Int. J. Advance Soft Compu. Appl, Vol. 8, No. 2, July 2016
ISSN 2074-8523

A Hybrid of Ant Colony Optimization and Flux Variability Analysis to Improve the Production of L-Phenylalanine and Biohydrogen

Amira Husna Azhar[1], Mohd Saberi Mohamad[1], Safaai Deris[2], Suhaimi Napis[3], and Richard Sinnott[4]

[1]Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai 81310, Malaysia
e-mail: amira_husna@ymail.com, saberi@utm.my

[2]Department of Creative Technology (Multimedia), Faculty of Technology Creative and Heritage, Universiti Malaysia Ketantan, 16300, Bachok, Kelantan, Malaysia

[3]Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

[4]Department of Computing and Information System, University of Melbourne, Victoria, 3010, Australia

Abstract

In silico metabolic engineering has shown many successful results in genome-scale model reconstruction and modification of metabolic network by implementing reaction deletion strategies to improve microbial strain such as production yield and growth rate. While improving the metabolites production, optimization algorithm has been implemented gradually in previous studies to identify the near-optimal sets of reaction knockout to obtain the best results. However, previous works implemented other algorithms that differ than this study which faced with several issues such as premature convergence and able to only produce low production yield because of ineffective algorithm and existence of complex metabolic data. The lack of effective genome models is because of the presence thousands of reactions in the metabolic network caused complex and high

dimensional data size that contains competing pathway of non-desirable product. Indeed, the suitable population size and knockout number for this new algorithm have been tested previously. This study proposes an algorithm that is a hybrid of the ant colony optimization algorithm and flux variability analysis (ACOFVA) to predict near-optimal sets of reactions knockout in an effort to improve the growth rates and the production rate of L-phenylalanine and biohydrogen in Saccharomyces cerevisiae and cyanobacteria Synechocystis sp PCC6803 respectively.

Keywords: *metabolic engineering; ant colony optimization; flux variability analysis; L-phenylalanine; biohydrogen.*

1 Introduction

Metabolic engineering (ME) has shown a big improvement and is getting more popular in these recent years. Metabolic engineering has been used to study and manipulate the biological microbial cell metabolism by many researchers in this field [1]. An example of a strategy that has been introduced by metabolic engineering method is by suggesting any genes or reactions from its complex metabolic network to be deleted [2,3]. This technique has shown many achievements towards genome-scale metabolic network (GSMN) model in addressing high yield of by-product secretion and cell growth rate.

The advancement of metabolic engineering has gained more attention as it is able to improve any desired metabolites strain that can further the process to become valuable products for market industries purposes [4]. With such results, more developments of quantitative models and algorithms using computational approach are growing in these recent years. However, the current production rate is still low than its theoretical maximum value probably due to the lack of effective computational technique developed to modify the metabolic model of microorganisms. Even modification toward biological network of an organism has been proven to be a successful technique, constructing a mutant strain of genome model is a challenging part to overproduce the interest metabolites beyond its wild type limit.

Computational time is increasing as the problem size increases, thus some computationally pre-process steps are required, which matches with the biological theory to have more suitable and compatible data. Other than that, the models need to undergo an optimization process in order to prevent the solution from being trapped in local optima, caused premature convergence. The innate potential for higher production of desired metabolites is obscure, and relates to the lack of effective genome models. This is mainly due to the fact that in producing

metabolites by microorganisms the presence of interaction among thousands of reactions in the metabolic network caused higher dimensional data size. In addition, the presence of competing pathway of non-desirable product may affect the desired metabolite production.

The reason of using advance computational approach in today's research is because it is powerful and able to save costs and time in manipulating phenotypes to enhance desired strain model compared to wet lab procedures. The goal of metabolic engineering is to develop effective methods in order to improve the metabolic capabilities producing desired metabolites in microorganism for industrial purpose [5]. In metabolic engineering, there are some recognized methods available that are currently being used widely to assist in simulating process to improve metabolite production genome-scale model, for example, optimization algorithm, modeling simulation algorithm and modeling framework.

Microbial strains are microorganisms that are widely used to produce biochemical products, antibiotics, drug targets, therapeutic proteins, food ingredients, vitamins, fuels, and other useful chemicals that are in demand in many markets and industries. Therefore, this research is concerned about the production of metabolites L-Phenylalanine (L-Phe) in *Saccharomyces cerevisiae* and biohydrogen (H_2) in cyanobacteria *Synechocystis sp* PCC6803. There is a growing concern about energy generation by fossil fuels to be continued such as biohydrogen because it is renewable and shows the tendency to be an alternative to fossil fuels for transportation [6]. While L-phenylalanine has been used as a precursor to vanillin production for food additives and also one of good nutritional supplements in pharmaceutical industry [7].

