Identifying Gene Knockout Strategies Using a Hybrid of Bees Algorithm and Flux Balance Analysis for *in Silico* Optimization of Microbial Strains

Yee Wen Choon, Mohd Saberi Mohamad, Safaai Deris, Chuii Khim Chong, Lian En Chai, Zuwairie Ibrahim, and Sigeru Omatu

Abstract. Genome-scale metabolic networks reconstructions from different organisms have become popular in recent years. Genetic engineering is proven to be able to obtain the desirable phenotypes. Optimization algorithms are implemented in previous works to identify the effects of gene knockout on the results. However, the previous works face the problem of falling into local minima. Thus, a hybrid of Bees Algorithm and Flux Balance Analysis (BAFBA) is proposed in this paper to solve the local minima problem and to predict optimal sets of gene deletion for maximizing the growth rate of certain metabolite. This paper involves two case studies that consider the production of succinate and lactate as targets, by using *E.coli* as model organism. The results from this experiment are the list of knockout genes and the growth rate after the deletion. BAFBA shows better results compared to the other methods. The identified list

Yee Wen Choon · Mohd Saberi Mohamad · Safaai Deris · Chuii Khim Chong · Lian En Chai
Artificial Intelligence and Bioinformatics Group, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: ywchoon2@live.utm.my, saberi@utm.my, safaai@utm.my, ckchong2@live.utm.my, lechai2@live.utm.my

Zuwairie Ibrahim
Department of Mechatronics and Robotics, Center for Artificial Intelligence and Robotics (CAIRO), Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia
e-mail: zuwairiee@fke.utm.my

Sigeru Omatu
Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan
e-mail: omatu@rsh.oit.ac.jp

S. Omatu et al. (Eds.): Distributed Computing and Artificial Intelligence, AISC 151, pp. 371–378.
springerlink.com

suggests gene modifications over several pathways and may be useful in solving challenging genetic engineering problems.

Keywords: Evolutionary Programming, Metabolic Engineering, Bees Algorithm, Gene Knockout, Optimization.

1 Introduction

Microbial strains are strains of microorganisms which have become popular for genome-scale metabolic networks reconstructions in recent years [1]. Reconstructions of the metabolic networks are found to be very useful in health, environmental and energy issues [2]. A vast numbers of high-throughput experimental data has expedited the development of computational models for simulating the actual processes inside the cell. One of the main goals in system biology is to construct an efficient and accurate pathway models that may be useful in predicting cellular responses and providing better understanding of complex biological functions.

Many algorithms were developed in order to identify the gene knockout strategies for obtaining improved phenotypes. Maranas *et al.* [3, 4] developed the first rational modeling frameworks (named OptKnock) for introducing gene knockout leading to the overproduction of a desired metabolite. OptKnock identifies a set of gene (reaction) deletions to maximize the flux of a desired metabolite without affecting the internal flux distribution such that growth is optimized.

OptKnock is implemented by using mixed integer linear programming (MILP) to formulate a bi-level linear optimization that is very promising to find the global optimal solution. OptGene is an extended approach of OptKnock which formulates the *in silico* design problem by using Genetic Algorithm (GA). These meta-heuristic methods are capable in producing near-optimal solutions with reasonable computation time, furthermore the objective function that can be optimized is flexible. SA is then implemented to allow the automatic finding of the best number of gene deletions for achieving a given productivity goal. However, SA faces the problem of falling into local minima far from the global optimum solution.

