Jurnal Teknologi

Full Paper

IN SILICO GENE DELETION OF ESCHERICHIA COLI FOR OPTIMAL ETHANOL PRODUCTION USING A HYBRID ALGORITHM OF PARTICLE SWARM OPTIMIZATION AND FLUX BALANCE ANALYSIS

Mei Jing Liew[a], Abdul Hakim Mohamed Salleh[a], Mohd Saberi Mohamad[a*], Yee Wen Choon[a], Safaai Deris[b], Azurah A Samah[a], Hairudin Abdul Majid[a]

[a]Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]Faculty of Creative Technology & Heritage, Universiti Malaysia Kelantan, Locked Bag 01, 16300 Bachok, Kota Bharu, Kelantan, Malaysia

Article history
Received
26 November 2015
Received in revised form
14 January 2016
Accepted
10 October 2016

*Corresponding author
saberi@utm.my

Abstract

Metabolic engineering of microorganism is widely used to enhance the production of metabolites that is useful in food additives, pharmaceutical, supplements, cosmetics, and polymer materials. One of the approaches for enhancing the biomass production is to utilize gene deletion strategies. Flux Balance Analysis is introduced to delete the gene that eventually leads the overproduction of the biomass and then to increase the biomass production. However, the result of biomass production obtained does not achieve the optimal production. Therefore, we proposed a hybrid algorithm of Particle Swarm Optimization and Flux Balance Analysis to attain an optimal gene deletion that is able to produce a higher biomass production. In this research, Particle Swarm Optimization is introduced as an optimization algorithm to obtain optimal gene deletions while Flux Balance Analysis is used to evaluate the fitness (biomass production or growth rate) of gene deletions. By performing an experiment on Escherichia coli, the results indicate that the proposed hybrid algorithm of Particle Swarm Optimization and Flux Balance Analysis is able to obtain optimal gene deletions that can produce the highest ethanol production. A hybrid algorithm is suggested due to its ability in seeking a higher ethanol production and growth rate than OptReg methods.

Keywords: Artificial Intelligence, Bioinformatics, Ethanol Production, Flux Balance Analysis, Gene Deletion Strategy, Metabolic Engineering, Particle Swarm Optimization

1.0 INTRODUCTION

In biological point of view, metabolism is a process within living organisms that is used to maintain life of the organisms and there is a huge amount of enzymatic reactions and transport processes are involved in metabolism. Specifically, these enzymatic reactions and transport processes are used to produce a metabolite from organic compounds [1]. Furthermore, the metabolite such as ethanol is widely used in the industrial application with various purposes and thus there is an increasing need for enhancing the metabolite production such as ethanol production. To fulfill these needs, the existence of metabolic engineering field is introduced as a technique to increase the production of metabolites. Generally, metabolic engineering is defined as experimental and mathematical tools that have been developed for genetic modification to enhance the production of desirable compounds [2]. Besides, metabolic

engineering is a technique that is used to modify the genes and metabolic pathways to enhance the production of a desired amino acid or byproduct in cell factory [3]. Hence, the purpose of metabolic engineering is to produce a high level biochemical product by introducing the modification of genes and metabolic pathway [4].

In addition, previous methods such as OptKnock and OptGene are the methods used in producing high metabolite production. Basically, OptKnock and OptGene are used to suggest gene knockout strategies for overproduction of metabolites such as biomass. According to [5], OptKnock is developed to suggest gene knockout strategies for biochemical overproduction and to recognize metabolic flux distributions that are governed by internal cellular objectives. In other words, it is used to search a set of gene deletions that maximizes the flux towards a desired product so that the identified gene deletions can force the microorganism to produce the desired product in order to achieve the maximal growth [2]. Next, OptGene is introduced after hybridizing the method of OptKnock and Genetic Algorithm. OptGene extends the applicability of OptKnock which is used to enhance the biomass production in metabolic engineering. OptGene is more advance than the OptKnock because OptGene is developed to overcome the limitations of the OptKnock.

However, there are some limitations found in these previous methods. First, OptGene does not perform well in binary implementations [2]. Hence, Particle Swarm Optimization is introduced as an optimization algorithm which consists of global and local search to this binary problem. Besides, the limitation found in OptKnock likes suggesting unrealistic flux distributions in some cases [5]. To overcome such limitation, Flux Balance Analysis is implemented. However, Flux Balance Analysis can only be used to evaluate the fitness after the genes are deleted. Thus, it cannot obtain the best gene deletion that leads to high biomass production. However, this limitation can be solved by implementing Particle Swarm Optimization in Flux Balance Analysis to rapidly identify the optimal set of gene deletions that enhances the production of biomass.