During last decades, computational approach gaining great attention from biologist to apply use to analyse genome scale model. According to [8], the constraints-based modeling is the most used method in elucidating the cell network. Besides that, constraint based modelling is tend to predict the intracellular fluxes of an objective function of stoichiometric model in steady-state condition. Flux Balance Analysis (FBA) is proven to be one of the effective tool that can be used to analyse fluxes of metabolic network and the effect of knockout on metabolism [9]. Simultaneous of the development FBA, there are some other method developed by researchers such as regulated FBA (rFBA), dynamic FBA (dFBA), integrated dynamic FBA (iFBA), and parsimonious FBA (pFBA) [10]. FVA is an extension of FBA which it is a promising technique for identification of possible minimum and maximum fluxes of reactions [11]. The enhancement of FVA, called fastFVA is then proposed to speed up the implementation of previous FVA [12]. MOMA is another method proposed that has been widely used by researches to analyse the effect of gene perturbation of mutant models [13]. ROOM is designed to analyse and predict the metabolic flux after the model is exposed to mutation. ROOM finds a flux distribution that satisfies the same constraints as FBA while minimizing the

number of significant flux changes. Same with MOMA, ROOM also aim to find minimize set of fluxes that are close to wild type without too concerning to minimize the growth rate [14].

Other than constraint-based methods, there are some famous and common rational modeling frameworks used by researchers in metabolic engineering area. These frameworks applies bi-level optimization strategy which it targets to solve two conflicts that biologically related which is the cell growth and maximum bioengineering objective. Modeling framework applied two key points in their calculation which is optimization strategy and modeling simulation. For example are OptKnock [15] and ReacKnock [16] which are aimed at identifying reaction (gene) knockout to obtain improved metabolites. OptKnock targets the set of reaction (gene) to be deleted without affecting the internal flux distribution of the genome model so that the flux of the metabolites could be optimized in a nonviable growth rate. OptKnock framework formulates the *in silico* design problem by using Genetic Algorithm (GA) and use FBA, MOMA or other method to calculate fitness if an individual [17]. While Reacknock applying FVA to predict the possible production rate under knockout strategies. RobustKnock [18] predicts more robust solution than OptKnock by predicting gene to be deleted that lead to the over-production of chemicals of interest. OptGene [19] is an another algorithm that search for reaction candidates that can be activated, inhibited or eliminated by using evolutionary search procedure for solving the resulting combinatorial optimization problem.

In this paper, a hybrid of Ant Colony Optimization and Flux Variability Analysis (ACOFVA) is proposed. This newly hybrid algorithm is developed to predict the reaction knockout strategy. Ant Colony Optimization (ACO) is a swarm intelligence based optimization algorithm which is inspired by the behavior of ants to find food which has been introduced in the year 1991 by [20]. While Flux Variability Analysis (FVA) is a constraint-based modeling approach developed by [21].

2 Methods

In this paper, the presence of competing reactions and genes in a metabolic network is suggested to be removed in order to overproduce interest metabolite beyond its wild type limit. The removal of selected reactions and genes is to reduce the size of genomes. This kind of strategy has been used by other researchers. However, a new algorithm is proposed here with some modifications in the way of calculating the objective function value.

2.1 Problem Formulation

The problem is to find the best set of reaction knockout including the associated genes of biological models which can be formulated as follows: A genome-scale model data is present by a stoichiometric matrix which consists of; where *m* and *n* are the representation of number of metabolites and reactions in the network respectively shows in equation 1.

$$S_{m \times n} \tag{1}$$

$$S.v = 0 \tag{2}$$

In equation 2, *v* represents the vector for overall fluxes including the internal, transport and the growth fluxes which are sure to achieve a steady state condition to bring the zero value. A steady state is referred to the total amount of any compound being produced that must equal to the total amount that is being consumed. A constraint-based method applying dynamic mass that can be formulated in equation 3 is as followed, where *t* is time:

$$\frac{dx}{dt} = Sv \tag{3}$$

There are two additional constraints that are commonly used in this modelling simulation comprising the upper and lower boundaries of internal fluxes denoted by *v* and exchange fluxes present by vector *b* as state in equation 4:

$$\begin{array}{c} \text{lowerbound} \quad \text{upperbound} \\ \overbrace{\quad}\quad\overbrace{\quad} \\ 0 < v < \infty \\ -\infty < b < \infty \end{array} \tag{4}$$

2.2 A hybrid of ACO and FVA

A hybrid of ACOFVA is a combination of optimization algorithm and constraint-based method is proposed to predict the best reaction knockout list. Table 1 shows the differences between existing algorithm and proposed method in this study. ACOFVA is proposed based on some potentials that can be handled by ACO and FVA in optimizing and modeling the metabolic network. For instance, ACO has the capability in finding the shortest path in constructing a best solution which is able to simultaneously perform local and global search to avoid local optima problem. Moreover, FVA is good in predicting the effects of the range of reaction's flux after the reaction deletion.

Table 1: Differences between the existing and proposed hybrid algorithm.

Algorithms	Optimization algorithm	Modeling simulation	Data	Target Metabolite
ACOFVA	ACO	FVA	*S. cerevisiae* and *Synechocystis*	L-Phe and biohydrogen
ACOFBA [22]	ACO	FBA	*S. cerevisiae*	vanillin
CBAFBA [23]	CBA	FBA	*S. cerevisiae*	vanillin
BAFBA [24]	BA	FBA	*E. coli*	succinate and lactate
GACOFBA [25]	GACO	FBA	*E. coli* and *S. cerevisiae*	succinate and lactate
BATFBA [26]	BAT	FBA	*E. coli*	succinate and lactate

Note: * The shaded column represents the hybrid algorithm proposed in this research.
* ACOFBA, Ant Colony Optimization and Flux Balance Analysis; CBAFBA, Continuous Bees Algorithm and Flux Balance Analysis; BAFBA, Bees Algorithm and Flux Balance Analysis; GACOFBA, Genetic Ant Colony Optimization and Flux Balance Analysis; BATFBA, Bat Algorithm and Flux Balance Analysis.