In this paper, a hybrid of Bees Algorithm and Flux Balance Analysis (BAFBA) is proposed to predict the gene knockout strategies. Bees Algorithm (BA) is a typical meta-heuristic optimization approach which was introduced by [5]. The search process of BA is based on the intelligent behaviors of honey bees. BA is proven to be efficient in solving optimization problems in the previous studies [5]. While the Flux Balance Analysis (FBA) approach which is used to calculate the fitness function is based on a steady state approximation to concentrations of the internal metabolites, which reduces the corresponding mass balances to a set of linear homogeneous equations. There are two advantages of BAFBA. First, BAFBA requires less computational time and thus it capable to solve larger size problems. Secondly, BA is capable of performing local and global search simultaneously and thus it works out the local minima problem. This paper presents the results obtained by BAFBA to two case studies where *E.coli* is the

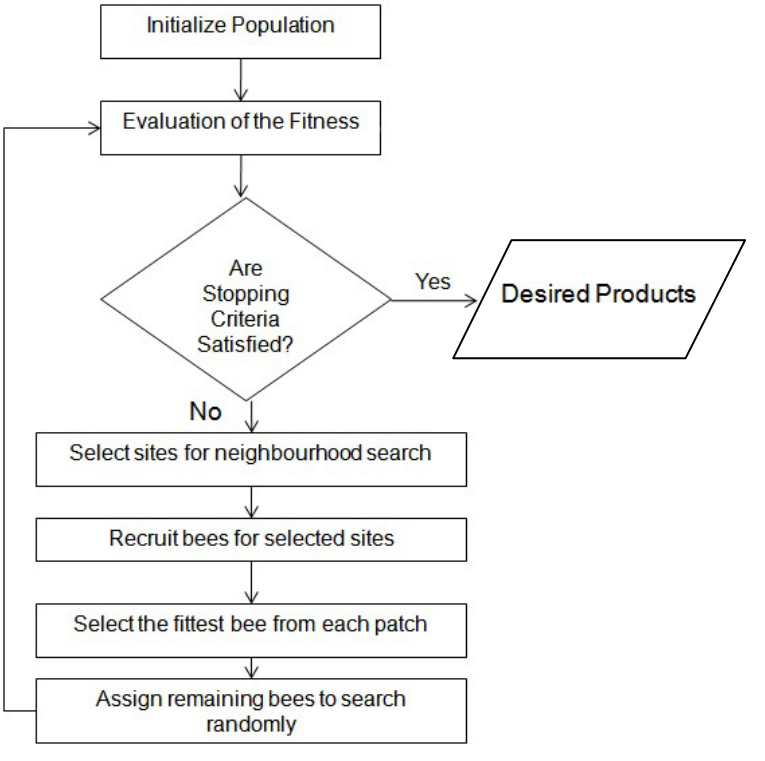
target microorganism. This paper also evaluates the performance of BAFBA for identifying gene knockout strategies with existing tools and compares the performance of BA with the existing methods within experimental approaches.

2 A Hybrid of Bees Algorithm and Flux Balance Analysis

In this paper, BAFBA is proposed to predict the gene knockout. Fig 2.1 shows the flow of a basic BA. The flow of BAFBA is presented in Fig 2.2. The important steps are explained in the following subsections.

2.1 Bee Representation of Metabolic Genotype

In the metabolic model, one or more genes can be found in each reaction. In this proposed method, each of those genes is represented by a binary variable indicating its absence or presence (0 or 1), these variables form a 'bee' representing a particular mutant that lacks some metabolic reactions when compared with the wild type (Fig 2.3.)



Note: Desired products represent the gene to be knockout.

Fig. 2.1 Flowchart of a basic BA.

2.2 Initialization of the Population

The algorithm starts with an initial population of n scout bees. Each bee is initialized as follows: assume that a reaction with n genes. Bees in the population can be initialized by assigning present or absent status to each gene randomly.

2.3 Scoring Fitness of Individuals

The fitness computation process is carried out for each site visited by a bee through FBA (Fig 2.4.). Cellular growth is defined as the objective function Z, vector **c** is used to select a linear combination of metabolic fluxes to include in the objective function, **v** is the flux map and i is the index variable (1, 2, 3, …, n).

Maximize Z, where

$$Z = \sum_i c_i v_i = \mathbf{c}.\mathbf{v} \tag{2.1}$$

where c = a vector that defines the weights for of each flux.