Therefore, a hybrid algorithm of Particle Swarm Optimization and Flux Balance Analysis is proposed in this research. As mentioned previously, the Flux Balance Analysis method is used to calculate the fitness of biomass production after gene deletions whereas Particle Swarm Optimization algorithm is used to attain the highest biomass production among all biomass productions that are calculated using Flux Balance Analysis methods. In this research, this proposed hybrid algorithm is able to suggest a suitable and optimal gene to be deleted for a particular biomass production if compared with OptKnock and OptGene. Besides, this hybrid algorithm is able to detect the optimal biomass production among the particular biomass production after the process of gene deletions. Hence, the hybrid algorithm is able to attain the optimal genes to be deleted that can eventually lead to high biomass production after the genes are deleted.

This paper is organized as follows. In Section 2, we describe a hybrid algorithm of Particle Swarm Optimization and Flux Balance Analysis. Then, the performance measurement of the proposed hybrid algorithm is discussed as well. Section 3 presents the dataset and experimental setup, and experimental results. Section 4 summarizes this paper by providing the conclusion and future work.

2.0 METHODOLOGY

2.1 A Hybrid Algorithm of Particle Swarm Optimization and Flux Balance Analysis (PSOFBA)

Flux Balance Analysis (FBA) is a mathematical method that uses for analyzing flows of metabolism of an organism through a metabolic network [6]. In this research, the metabolic network is represented by using stoichiometry matrix. Basically, FBA calculates the metabolic network (stoichiometry matrix) of an organism to predict the fitness (biomass production or growth rate). The input of FBA is a stoichiometry matrix (S). The row of S represents as a substrate or metabolite concentration and the column of S represents the internal flux and external flux. When positive number is returned to matrix S, it shows the metabolite is being produced. In contrast, if a negative number is returned to matrix S, it indicates that the participate metabolite is consumed. Furthermore, the FBA is able to predict the biomass production and growth rate based on the matrix S [6]. Hence, FBA has been widely used in metabolic engineering due to its ability to calculate the fitness (biomass and growth rate) after genes are deleted. However, FBA cannot obtain an optimal set of gene to delete for producing optimal biomass production.

Hence, the algorithm of Swarm Intelligence is proposed to solve this limitation of FBA. In this research, Particle Swarm Optimization (PSO) is introduced as an algorithm of swarm intelligence to search the optimal set of gene that can produce a higher biomass production. PSO is invented by Kennedy and Eberhart in 1995. This method simulates the behavior of birds or fishes to maintain an optimum distance between themselves [7] and it is widely used in data mining, signal processing, and function optimization. PSO is user-friendly because it can set few parameters. Moreover, it performs well in global search.

The steps for PSO are discussed briefly in the following section. First of all, the number of iterations, number of particles, and inertia are set. Noted that the inertia is used to control the momentum of particles. After that, random position and velocity of each particle is initialized. The value of initial velocity is usually set to a value of 0. Next, the position (fitness) of each particle is evaluated.

In addition, PSO consists of local search and global search. In this research, the best value of local search

is defined as personal best (pbest) whiles the best value of global search is defined as global best (gbest). The value of pbest is obtained by comparing the fitness of each particle at different iterations. Besides, the gbest is obtained by comparing the fitness of all particles in the iteration.If current fitness is greater than pbest, the current fitness is set to pbest and if current fitness is greater than gbest, current fitness is set to gbest. After that, the velocity and position of each particle are updated.

The Eq. (2.1) is the formula of velocity update whereas Eq. (2.2) is the formula of position update. Above step is repeated for each particle until the termination criterion is met. Noted that the termination criterion is the number of iterations reached the maximum number of iterations. Lastly, the algorithm returned the best solution found [8]. The Eq. (2.1) and Eq. (2.2) are defined as follows:

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (2.1)$$

$$V_{id}(t+1) = \omega X_{id}(t) + C_1\phi_1 \cdot \left(P_d(t) - X_{id}(t)\right) + C_2\phi_2 \cdot \left(g_d(t) - X_{id}(t)\right) \quad (2.2)$$

In addition, the explanation of each variable found in Eq.(2.1) and Eq. (2.2) is explained as follows:

$X_i = (X_{i1}, X_{i2}, \ldots, X_{id})$: i^{th} particle's position in search space.

$V_i = (V_{i1}, V_{i2}, \ldots, V_{id})$: i^{th} particle's velocity.