The proposed ACOFVA attempts to improve the traditional approach which can simultaneously combine the optimization technique and modeling method in modifying the metabolic network in order to improve the metabolite production in microorganisms. The main differences between existing algorithm and this newly proposed algorithm is the implementation of FVA for the modeling simulation which has not been used before. Next, the optimal percentage of a solution constructed is set to achieve the highest solution which reflects to the fitness of the objective function. This optimal percentage can only be set with the implementation of FVA. The *in silico* research about the interest metabolites production focused in this study is still lack. Thus this study gives deeper view on optimizing the production of desired metabolites. Figure 1 shows the flowchart of ACOFVA.

2.2.1 Model pre-processing

The first step before experiments are carried out, the dataset is undergoing pre-processing. The purpose of pre-processing model is to remove unnecessary reaction(s) and all dead end(s) involved in the model. These unused reactions that have been removed are not going to be used in this research in which the presence

may affect the result of this experiment and also the accuracy. In addition, this step also perform a check for consistency of the model after it has undergone the pre-process step.

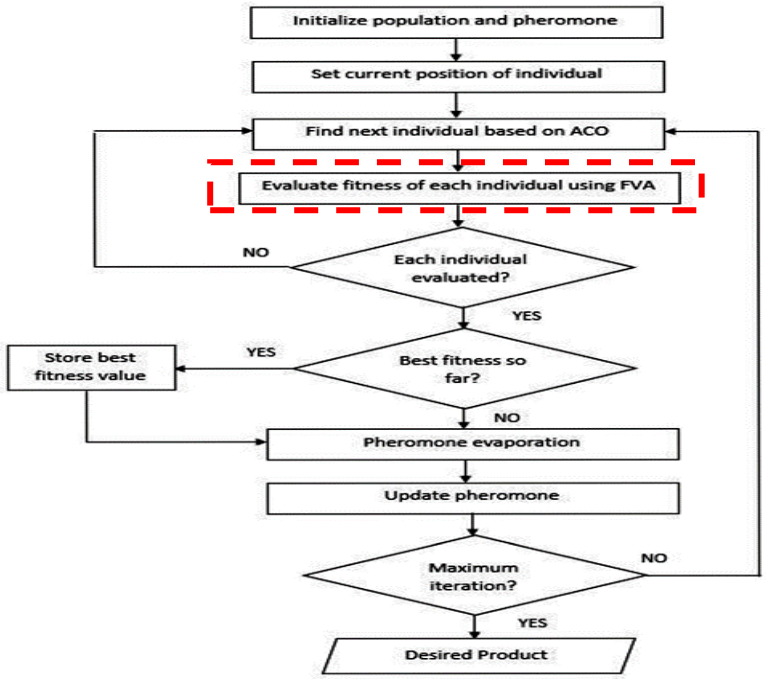

Fig. 1. Flowchart of ACOFVA.

2.2.2 Ant Representation

Each ant is then placed randomly on the nodes as the starting point. The nest is set to be the initial point from where the ants start their searching for the food source. From the nest, an ant travels along the nodes to reach the target point while constructing the best solution. In general, the concept of this algorithm is represented in Figure 2. Using this representation, ant traverses the routes to search for individual that manages to find the optimal reaction knockout. Ant uses transition rule and pheromone update rule as a guide to select for the next movement. Pheromone refers to the marker that is left at the trail by the ant colony. Each of the individual ant has its own pheromone value which actually shows the quality of it. Individual ant with higher pheromone is said to be the better solution.

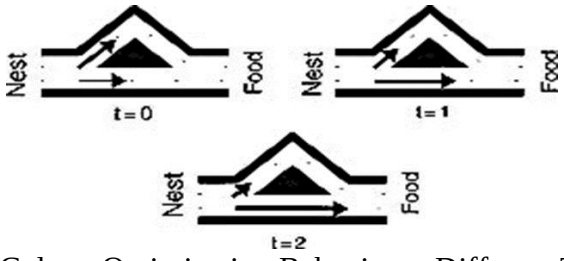

Fig. 2. Basic Ant Colony Optimization Behavior at Different Time Stamps [27]

2.2.3 Initialize population

The algorithm starts by initializing populations of *m* ants. Each ant is initialized as by assuming a reaction with *n* genes. A population is initialized by a predefined number of individuals by assigning present and absent status. The matrix creates the status randomly using bit '0' or '1' among them. The suggested reaction to be knocked out is represented by '1' while reaction that is not deleted and maintained is represented by '0'. The purpose of assigning those value is to select the best set of reaction to be deleted in order to enhance the production of a metabolite. Figure 3 shows the representation of gene and reaction of metabolic model to give a clearer view of the reaction deletion. ACO is applied at this stage where the number of ants is initialized as mentioned using bit number. This step is important to find the essential or non-essential reaction to optimize the desired metabolite. The non-desired reaction is selected to be knocked out and needs to make sure that the deletion of the set of reactions not defect the growth of a microorganism based on the biological facts.

