(1), 499.

[10] Wang, F. S., & Wu, W. H. (2015). Optimal design of growth-coupled production strains using nested hybrid differential evolution. Journal of the Taiwan Institute of Chemical Engineers, 54, 57-63.

[11] Chowdhury, A., Zomorrodi, A. R., & Maranas, C. D. (2015). Bilevel optimization techniques in computational strain design. Computers & Chemical Engineering, 72, 363-372.

[12] Maia, P., Rocha, M., & Rocha, I. (2016). In silico constraint-based strain optimization methods: the quest for optimal cell factories. Microbiology and Molecular Biology Reviews, 80(1), 45-67.

[13] Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. Nature biotechnology, 28(3), 245.

[14] Civicioglu, P. (2012). Transforming geocentric cartesian coordinates to

geodetic coordinates by using differential search algorithm. Computers & Geosciences, 46, 229-247.

[15] Kumar, V., Chhabra, J. K., & Kumar, D. (2016). Data Clustering using Differential Search Algorithm. PERTANIKA JOURNAL OF SCIENCE AND TECHNOLOGY, 24(2), 295-306.

[16] Liu, J., Teo, K. L., Wang, X., & Wu, C. (2016). An exact penalty function-based differential search algorithm for constrained global optimization. Soft Computing, 20(4), 1305-1313.

[17] Liu, J., Wu, C., Cao, J., Wang, X., & Teo, K. L. (2016). A Binary differential search algorithm for the 0–1 multidimensional knapsack problem. Applied Mathematical Modelling, 40(23-24), 9788-9805.

[18] Kiran, M. S. (2015). The continuous artificial bee colony algorithm for binary optimization. Applied Soft Computing, 33, 15-23.

[19] Orth, J. D., Fleming, R. M., & Palsson, B. O. (2010). Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. EcoSal plus.

[20] Feist, A. M., et al. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular systems biology, 3(1), 121.

[21] Motamedian, E., Saeidi, M., & Shojaosadati, S. A. (2016). Reconstruction of a charge balanced genome-scale metabolic model to study the energy-uncoupled growth of Zymomonas mobilis ZM1. Molecular BioSystems, 12(4), 1241-1249.

[22] Joyce, A. R., & Palsson, B. Ø. (2008). Predicting gene essentiality using genome-scale in silico models. In Microbial Gene Essentiality: Protocols and Bioinformatics (pp. 433-457). Humana Press.

[23] Joyce, A. R., et al. (2006). Experimental and computational assessment of conditionally essential genes in Escherichia coli. Journal of bacteriology, 188(23), 8259-8271.

[24] Thakker, C., Martínez, I., San, K. Y., & Bennett, G. N. (2012). Succinate production in Escherichia coli. Biotechnology journal, 7(2), 213-224.

[25] Donnelly, M. I., Millard, C. S., Chen, M. J., Rathke, J. W., & Clark, D. P. (1998). A novel fermentation pathway in an Escherichia coli mutant producing succinic acid, acetic acid, and ethanol. In Biotechnology for Fuels

and Chemicals (pp. 187-198). Humana Press, Totowa, NJ.

[26] Bunch, P. K., Mat-Jan, F., Lee, N., & Clark, D. P. (1997). The IdhA gene encoding the fermentative lactate dehydrogenase of Escherichia coli. Microbiology, 143(1), 187-195.

[27] Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., Moore, J. C., Shanmugam, K. T., & Ingram, L. O. (2008). Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of Escherichia coli C that produce succinate and malate. Biotechnology and bioengineering, 99(5), 1140-1153.

[28] Noda, S., Shirai, T., Oyama, S., & Kondo, A. (2016). Metabolic design of a platform Escherichia coli strain producing various chorismate derivatives. Metabolic engineering, 33, 119-129.

[29] Zhao, Y., Wang, C. S., Li, F. F., Liu, Z. N., & Zhao, G. R. (2016). Targeted optimization of central carbon metabolism for engineering succinate production in Escherichia coli. BMC biotechnology, 16(1), 52.

[30] Zhao, J., Baba, T., Mori, H., & Shimizu, K. (2004). Effect of zwf gene knockout on the metabolism of Escherichia coli grown on glucose or acetate. Metabolic Engineering, 6(2), 164-174.

[31] Jiao, Z., Baba, T., Mori, H., & Shimizu, K. (2003). Analysis of metabolic and physiological responses to gnd knockout in Escherichia coli by using C-13 tracer experiment and enzyme activity measurement. FEMS microbiology letters, 220(2), 295-301.

[32] Yu, B. J., Sung, B. H., Lee, J. Y., Son, S. H., Kim, M. S., & Kim, S. C. (2006). sucAB and sucCD are mutually essential genes in Escherichia coli. FEMS microbiology letters, 254(2), 245-250.

[33] Yang, J., Wang, Z., Zhu, N., Wang, B., Chen, T., & Zhao, X. (2014). Metabolic engineering of Escherichia coli and in silico comparing of carboxylation pathways for high succinate productivity under aerobic conditions. Microbiological research, 169(5-6), 432-440.