$P_d = (P_1, P_2, \ldots, P_d)$: Best position of the d^{th}.

$g_d = (g_1, g_2, \ldots, g_d)$: Best position in the whole swarm.

$i = 1, 2, \ldots, m$

$d = 1, 2, \ldots, D$

C_1, C_2: self confidence and swarm confidence respectively.

ϕ_1, ϕ_2: random number between 0 and 1.

$V_{id} \in \{-V_{max}, V_{max}\}$, V_{max}decided by user.

In this research, a hybrid algorithm of PSO and FBA called PSOFBA is proposed to solve the limitation found. As mentioned previously, FBA is used to calculate the fitness, whereas PSO is used to find the optimal set of gene to delete and it's corresponding ethanol production. There are four phases found in hybrid algorithm of PSOFBA which are initialization phase, search food phase, neighbour search phase and global search phase. Figure 1 is showing the flowchart of the hybrid algorithm of PSOFBA. The difference between PSOFBA flowchart and original PSO flowchart is that the fitness of the population for PSOFBA is evaluated and calculated using FBA. This modification is indicated by the dotted box found in Figure 1.

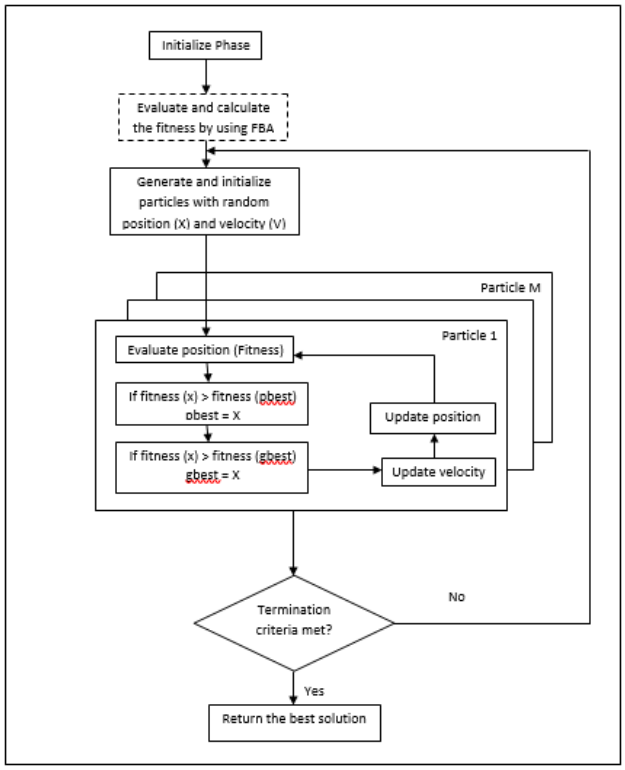

Figure 1 The flowchart of hybrid algorithm of PSOFBA The dotted box (modification area if compared with the original PSO algorithm) represents the used of FBA to evaluate and calculate the fitness of the population

The first phase is initialization phase. All needed parameters in experiment are set in this phase. The detail of parameters setting is mentioned in section 3.1. Then, a number of populations which are used for genes knockout are generated. The reactions within microorganism are carried out by the genes. The hybrid algorithm of PSOFBA is randomly created the 2000 population (the population is represented as 1532 x 2000 matrix and this represents matrix S) with the value zero and one. The value of 0 means the reaction has been knocked out whiles the value of one means the reaction does not knockout. The value of row represents reactions whereas the value of column represents the number of populations. The matrix S is the input of FBA. Lastly, FBA evaluates the fitness of the population (shows in the dotted box in Figure 1). The fitness is ethanol production and growth rate.

The second phase of the hybrid algorithm of PSOFBA is a search food phase. In this phase, the position and velocity of each particle are initialized first. The initial velocity of each particle for the first iteration is set at a value of 0. During the first iteration, the position of particles (which represents the fitness of each population) for the first hundred populations of Escherichia coli is set. In original PSO, the position of particles can represent as the position of birds or fish when they search for their food. This step is repeated for the next iteration; the position of particles is set to next hundred populations.

The next phase in PSOFBA is neighbor search phase. The neighbor search utilized ring topology. The fitness (growth rate and ethanol production) of each population is calculated using FBA before neighbor search phase. The fitness of particles is set as the particle's position. No neighbour search is located on the first iteration. For the second iteration, the position of each particle is compared to the previous position. If the current position is better (greater) than previous position, then the current position is set as pbest with the growth rate must be greater than 0.01 hr-1. This condition indicates that the Escherichia coli survives if the growth rate is greater than 0.01 hr-1. Otherwise, the previous position is set as pbest. For the next iterations, the current position is compared to the pbest only.