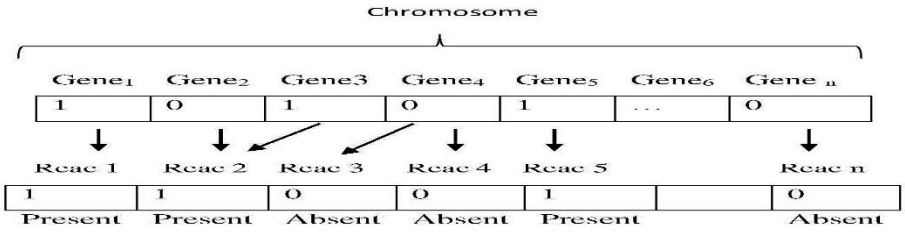

Fig. 3. The representation of reaction and gene of metabolic model.

2.2.4 Solution construction

Solution construction is important because optimal solution constructed is mainly contribute in optimizing the desired production. In this research, the solution construction is evaluated by fitness function of each individual. The next individual to be evaluated is decided by the previous ant. This involved the probabilistic calculation so that the next solution constructed is better. The constraint-based modeling simulation method of FVA is selected to calculate the scoring fitness of

individuals and is continuously repeated until all individuals are evaluated. Moreover, the bases for probabilistically constructing solutions is the implementation and use of heuristic information and in this study, FVA and the optimization rule of pheromone value are combined.

2.3.5 Scoring fitness of an individual

The fitness evaluation was carried out by using FVA. Figure 4 shows the basic flowchart of FVA to obtain the best production rate. FVA calculates the fitness of a mutant individual after undergoing a deletion reaction strategy. The calculation performed is based on steady state approximation of the internal metabolites concentration, which reduces the corresponding mass balances in an effective manner. When a list of reactions that is selected to knock out has been found and removed from current solution, FVA examines the model after the knockout process using linear objective function to determine the flux distribution.

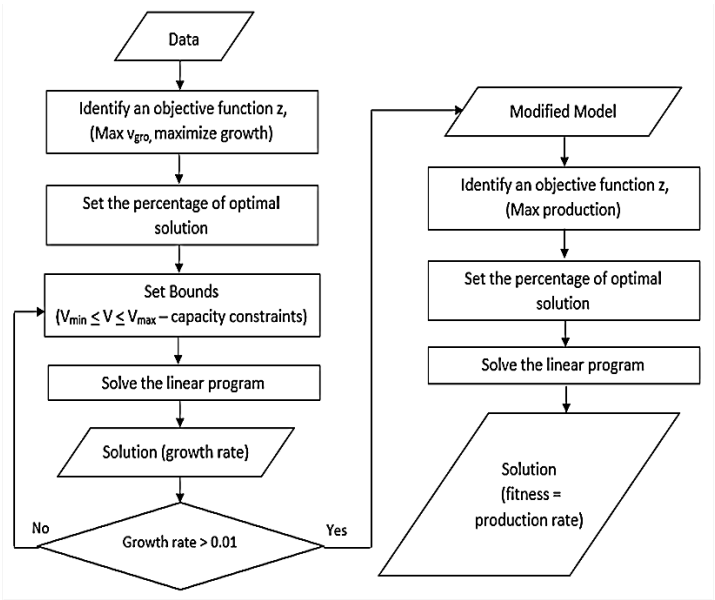

Fig. 4. The flow of FVA to obtain the best production rate.

Firstly, FVA is set to calculate the cellular growth rate of a model of the new solution referring to the mutant model constructed by ACO. The percentage of optimal solution can be set up to 100 %. In this study, the optimal solution is set to the highest percentage which is 100%. The growth rate that is higher than 0.01 is chosen to further continue the process to optimize the desired product flux at a fixed cellular growth rate which indicates that the new model has a viable growth rate. Then, the process is now focused on the production rate that has been set as the new objective function. The final solution is obtained at the last step which carries the

best production rate value. This production rate is used as the fitness. The flow of fitness evaluation is showed in Figure 4. The aim is to maximize the objective function of *f(x)* can be formulated as equation 6.

$$f(x) = production\ rate \quad \text{mmol g/DW/ h} \tag{6}$$

3 Datasets and Experimental Setup

Genome-scale model has been chosen as input data to perform the operation of the algorithm of ACOFVA. The two developed datasets Yeast (*Saccharomyces cerevisiae*) and Cyanobacteria (*Synechocystis sp* PCC6803) are selected. These models are well established and have been used by many researchers in metabolic engineering area to study the bacterium's metabolism and phenotypic behaviour. The *S. cerevisiae* model used is downloaded from a published literature by [28] derived from http://sourceforge.net/projects/yeast/files/yeast_4.05.zip/download and known as yeast.4.05 which contains 1865 reactions and 1319 metabolites. The other model is *Synechocystis sp* PCC6803 version iJN678 developed by [29] downloaded from bigg.ucsd.edu/ models/iJN678 that contains 863 reactions and 795 metabolites.

Before all the models are applied to the constructed ACOFVA algorithm, the pre-processed steps are carried out towards the models based on biological assumptions and computational approaches. During this step, some non-informative data inside the model are removed to obtain better results. For example, dead-end reactions are removed from the model to elude it from affecting the accuracy of the result while testing the model with the algorithm. In addition, this step also perform a check for consistency of the model after it has undergone the pre-process step. Table 2 shows the reducing of reactions and metabolites number after those model undergo the pre-processing steps.