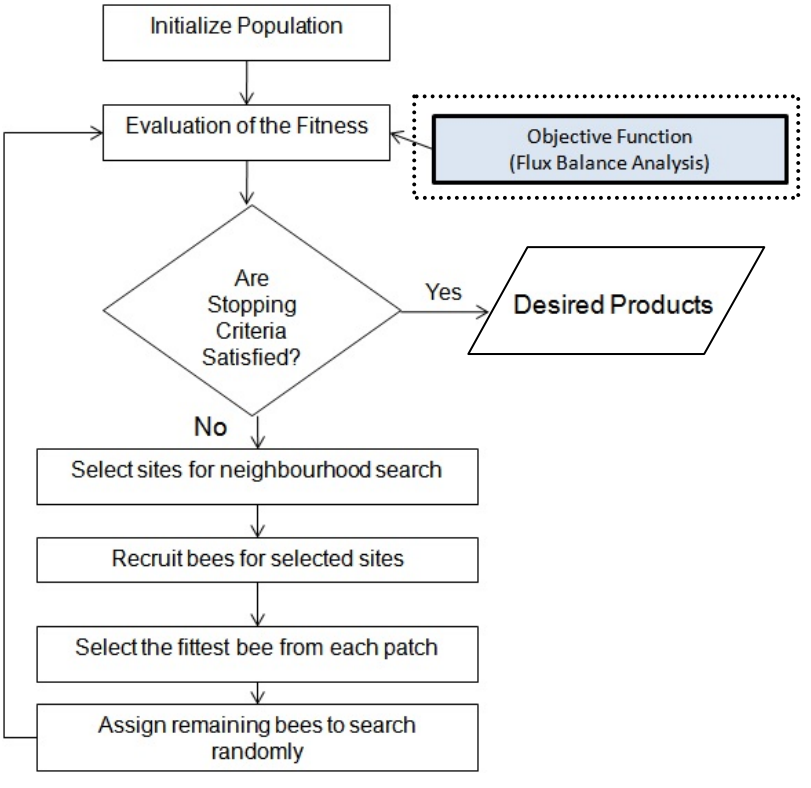
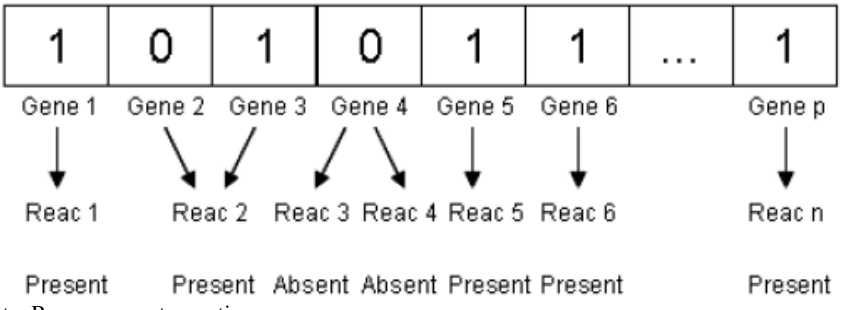


Fig. 2.2 The flow of BAFBA.



Note: Reac represents reaction.