The last phase is the global search phase. The global search utilized start topology. For each iteration, the highest fitness among all particles in first hundred populations is set as gbest. Subsequently, the velocity and position are updated using Eq. (2.1) and Eq. (2.2). The steps are repeated until the termination criterion is met. Terminating the criterion in PSOFBA when reaching the maximum number of iterations. Hence, the last iteration is returned as the best solution with the reaction list that needed to be knocked out. The best solution represents best fitness among populations

2.2 Performance Measurement

In this research, standard deviation is used to measure the performance of PSOFBA. The Eq. (2.3) is the formula for standard deviation. This formula is used to calculate the standard deviation of growth rate for Escherichia coli after a gene knockout. In this experiment, the number of gene knockouts is performed from two to seven. Each number of gene knockouts is running for 50 times. Hence, the total number of samples is 300 samples.

$$S_N = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}} \qquad (2.3)$$

Where S_N= Standard deviation of growth rate for *Escherichia coli* after genes knockout
N = Total number of samples
X = Growth rate
$\overline{X}$ = Means of growth rate

3.0 RESULTS AND DISCUSSION

3.1 Dataset and Experimental Setup

The dataset used in this experiment is iAF1260 model [9]. The iAF1260 model is a metabolic network reconstruction for Escherichia coli K-12 MG1655.This model consists of 1261 genes, 2382 reactions, and 1668 metabolites. The iAF1260 model is downloaded from BioModels Database (www.ebi.ak.uk/biomodelsmain/MODEL3023609334) in Systems Biology Markup Language (SBML) format. The reason of using this model as the dataset in this experiment is that the characterization and quantification of the biomass components and maintenance requirements is associated with the growth of Escherichia coli [9].

In this experiment, the raw model of iAF1260 model is preprocessed before it is used in hybrid algorithm of PSOFBA. This is because the huge number of reactions and metabolites increase the computational time. The preprocessed model consists of two steps which are removing dead ends in the model and reducing the unused reactions in the model. The first step is to remove dead ends of reactions and metabolites in the iAF1260 model. The dead ends of metabolites blocked multiple reactions and resulted in extra or missing reactions. Next, the unused reactions in iAF1260 model are decreasing. This step removed all of the functions that had never used in this model. After model preprocessing, the number of reactions is reduced from 2382 reactions into 1532 reactions whereas the number of metabolites is reduced from 1668 metabolites into 1032 metabolites.

In addition, the hybrid algorithm of PSOFBA runs on Windows platform. The software used in the experiment is Matrix Laboratory (MATLAB) R2010b, Constraint-Based Reconstruction and Analysis (COBRA) toolbox version 2.0.3, and glpkmex version 2.11. The required random-access memory is 512 megabytes.The parameters set in this experiment are the number of iterations is set to 20, inertia is set to 1.0, the correction factor is set to 2.0 and the number of samples is set to 100. Besides, the number of populations of Escherichia coli are set to 2000 because the number of populations must be greater than the number of reactions list (1532 reactions) in the dataset. Thus, all reactions have the probability to be knocked out in the experiment.

3.2 Experimental Results

The outputs of the hybrid algorithm of PSOFBA are the name of the reactions, ethanol production, and growth rate after genes of Escherichia coli has been knocked out. Most of the reactions in the knockout lists are involved in glycolysis pathway, citrate cycle and the pentose phosphate pathway. Table 1 shows the knockout list of reactions, ethanol production and growth rate after gene knockouts for each number of genes knockout. These results show the highest ethanol production for each number of genes knockout. The results of ethanol production and growth rate are consider a better category when the value of the results is high. The number of genes knockout that has been chosen are two to seven genes knockout. This was because the ethanol production was found after one gene knockout was below than ten mmol gDW-1 hr-1. In addition, the number of genes knockout was greater than seven genes knockout because the growth rate of

Escherichia coli has been decreased dramatically. Hence, the results do not show the number of genes knockout that are less than two and more than seven. Figure 2 shows the glycolysis pathway, citrate cycle and the pentose phosphate pathway. Moreover, the explanation for each number of genes knockout on how to improve the ethanol production is explained in the table below except the two gene knockouts.