Table 2. Comparison between raw and pre-processed model.

Model	**Number of Reactions**		**Number of Metabolites**	
	Raw Model	Pre-processed Model	Raw Model	Pre-processed Model
S. cerevisiae	1865	1461	1319	881
Synechocystis sp PCC6803	863	739	795	636

Three results are generated during this work which is a list of reactions knocked out, growth rate, and production rate, which are going to be explained in details in

the next section. The experiment was conducted using quad core of 3.60 GHz Intel Core i7 processor with 16 GB RAM workstation.

Parameter Setting for ACOFVA

The performance of any algorithm is largely depend on the setting of its algorithm dependent parameter [30]. The optimal setting should allow the algorithm to achieve the best performance for solving a range of optimization problems.

The glucose uptake rate is used as the sole carbon source for both models in this study. This research is emphasized to improve the metabolites production in heterotrophic condition. The glucose uptake rate is set at 10 mmol g DW-1 h-1 as it is the common glucose value that is suitable with Yeast organism. In addition, the amount is also the highest value of glucose uptake rate for Yeast *S. cerevisiae* [31]. According to [32], 0.85 mmol g DW-1 h-1 is the maximum bounds of glucose uptake rate for the heterotrophic growth condition for Cyanobacteria *Synechocystis* sp. PCC6803 model. Therefore, by setting high glucose uptake rate as sole carbon for the model to grow, the desirable production can be enhanced. The set amount has been used widely by computational and also wet lab experiment.

In addition, the population size and number of knockout for this new proposed algorithm are 100 and 5 respectively. These values have been tested and showed good results. Thus, it is selected as the most suitable parameter to have a better growth and production rates. All the results showed are from 30 number of runtime and iteration for each runtime is 100. The performance of any algorithm is largely depend on the setting of its algorithm dependent parameter [30]. The optimal setting should allow the algorithm to achieve the best performance for solving a range of optimization problems.

4 Results and Discussion

In this paper, a hybrid of ACOFVA is proposed to improve the result of some previous existing methods in elucidating phenotype behavior after modification towards metabolic network is carried out while improving the interest metabolite production. For the benchmarking function, mean and standard deviation (STD) of growth rate and production rate are calculated. Table 3 shows the result of mean and STD. Both datasets show low STD when tested with ACOFVA indicate that the result of each run is very close to the mean. It is apparent that the stability of the proposed algorithm is high as the difference between the results of each runtime is small. In addition, due to the set of reactions and genes knockouts performed, it could be hypothesized that each set of reactions might cause a specific and varying range of interactions that can affect the final cascading pathway for L-Phe and H_2 production.

Table 3. Mean and standard deviation for the growth rate of each metabolite.

Data	Mean	Standard Deviation
L-phenylalanine (*S. cerevisiae*)	1.1547	0.5054
Biohydrogen (*Synechocystis*)	0.0374	0.0024

Production of L-Phenylalanine in S. cerevisiae

Table 4 shows reactions and genes information that are suggested to be knocked out for L-Phenylalanine in *S. cerevisiae.* The list of reaction knocked out that can be used in real laboratory experiment to test its potential to improve L-Phe production is also provided. Some justifications towards the suggested reaction to be deleted are explained according to biological information.

Table 4. Reactions and genes information suggested to be knocked out for L-Phenylalanine in *Saccharomyces cerevisiae.*

Reaction Id	Reaction Description	Genes	Pathway
PC	Pyruvate carboxylase	PYC2	atp + hco3 + pyr --> adp + h + oaa + pi
ACACT4p	Acetyl CoA C-acetyltransferase	POT1	3odcoa + coa --> accoa + occoa
ACALDCD	Acetaldehyde condensation	PDC1	acald --> actn-R
GAT2_SC	Glycerol-3-phospate	GPT2	dcacoa + ddcacoa + dhap + hdcoa + ocdycacoa + odecoa + pmtcoa + stcoa + tdcoa --> 1agly3p_SC + coa
CERS	Fatty acid	SCS7	cer1_24 + h + nadph + o2 --> cer2'_24 + h2o + nadp

Comparison of the results obtained for L-Phe production using ACOFVA and other methods are showed in Table 5. As can be seen from the results, this method shows better results in terms of growth rate and production rate compared to other methods, including the previous researcher's work. The same type and version of the dataset model is applied for all the mentioned methods to make the comparison more reliable and fair. The results return by FVA is by the wild type (WT) model. It can be clearly seen that the value of growth rate is a bit lower than the WT. However,

the production rate of the modified model is far higher than the WT. This is because the modification towards the biological network is set to optimize the production of interest metabolite while maintaining the good growth. From the results, the growth rate of *S. cerevisiae* at 1.1547 h^{-1} it able to produce 5.7778 mmol gDW^{-1} h^{-1} of L-Phe.