[34] Zeikus, J. G., Jain, M. K., & Elankovan, P. (1999). Biotechnology of succinic acid production and markets for derived industrial products. Applied Microbiology and Biotechnology, 51(5), 545-552.

[35] Lee, S. J., Lee, D. Y., Kim, T. Y., Kim, B. H., Lee, J., & Lee, S. Y. (2005). Metabolic engineering of Escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout

simulation. Applied and environmental microbiology, 71(12), 7880-7887.

[36] Han, K., Lim, H. C., & Hong, J. (1992). Acetic acid formation in Escherichia coli fermentation. Biotechnology and bioengineering, 39(6), 663-671.

[37] Ren, S., Zeng, B., & Qian, X. (2013, January). Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. In BMC bioinformatics (Vol. 14, No. 2, p. S17). BioMed Central.

[38] Hädicke, O., & Klamt, S. (2010). CASOP: a computational approach for strain optimization aiming at high productivity. Journal of biotechnology, 147(2), 88-101.

[39] Valderrama-Gomez, M. A., Kreitmayer, D., Wolf, S., Marin-Sanguino, A., & Kremling, A. (2017). Application of theoretical methods to increase succinate production in engineered strains. Bioprocess and biosystems engineering, 40(4), 479-497.

[40] Xu, Z., Zheng, P., Sun, J., & Ma, Y. (2013). ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. PloS one, 8(12), e72150.

[41] Koo, B. M., et al. (2017). Construction and analysis of two genome-scale deletion libraries for Bacillus subtilis. Cell systems, 4(3), 291-305.

[42] Lee, J. W., Yi, J., Kim, T. Y., Choi, S., Ahn, J. H., Song, H., Lee, M. H., & Lee, S. Y. (2016). Homo-succinic acid production by metabolically engineered Mannheimia succiniciproducens. Metabolic engineering, 38, 409-417.

[43] Bar-Even, A., Flamholz, A., Noor, E., & Milo, R. (2012). Rethinking glycolysis: on the biochemical logic of metabolic pathways. Nature chemical biology, 8(6), 509.

[44] Walfridsson, M., Bao, X., Anderlund, M., Lilius, G., Bülow, L., & Hahn-Hägerdal, B. (1996). Ethanolic fermentation of xylose with Saccharomyces cerevisiae harboring the Thermus thermophilus xylA gene, which expresses an active xylose (glucose) isomerase. Applied and environmental microbiology, 62(12), 4648-4651.

[45] Yang, S., et al. (2016). Zymomonas mobilis as a model system for production of biofuels and biochemicals. Microbial biotechnology, 9(6), 699-717.

[46] He, M. X., et al. (2014). Zymomonas mobilis: a novel platform for future biorefineries. Biotechnology for biofuels, 7(1), 101.

[47] Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J., & Siika-Aho, M. (2007). Thermostable enzymes in lignocellulose hydrolysis. In Biofuels (pp. 121-145). Springer Berlin Heidelberg.

[48] Kalnenieks, U., Pentjuss, A., Rutkis, R., Stalidzans, E., & Fell, D. A. (2014). Modeling of Zymomonas mobilis central metabolism for novel metabolic engineering strategies. Frontiers in microbiology, 5, 42.

[49] Lee, K. Y., Park, J. M., Kim, T. Y., Yun, H., & Lee, S. Y. (2010). The genome-scale metabolic network analysis of Zymomonas mobilis ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. Microbial cell factories, 9(1), 94.

[50] Nissen, T. L., Kielland-Brandt, M. C., Nielsen, J., & Villadsen, J. (2000). Optimization of ethanol production in Saccharomyces cerevisiae by metabolic engineering of the ammonium assimilation. Metabolic engineering, 2(1), 69-77.

[51] Seo, J. S., Chong, H. Y., Kim, J. H., & Kim, J. Y. (2009). U.S. Patent Application No. 12/279,692.

[52] Navarrete, C., Nielsen, J., & Siewers, V. (2014). Enhanced ethanol production and reduced glycerol formation in fps1Δ mutants of Saccharomyces cerevisiae engineered for improved redox balancing. AMB Express, 4(1), 86.

[53] Romero, S., Merino, E., Bolívar, F., Gosset, G., & Martinez, A. (2007). Metabolic engineering of Bacillus subtilis for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. Applied and environmental microbiology, 73(16), 5190-5198.

[54] Biswas, R., Prabhu, S., Lynd, L. R., & Guss, A. M. (2014). Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of Clostridium thermocellum. PLoS One, 9(2), e86389.

[55] Sekar, B. S., Seol, E., & Park, S. (2017). Co-production of hydrogen and ethanol from glucose in Escherichia coli by activation of pentose-phosphate pathway through deletion of phosphoglucose isomerase (pgi) and overexpression of glucose-6-phosphate dehydrogenase (zwf) and 6-phosphogluconate dehydrogenase (gnd). Biotechnology for biofuels, 10(1), 85.