Table 1 Knockout list, production of ethanol, and growth rate after genes knockout

No.	Number of Genes Knockout	Knockout List	Ethanol (mmol gDW^{-1} hr^{-1})	Growt Rate (hr^{-1})
1	2	O2t, PFL	16.5863	0.1780
2	3	ACKr, ATPS4r, O2t.	20.4526	0.1676
3	4	ATPS4r, CYTBD, FUMt2_2, PTAr.	20.4526	0.1676
4	5	GLUDy, GND, NADH16, O2t, PTAr.	20.3062	0.1772
5	6	CYTBD, FBP, FUMt2_2, PTAr, RPE, THD2.	20.3060	0.1732
6	7	ACKr, AKGt2r, FUM, GLUDy, NADH16, PPS, TKT2.	20.2065	0.1775

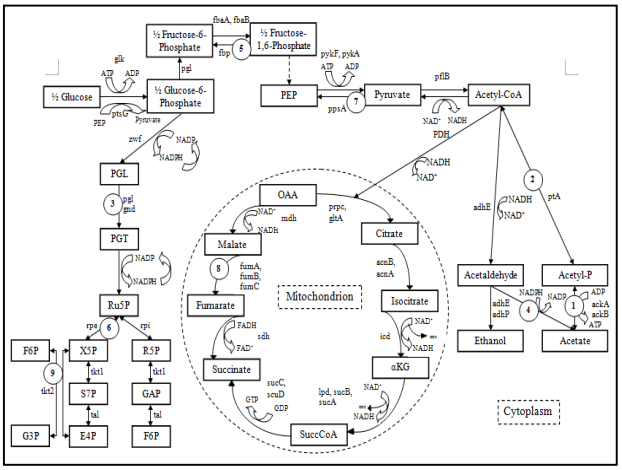

Figure 2 Glycolysis pathway, citrate cycle and pentose phosphate pathway

The knockout list of three genes knockout is acetate kinase (ACKr), Adenosine Triphosphate synthase (ATPS4r) and oxygen transport via diffusion (O2t). The reaction of O2t is knockout due to the anaerobic growth of Escherichia coli . In anaerobic conditions, pyruvate formate lyase (pfl) is a mutant. This causes the pyruvate cannot be converted into acetyl-CoA [10]. In the process of converting acetyl-P (reaction labelled in Circle 1 in Figure 2) to acetate, the ACKr and ATPS4r are needed. If the reaction of ACKr is knockout, the process of converting acetyl-P to acetate will not occur [11]. In addition, if the ATPS4r is knockout, then protons cannot pump into cell [6]. In conclusion, adenosine diphosphate (ADP) cannot turn phosphorylate to ATP without protons and hence no process from acetyl-P to acetate but the acetate can still produce a little amount. This phenomenon causes a high amount of acetyl-P [11]. The acetyl-P can be reconverted into acetyl-CoA. The ethanol can produce from acetyl-CoA. Acetylaldehyde acts as precursor for ethanol production due to the pfl mutant that causes the conversion of pyruvate directly into acetylaldehyde. Therefore, ethanol can be produced in a high amount.

Next, the knockout list of four genes knockout is ATP synthase (ATPS4r), cytochrome oxidase bd (CYTBD), Fumarate transport via proton symport (FUMt2_2) and phosphotransacetylase (PTAr). The ATPS4r leads the protons pumped into cell [6]. Then, the cytochrome d oxidase is required protons to catalyse the oxidation of ubiquinol-8. The cytochrome d oxidase has an affinity for oxygen and their sensitivity to respiratory inhibitors [12]. Therefore, the cytochrome d oxidase cannot catalyze the oxidation of ubiquinol-8. Besides, it is indicated that the Escherichia coli is in anaerobic growth due to limited oxygen of cytochrome d oxidase performed. Furthermore, the knockout of FUMt2_2 shows a poor aerobic growth on succinate [13]. Lastly, the process of converting acetyl-CoA to acetyl-P (reaction labelled in Circle 2 in Figure 2) is needed PTAr [14]. In conclusion, pyruvate cannot be converted into acetyl-CoA in anaerobic condition; moreover, the lack of PTAr also leads to form pyruvate. Therefore, the ethanol production is high due to the accumulation of pyruvate that acts as a precursor.