Table 5. Comparison between different methods for growth rate and production rate of L-Phenylalanine in *Saccharomyces cerevisiae.*

Max. theoretical yield: 6.000

Method	Growth rate (h^{-1})	Production rate (mmol gDW^{-1} h^{-1})	List of knockout reactions and genes
ACOFVA	1.1547	**5.7778**	Reactions Id: PC, ACACT4p, ACALDCD, GAT2_SC, CERS Genes: PYC2, POT1, PDC1, GPT2, SCS7
FVA	**1.7023**	0.1927	N/A
OptGene [17]	0.57	N/A	Genes: PDC1, GDH1
CBAFBA [23]	**1.7023**	0.19466	Genes: ARO4, BDH1, ARO10
ACOFBA [22]	**1.7023**	0.1947	Genes: BAT1, ARO10, MDH1

Note: * The bold numbers represent the best result. N/A - Not applicable.
* mmol gDW^{-1} h^{-1} is millimoles per gram dry cell weight per hour.
* The shaded column represents the hybrid algorithm proposed in this research.

Aforementioned, a set of reactions with the associated genes is knocked out to obtain the better value of the objection function. Reaction pyruvate carboxylase (PC) with a gene PYC is involved in a cytoplasmic enzyme to convert pyruvate to oxaloacetate. The removal of PYC, makes the production of competitive metabolite become inactive such as succinate as NADH is not utilized in the TCA cycle. Gene POT1 plays a role in various pathways, including valine, leucine, and isoleucine degradation. Thus, by removing this reaction and its corresponding genes, the mentioned metabolites cannot be produced. Pyruvate decarboxylase (PDC) involved in amino acid catabolism and the fermentation of glucose to ethanol and the removal of PDC caused the ethanol formation cannot be carried out. GPT2 and SCS7 are another genes suggested to be knockout. This information is obtained from known databases such as Saccharomyces Genome Database (SGD) and Kyoto Encyclopedia of Genes and Genomes (KEGG). In addition, Figure 5 shows the convergence graph for L-Phenylalanine production. The graph shows that this

algorithm can converge faster which takes nearly at iterations 30 although the algorithm is allowed to run until 100 iterations. These results are used as the benchmark function of the algorithm.

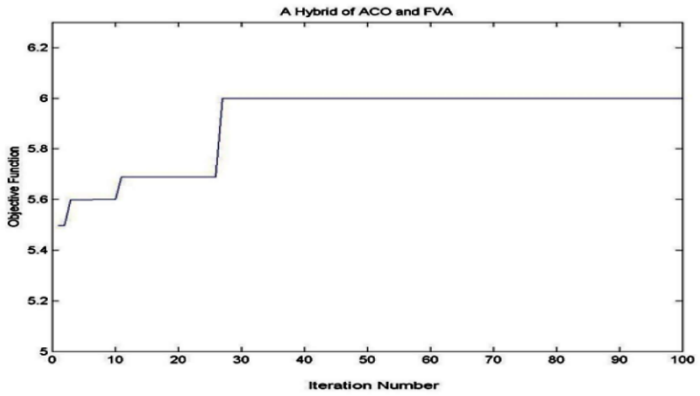

Fig. 5. ACOFVA convergence graph for production of L-Phenylalanine.

Production of biohydrogen in Synechocystis sp PCC6803

Table 5 shows reactions and genes information that are suggested to be knocked out for biohydrogen in *Synechocystis*.

Table 5. Reactions and genes information suggested to be knocked out for biohydrogen in *Synechocystis sp* PCC6803.

Reaction Id	Reaction description	Genes	Pathway
PDH	Pyruvate dehydrogenase	sll1721	nad + coa + pyr -> nadh + co2 + accoa
ENO	Enolase	slr0752	2pg <=> h2o + pep
ACKr	Acetate kinase	sll1299	acetyl-CoA => acetate
SUCCtpp	Succinate transport via diffusion (periplasm to cytosol)	N/A	h + succ <=> h + succ
ALCD2y	Alcohol dehydrogenase	slr0942	nadp[c] + etoh[c] <=> h[c] + nadph[c] + acald[c] an alcohol + NADP+ = an aldehyde + NADPH + H+
Note: N/A - Not applicable.			

Comparison of the results obtained for H2 production using ACOFVA and other methods are showed in Table 6. The results return by FVA is by the wild type (WT)

model. From the table, it can be clearly seen that the value of growth date is a bit lower than the WT. However, the production rate of the modified model is far higher than the WT. For the *Synechocystis* model, at growth 0.0374 h-1 this model can produce 4.0607 mmol gDW-1 h-1 of biohydrogen. This is because the modified model has removed some competitive reactions that could affect the acquisition of high production of interest metabolite. Figure 6 shows the convergence graph for biohydrogen production. The graph shows that this algorithm can converge faster which takes nearly iterations 40 although the algorithm is allowed to run until 100 iterations.

Table 6. Comparison between different methods for growth rate and production rate of biohydrogen in *Synechocystis sp* PCC6803.

Max. theoretical yield: 9.7665

Method	Growth rate (h^{-1})	Production rate (mmol gDW^{-1} h^{-1})	List of knockout reactions and genes
ACOFVA	0.0374	**4.0607**	Reactions Id: PDH, ENO, ACKr, SUCCtpp, ALCD2y Genes: sll1721, slr0752, sll1299, slr0942
FVA	**0.0632**	9.19E-05	N/A
FBA [33]	0.019	3.195	N/A

Note: * The bold numbers represent the best result. N/A - Not applicable.
* mmol gDW^{-1} h^{-1} is millimoles per gram dry cell weight per hour.
* The shaded column represents the hybrid algorithm proposed in this research.