Fig. 2.3 Bee representation of metabolic genotype

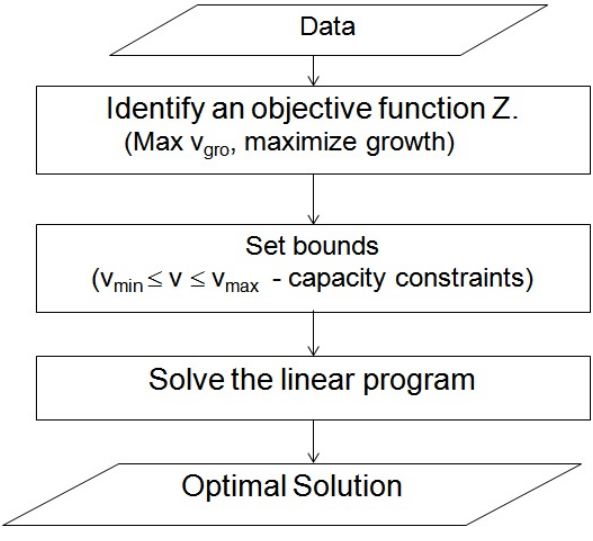


Fig. 2.4 Steps in FBA

2.4 Neighbourhood Search

The algorithm carries out neighbourhood searches in the selected sites, assigning more bees to search near the best sites. The bees can be chosen directly according to their fitnesses associated with the sites they are visiting. Searches in the neighbourhood of the best sites which represent more promising solutions are further more detailed by recruiting more bees to follow them than other selected bees.

2.5 Randomly Assigned and Termination

The remaining bees in the population are assigned randomly around the search space scouting for new potential solutions. These steps are repeated until a

stopping criteria is met. The stopping criteria are either the maximum loop value is met or the fitness function has converged. At the end of each iteration, the colony produces two parts to its new population – representatives from each selected patch and other scout bees assigned to conduct random searches.

3 Experimental Results

In this paper, the *E.coli* dataset is used to test on the operation of BAFBA. All simulations were performed for aerobic minimal media conditions. The glucose uptake rate was fixed to 10 mmol/gDW/hr while a set non-growth associated maintenance of 7.6 mmol ATP/gDW/hr. The results obtained are compared to the previous works reported in the literature studies [3, 6]. Millimole (mmol) is the unit of concentration whereas millimoles per hour (mmol/hr) is used as the unit measurement in the experiments.

Table 3.1 and Table 3.2 summarize the results obtained from BAFBA. As shown from the results, this method has produced better results to the previous works. In this paper, potential genes which can be removed are identified.

BAFBA suggests the removal of three reactions from the network results in succinate growth rate reaching 0.91665 which is better than the other two methods. The list obtained is to disable the phosphotransferase system which causes the network to rely solely on glucokinase for glucose uptake [3].

Table 3.1 Comparison between different methods for production of Succinate

Method	Growth Rate (mmol/hr)	List of knockout genes
BAFBA	0.91665	ACALD, ACKr, ATPM
SA + FBA [6]	0.35785	MALS, ORNDC, FUM, GLYCL, GHMT2, ADPT, DCYTD, DUTPDP, URIDK2r, NTD8, PUNPI, THD2, GND, PFL, SUCFUMt
OptKnock [3]	0.31	PYK, ACKr, PTAr, Phosphotransferase system

Note: The shaded column represents the best result.

Next, BAFBA is applied to identify knockout strategy for producing lactate. Table 3.2 shows the best result is obtained from this method is 0.91665. From the list of knockout genes, it can be concluded that the flux toward lactate at the maximum biomass yield is redirected by blocking acetate and ethanol production [3]. BAFBA produced the best results in both cases, and this is because BA performs local and global search simultaneously to avoid being trapped at locally optimal solutions. BA splits the search into exploration and exploitation, which are then executed parallely rather than serially like SA. Thus, BA performs better than SA where it solves the local minima problem faced by SA.

In addition, Table 3.3 and Table 3.4 show the results of three of the identified gene knockout strategies for succinate and lactate overproduction.

Table 3.2 Comparison between different methods for production of Lactate

Method	Growth Rate (mmol/hr)	List of knockout genes
BAFBA	0.91665	ACALDt, ALCD2x, ATPM
SA + FBA [6]	0.39850	ACLD19, DRPA, GLYCDx, F6PA, TPI, LDH_D2, EDA, TKT2, LDH_D-
OptKnock [3]	0.28	ACKr, PTAr, ACALD

Note: The shaded column represents the best result.