The knockout list of five genes knockout is glutamate dehydrogenase (GLUDy), phosphogluconate dehydrogenase (GND), Nicotinamide Adenine Dinucleotide dehydrogenase (NADH16), oxygen transport via diffusion (O2t) and phosphotransacetylase (PTAr). The GLUDy and NADH16 are associated together in the glutamate metabolism, which is between the processes of isocitate to a-ketoglutarate. The GND is required in the process of 6-phosphogluconolactonase (PGL) to 6-phosphogluconate (PGT) which is shows in Figure 2 that labelled in Circle 3. If the gene for GND is knockout, then glucose-6-phosphate (G6P) is increased [15]. If the reaction of O2t is knockout, it will cause the Escherichia coliis in anaerobic growth. In anaerobic condition, the pyruvate formate lyase (pfl) is a mutant. This is because pyruvate cannot be converted into acetyl-CoA [10]. Lastly, the process of acetyl-CoA to acetyl-P (reaction labelled in Circle 2 in Figure 2) is needed in PTAr [14]. In conclusion, the amount of glucose has the equal amount of G6P due to the knockout of GND. Then, the anaerobic condition and knockout of PTAr can lead to form pyruvate if compared to acetyl-CoA. Hence, the pyruvate can directly form acetylaldehyde and then the ethanol can be produced in a higher amount.

The knockout list of six genes knockout is cytochrome oxidase bd (CYTBD), fructose-biphosphate (FBP), Fumarate transport via proton symport (FUMt2_2), phosphotransacetylase (PTAr), ribulose 5-phosphate 3-epimerase (RPE) and NAD(P) transhydrogenase (THD2). The cytochrome d oxidase has an affinity for oxygen and their sensitivity to

respiratory inhibitors [12]. This is indicated the Escherichia coli is in anaerobic growth because cytochrome d oxidase can only perform in limited oxygen. Then, the FBP is avoiding the reformation of fructose-1, 6-diphosphate into fructose-6-phosphate (F6P) which is shown in Figure 2 that labelled with Circle 5. Besides, knockout of FUMt2_2 shows a poor aerobic growth on succinate [13]. Furthermore, the process of acetyl-CoA to acetyl-P (reaction labelled in Circle 2 in Figure 2) is needed in PTAr [14]. In addition, the process of ribulose-5-phosphate (Ru5P) to xylulose-5-phosphate (X5P) will be blocked if without RPE which is shown in Figure 2 that labelled with Circle 6. Lastly, the THD2 affects the process of acetaldehyde to form acetate (reaction labelled in Circle 4 in Figure 2). In conclusion, the pyruvate formate lyase (pfl) is a mutant in anaerobic condition. This phenomenon causes pyruvate cannot convert into acetyl-CoA. Hence, the pyruvate can form acetaldehyde directly. Therefore, the ethanol can be produced from acetaldehyde in a large amount due to the process of acetaldehyde to acetate is blocked.

The last knockout list is an acetate kinase (ACKr), 2-oxoglutarate reversible transport via symport (AKGt2r), fumarase (FUM), glutamate dehydrogenase (GLUDy), Nicotinamide Adenine Dinucleotide dehydrogenase (NADH16), phosphoenol pyruvate synthase (PPS) and transketolase II (TKT2). In the process of converting acetyl-P to acetate (reaction labelled in Circle 1 in Figure 2) in pyruvate metabolism, the ACKr and ATPS4r are needed. If the reaction of ACKr is knockout, the process of converting acetyl-P to acetate cannot occur [11]. The AKGt2r, GLUDy and NADH16 are associated together in the glutamate metabolism, which is between the processes of isocitate to a-ketoglutarate. Besides, knockout of FUM causes the malate and fumarate (reaction labelled in Circle 8 in Figure 2) cannot form into each other. Moreover, the PPS prevents the pyruvate being reformed into phosphoenol pyruvate (reaction labelled in Circle 7 in Figure 2). Lastly, the TKT2 causes F6P and glyceraldehyde-3-phosphate (reaction labelled in Circle 9 in Figure 2) cannot form into each other. This is because the glycolysis pathways occur smoothly without reformation. In conclusion, more ethanol production is produced.

Among the ethanol production, the highest ethanol production is produced in three and four genes knockout which is 20.4526 mmol gDW-1 hr-1. Besides, the highest growth rate of Escherichia coli after gene knockouts is 0.1780 hr-1 which is performed by 2 gene knockouts. For each type of knockouts, the average ethanol production, the average of growth rate and the standard deviation of growth rate for Escherichia coli after genes knockout are calculated. Table 2 shows the average of ethanol production, average of growth rate, and the standard deviation of growth rate for Escherichia coli after genes knockout. The formula of standard deviation is shown in section 2.1.