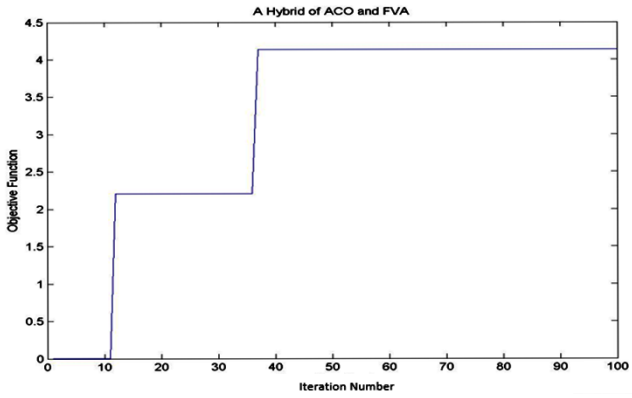

Fig. 6. ACOFVA convergence graph for production of biohydrogen.

Pyruvate dehydrogenase (PDH) is located in the mitochondria. This reaction consists gene sll1721. PDH is necessary to convert Pyruvate to acetyl CoA in the presence of oxygen so that carbons from pyruvate can go through TCA cycle to

produce ATP. Thus by removing this reaction, conversion of pyruvate to acetyl-CoA cannot be carried out as reflected in the production of other competitive metabolites that are present in that pathway. Removing ACKr also causes the reaction to carry acetyl-CoA as unable. Thus, pyruvate can only focus on the production of H_2. slr0752 is a protein-coding gene which is involved in glycolysis. This gene helps in converting glycerate to phosphoenolpyruvate. By deleting the enolase reaction and its gene, phosphoenolpyruvate cannot be used to catalyse the pyruvate production which then lactates or ethanol is not able to be produced. SUCCtpp is a reaction for subsystem of transportation of succinate via diffusion from periplasm to cytosol. This reaction is removed because it is not an important path to overproduce the interest metabolites. However, this reaction does not contain any gene because it is a transport reaction as mentioned. Reaction alcohol dehydrogenase with gene slr0942 is other reaction (gene) suggested to be knocked out. All of this information is according to KEGG.

Conclusion and Future Works

The advancement of computational biology use in metabolic engineering area to predict study about genome-scale metabolic model has become more popular by the days. Metabolic flux analysis shows many successful results in studying genome data. However, the accuracy and efficiency of existing methods are still not viable, thus setting a challenge that calls for an action for improvement. These two points are the crucial keys in developing any computational methods to obtain better results in predicting and observing the phenotype of interest in the biotechnology field. Consequently, it fulfills the industrial market needed such as for biofuel, food and the pharmaceutical sectors. In this paper, a hybrid of modeling simulation of constraint-based method and optimization algorithm which is ACOFVA has been proposed to overcome the issues mentioned. This algorithm has successfully predicted several reactions and genes that can be knocked out in order to improve and optimize the production of desired metabolites in a viable growth rate. This particular research only focuses on the production rate of L-Phe and H_2 with some information of reaction and genes suggested to knockout. The list of reaction and genes knockout need to be further validated through wet lab experiment. ACOFVA has successfully improved the performance of other existing methods such as FBA, FVA, CBAFBA and also ACOFBA. This is because the newly developed algorithm has implemented some suitable parameters in the algorithm that was tested earlier. It directly improves the accuracy while searching for the best value of an objective function. Based on the experiment, the results showed that the L-Phe and H_2 produced were higher than its wild-type models. For the future plan, Biomass per Coupled Yield (BPCY) as a fitness evaluation can be calculated to test the accuracy

of phenotype for other metabolites, too.

ACKNOWLEDGEMENT
We would like to express gratitude to the editors and reviewers for helpful suggestions and Universiti Teknologi Malaysia for sponsoring this research by the GUP Research Grant (Grant Number: Q.J130000.2528.12H12) and FRGS Research Grant (Grant Number: R.J130000.7828.4F720).

References

[1] Chartrain, M., Salmon, P. M., Robinson, D. K., & Buckland, B. C. (2000). Metabolic engineering and directed evolution for the production of pharmaceuticals. *Current Opinion In Biotechnology*. *11*(2), 209-214.
[2] Mahalik, S., Sharma, A. K., & Mukherjee, K. J. (2014). Genome engineering for improved recombinant protein expression in Escherichia coli. *Microbial Cell Factories, 13*(1), 1.
[3] Kumar, R.R., & Prasad, S. (2011). Metabolic Engineering of Bacteria. *Indian Journal of Microbiology, 51*(3), 403-409.
[4] Atsumi, S., & Liao, J. C. (2008). Metabolic engineering for advanced biofuels production from Escherichia coli. *Current Opinion In Biotechnology*. *19*(5), 414-419.
[5] Wiechert, W. (2002). Modeling and simulation: tools for metabolic engineering. *Journal of Biotechnology*, *94*(1), 37-63.
[6] Veziroglu, A., & Macario, R. (2011). Fuel cell vehicles: State of the art with economic and environmental concerns. *International Journal of Hydrogen Energy*. *36*(1), 25-43.
[7] Liu, S., Xiao, M., Zhang, L., Xu, J., Ding, Z., Gu, Z., & Shi, G. (2013). Production of l-phenylalanine from glucose by metabolic engineering of wild type Escherichia coli W3110. *Process Biochemistry*. *48*(3), 413-419.
[8] Price, N. D., Papin, J. A., Schilling, C. H., & Palsson, B.O. (2003). Genome-Scale Microbial In Silico Models: The Constraints-Based Approach. *Trends in biotechnology*. *21*(4), 162-169.
[9] Lee, J. M., Gianchandani, E. P., & Papin, J. A. (2006). Flux Balance Analysis in the Era of Metabolomics. *Briefings in Bioinformatics, 7*(2), 140-150.
[10] Martins Conde, P. D. R., Sauter, T., & Pfau, T. (2016). Constraint Based Modeling Going Multicellular. *Frontiers In Molecular Biosciences, 3.*

[11] Mahadevan, R., & Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering, 5*(4), 264-276.