Table 3.3 Result of different knockout strategies for production of Succinate

Mutants	Growth Rate (mmol/hr)	List of knockout genes
A	0.87392	ACALD, ACKr
B	0.91665	ACALD, ACKr, ATPM
C	0.87392	ACALD, ACALDt, ACKr, ALCD2x

Table 3.3 shows three of the identified gene knockout strategies (i.e., mutants A, B, and C). For the production of succinate, acetate kinase (ACKr) which contributes to the phosphotransferase system for all three mutant A, B, and C is disabled, this causes the network to rely exclusively on glucokinase for glucose uptake [3]. The deletion of acetaldehyde dehydrogenase (ACALD) results earlier coupling of succinate with biomass yields. For mutant C, the additional deletion of alcohol dehydrogenase (ALCD2x) eliminated the production of ethanol.

Table 3.4 Result of different knockout strategies for production of Lactate

Mutants	Growth Rate (mmol/hr)	List of knockout genes
D	0.87392	ACALD, ACKr
E	0.91665	ACALDt, ALCD2x, ATPM
F	0.91665	ACALD, ACALDt, ACKr, ATPM

Table 3.4 shows the result of different knockout strategies for the production of lactate, phosphotransferase system for mutant D and mutant F are disabled. The additional deletion of ACALD results earlier coupling of lactate with biomass yields. For mutant E, the knockout strategy eliminated the competing byproduct (i.e, ethanol). In conclusion, the phosphotransferase system and ethanol affect greatly to both production of succinate and lactate.

4 Conclusion and Future Works

In this paper, BAFBA is proposed to predict optimal sets of gene deletion in order to maximize the production of certain metabolite. This method is based on BA, which is capable of performing local and global search simultaneously where the local minima problem faced by SA is worked out. The FBA approach is used as a fitness function whereby it is based on a steady state approximation to concentrations of the internal metabolites, which reduces the corresponding mass balances to a set of linear homogeneous equations.

Experimental results on *E.Coli* core model dataset obtained from literature [4] showed that BAFBA is effective in generating optimal solutions to the gene knockout prediction, and is therefore a useful tool in Metabolic Engineering.

In regard to further improve the performance of BAFBA, we are interested in applying an automated pre-processing operation in BAFBA to simplify the genome-scale metabolic model. Another interesting feature is the development of multi-objective optimization algorithms in a single run to achieve two goals, for example, maximizing the biomass and the desired product. Lastly, as BA employs many tunable parameters which are difficult for the user to select, it is important to find ways to help the user choose appropriate parameters.

Acknowledgments. This work is financed by Institutional Scholarship MyPhD provided by the Ministry of Higher Education of Malaysia. We also would like to thank Universiti Teknologi Malaysia for supporting this research by UTM GUP research grants (vot number: Q.J130000.7123.00H67 and Q.J130000.7107.01H29).

References

[1] Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B.O.: Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 72, 129–143 (2009), doi:10.1038/nrmicro1949

[2] Chandran, D., Copeland, W.B., Sleight, S.C., Sauro, H.M.: Mathematical modeling and synthetic biology. Drug Discovery Today Disease Models 5(4), 299–309 (2008), doi:10.1016/j.ddmod.2009.07.002

[3] Burgard, A.P., Pharkya, P., Maranas, C.D.: Opt. Knock: A bilevel programming framework for identifying gene knockout strategies for microbial strains optimization. Biotechnol. Bioeng. 84, 647–657 (2003), doi:10.1002/10803

[4] Pharkya, P., Burgard, A.P., Maranas, C.D.: Opt. Strain: a computational framework redesign of microbial production systems. Genome. Res. 14, 2367–2376 (2004), doi:10.1101/gr.2872004

[5] Pham, D.T., Ghanbarzadeh, A., Koç, E., Otri, S., Zaidi, M.: The bees algorithm – a novel tool for complex optimization problems. In: Proceedings of the Second International Virtual Conference on Intelligent Production Machines and Systems, pp. 454–461 (2006)

[6] Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., Rocha, I.: Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinformatics 9, 499 (2008), doi:10.1186/1471-2105-9-499

[7] Segre, D., Vitkup, D., Church, G.M.: Analysis of optimality in natural and perturbed metabolic networks. Proceedings of National Academy of Sciences of the United States of America 99, 15112–15117 (2002), doi:10.1073/pnas.232349399