Table 2 Average of ethanol production, average of growth rate, and the standard deviation of growth rate for Escherichia coli after genes knockout

Number of Genes Knockout	2	3	4	5	6	7
Average of Ethanol Production (mmol gDW^{-1} hr^{-1})	19.4382	19.7671	19.9055	19.9001	19.9738	19.8677
Average of Growth Rate (hr^{-1})	0.1841	0.181	0.1805	0.1828	0.1813	0.183
Standard Deviation of Growth Rate (hr^{-1})	0.0133	0.0067	0.0086	0.0076	0.012	0.0088

Based on the Table 2, the five genes knockout performed better than other number of gene knockouts. This number of gene knockouts has produced a higher average ethanol production, which is 19.9055 mmol gDW-1 hr-1. Besides, the average of growth rate after gene knockouts is also high which is 0.1830 hr-1. In addition, the standard deviation for growth rate after gene knockouts is low which is 0.0080 hr-1. In conclusion, 5 genes knockout can produce a high amount of ethanol and also high growth rate for the survival. Furthermore, the bias within the average growth rates after gene knockouts is very small (0.0076 hr-1).

Table 3 shows the comparison of ethanol production and growth rate after gene knockouts between the hybrid algorithm of PSOFBA and OptReg [16]. The OptReg has done experiments for two and three genes knockout and also two and three genes regulation. The result of the hybrid algorithm of PSOFBA can only compare with two and three genes knockout. This is because the gene knockout strategy (in silico gene deletion) is not same as the gene regulation. Hence, the results of the hybrid algorithm of PSOFBA cannot make the comparison with gene regulation that has been done in OptReg.

Table 3 The comparison of ethanol production and growth rate after gene knockouts between PSO and OptReg

Method	Number of Knockout	Knockout List	Ethanol (mmol gDW^{-1} hr^{-1})	Growth Rate (hr^{-1})
PSOFBA	2	O2t, PFL	16.5863	0.1780
	3	ACKr, ATPS4r, O2t.	20.4526	0.1676
OptReg [16]	2	O2t, PTAr.	16.3000	0.1900
	3	PGI, PFL, O2t.	18.7400	0.0800

Based on the Table 3, all of the ethanol production that is found by the hybrid algorithm ofPSOFBA is better than OptReg. The two genes knockouts from hybrid algorithm of PSOFBA is produced 16.5863 mmol gDW-1 hr-1 of ethanol whiles two genes knockouts from OptReg is produced 16.3000 mmol gDW-1 hr-1 of ethanol. Besides, the three genes knockouts from

hybrid algorithm of PSOFBA produced ethanol in 20.4526 mmol gDW-1 hr-1 whiles three genes knockouts from OptReg produced ethanol in 18.7400 mmol gDW-1 hr-1. Furthermore, all of the growth rates from hybrid algorithm of PSOFBA are better than the growth rate from OptReg with three genes knockouts, but lower than the growth rate from OptReg with 2 genes knockouts. The reason for this result may be because the two gene knockouts in OptReg of Escherichia coli are not the essential genes. Therefore, the growth rate of two gene knockouts in OptReg is high.

In conclusion, the hybrid algorithm of PSOFBA is found to have the high ethanol production when the numbers of genes knockout are set to three and four gene knockouts. Furthermore, the five genes knockout of Escherichia coli perform better than the other number of genes knockout in terms of average of ethanol production, average of growth rate, and standard deviation of growth rate. Furthermore, the ethanol production that is found perform better than OptReg using the hybrid algorithm of PSOFBA.

4.0 CONCLUSION

Gene deletion strategies are very useful methods in the field of metabolic engineering to increase the biomass production. In this research, gene deletion strategies are introduced by using the hybrid algorithm of Particle Swarm Optimization and Flux Balance Analysis (which is called PSOFBA). Basically, this hybrid algorithm is proposed to overcome the limitations found in FBA and previous work such as OptKnock and OptGene.

In this research, the hybrid algorithm is applied into the ethanol case to find a suitable set of gene knockouts that leads to the highest ethanol production. In this case, the highest ethanol production that has been achieved by the hybrid algorithm of PSOFBA was 20.4526 mmol gDW-1 hr-1 . This production is higher than the highest ethanol production identified by OptReg which is 18.7400 mmol gDW-1 hr-1. Hence, the performance of the hybrid algorithm of PSOFBA is better than OptReg in terms of ethanol production.This result also shows that the hybrid algorithm PSOFBA is able to obtain optimal ethanol production after the genes are deleted.