[12] Gudmundsson, S., & Thiele, I. (2010). Computationally Efficient Flux Variability Analysis. *BMC bioinformatics, 11*(1), 489.

[13] Segre, D., Vitkup, D., & Church, G. (2002). Analysis of Optimality in Natural and Perturbed Metabolic Networks. *Proceedings of the National Academy of Sciences* (Vol. *99*(23), pp. 15112-15117).

[14] Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off Minimization of Metabolic Flux Changes After Genetic Pertubations. *Proceedings Of The National Academy Of Sciences* (Vol. *102*(21), pp. 7695-7700).

[15] 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.

[16] 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.

[17] Patil, K., Rocha, I., Förster, J., & Nielsen, J. (2005). Evolutionary Programming as a Platform for In Silico Metabolic Engineering. *BMC bioinformatics, 6*(1), 1.

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

[19] Pharkya, P., & Maranas, C. D. (2006). An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering, 8*(1), 1-13.

[20] Colorni, A., Dorigo, M., & Maniezzo, V. (1991). Distributed optimization by ant colonies. In *Proceedings of the first European conference on artificial life* (Vol. 142, pp. 134-142).

[21] Xue, X. D., Cheng, X. D., Xu, B., Wang, H.L., & Jiang, C.P. (2010). The basic principle and application of Ant Colony Optimization algorithm. In *Artificial Intelligence and Education (ICAIE) International Conference* (pp. 358-360). IEEE.

[22] Lu, S. J., Salleh, A. H. M., Mohamad, M. S., Deris, S., Omatu, S., & Yoshioka, M. (2014). Identification of gene knockout strategies using a hybrid of an ant colony optimization algorithm and flux balance analysis to optimize microbial strains. *Computational Biology And Chemistry. 53,* 175-183.

[23] Yin, L. H., Choon, Y. W. , Mohamad, M. S. , Chai, L. E. , Chong, C. K., Abdullah, A., Deris, S., & Illias, R. M. (2013). Prediction of Vanillin and Glutamate Productions in Yeast Using a Hybrid of Continuous Bees Algorithm and Flux Balance Analysis (CBAFBA). *Current Bioinformatics*. *9*(3), 284-294.

[24] Choon, Y. W., Mohamad, M. S., Deris, S., & Illias, R. M. (2014). A hybrid of bees algorithm and flux balance analysis (BAFBA) for the optimisation of microbial strains. *International Journal of Data Mining and Bioinformatics*. *10*(2), 225.

[25] 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.

[26] Chua, P. S., Salleh, A. H. M., Mohamad, M. S., Deris, S., Omatu, S., & Yoshioka, M. (2015). Identifying a gene knockout strategy using a hybrid of the bat algorithm and flux balance analysis to enhance the production of succinate and lactate in Escherichia coli. *Biotechnology And Bioprocess Engineering*. *20*(2), 349-357.

[27] Tavares, Neto, R., & Godinho Filho, M. (2011). An Ant Colony Optimization approach to a permutational flowshop Scheduling Problem with outsourcing allowed. *Computers and Operations Research*. *38*(9), 1286-1293.

[28] Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W. B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N. J., Kell, D. B., King, R. D., Oliver, S. G., Stevens, R. D., & Mendes, P. (2010). Further developments towards a genome-scale metabolic model of yeast. *BMC Systems Biology. 4*(1), 145.

[29] Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings Of The National Academy Of Sciences* (Vol. *109*(7), pp. 2678-2683).

[30] Yang, X. S., Deb, S., Loomes, M., & Karamanoglu, M. (2013). A framework for self-tuning optimization algorithm. *Neural Computing & Applications. 23*(7-8), 2051-2057.

[31] Borodina, I., Kildegaard, K. R., Jensen, N. B., Blicher, T. H., Maury, J., Sherstyk, S., Schneider, K., Lamosa, P., Herrgård, M. J., Rosenstand, I., Öberg, F., Forster, J., & Nielsen, J. (2015). Establishing a synthetic pathway for high-

level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via β-alanine. *Metabolic Engineering. 27,* 57-64.

[32] Mao, L. & Verwoerd, W.S. (2013). Genome-Scale Stoichiometry Analysis to Elucidate the Innate Capability of the Cyanobacterium Synechocystis for Electricity Generation. *Journal of industrial microbiology & biotechnology*. *40*(10), 1161-1180.

[33] Nogales, J., Gudmundsson, S., & Thiele, I. (2013). Toward systems metabolic engineering in cyanobacteria. *Bioengineered. 4*(3), 158-163.