In addition, there are some strengths and weaknesses of the proposed hybrid algorithm. The strength of this hybrid algorithm is that hybrid algorithm of PSOFBA can obtain the optimal ethanol production in a huge amount of samples. However, the weakness found in the hybrid algorithm of PSOFBA which is it takes a longer computational time when the number of samples is large. In future, this limitation may be overcomed by introducing parallel computing in the hybrid algorithm of PSOFBA for shorter computational time.

Acknowledgement

We would like to thank the Universiti Teknologi Malaysia for supporting this research through a GUP research grant (Grant number: Q.J130000.2528.12H12) and FRGS research grant (Grant number: R.J130000.7828.4F720).

References

[1] Kauffman, K. J., Prakash, P., Edwards, J. S. 2003. Advances In Flux Balance Analysis. *Curr Opin Biotechnol*. 14: 491-496.
[2] Patil, K. R., Rocha, I., Förster, J., Nielsen, J. 2005. Evolutionary Programming As A Platform For In Silico Metabolic Engineering. *BMC Bioinforma.* 6: 1-12.
[3] Park, J. H., Lee, S. Y. 2008. Towards System Metabolic Engineering Of Microorganisms For Amino Acids Production. *Curr Opin Biotechnol.* 19: 454-460.
[4] Curran, K. A., Crook, N. C., Alper, H. S. 2012. Using Flux Balance Analysis to Guide Microbial Metabolic Engineering. *Microb. Metab. Eng*. Cheng Q (eds.). 83: 197-216.
[5] Burgard, A. P., Pharkys, P., Maranas, C. D. 2003. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization, *Biosci. Bioeng*. 85: 647-657.
[6] Orth, J. D., Thiele, I., Palsson, B. O. 2010. What is Flux Balance Analysis. *Nat Biotechnol*. 28: 245-248.
[7] Kennedy, J., Eberhart, R. 1995. Particle Swarm Optimization. *Proceeding of the IEEE on Neural Network*. 27 Nov-1 Dec, Perth, WA: IEEE, 1942-1948.
[8] Das, S., Abraham, A., Konar, A. 2008. Swarm Intelligence Algorithms in Bioinformatics. *SCI*. 94: 113-147.
[9] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbell, L. J., Hatzimanikatis, V., Palsson, B. Ø. 2007. A Genome-Scale Metabolic Reconstruction for Escherichia coli K-12 MG1665 That Accounts for 1260 ORFs and Thermodynamic Information. *Mol Syst Biol*. 3: 121-138.
[10] Hasona, A., Kim, Y., Healy, F. G., Ingram, L. O., Shanmuga, K. T. 2004. Pyruvate Formate Lyase and Acetate Kinase Are Essential for Anaerobic Growth of Escherichia coli on Xylose. *J. Bacteriol*. 186: 7593-7600.
[11] Klein, A. H., Shulla, A., Reimann, S. A., Keating, D. H., Wolfe, A. J. 2007. The Intracellular Concentration of Acetyl Phosphate in Escherichia coli Is Sufficient for Direct Phosphorylation of Two-Component Response Regulators, *J. Bacteriol.* 189: 5574-5581.
[12] Cotter, P. A., Chepuri, V., Gennis, R. B., Gunsalus, R. P. 1990. Cytochrome o (cyoABCD) and d (cydAB) Oxidase Gene Expression in Escherichia coli Is Regulated by Oxygen, pH, and the fnr Gene Product. *J. Bacteriol.* 172: 6333-6338.
[13] Janausch, I. G., Kim, O. B., Unden, G. 2001. DctA- and Dcu-Independent Transport of Succinate in Escherichia coli: Contribution of Diffusion and of Alternative Carriers. *Arch Microbiol*. 176: 224-230.
[14] Compos-Bermudez, V. A., Bologna, F. P., Andreo, C. S., Drincovich, M. F. 2010. Functional Dissection of Escherichia coli Phosphotransacetylase Structural Domains and Analysis of Key Compounds Involved in Activity Regulation. *FEBS J.* 277: 1957-1966.
[15] Miller, R. D., Dykhuizen, D. E., Green, L., Hartl, D. L. 1984. Specific Deletion Occurring In The Directed Evolution of 6-Phosphogluconate Dehydrogenase in Escherichia coli. *Genetics*. 108: 765-772.
[16] Pharkya, P., Maranas, C. D. 2005. An Optimization Framework for Identifying Reaction Activation/Inhibition or Elimination Candidates for Overproduction in Microbial Systems. *Metab. Eng.* 8: 1-1.