Comparison of Optimization-Modelling Methods for Metabolites Production in *Escherichia coli*

Mee K. Lee / Mohd Saberi Mohamad / Yee Wen Choon / Kauthar Mohd Daud / Nurul Athirah Nasarudin / Mohd Arfian Ismail / Zuwairie Ibrahim / Suhaimi Napis / Richard O. Sinnott

**Abstract:**

The metabolic network is the reconstruction of the metabolic pathway of an organism that is used to represent the interaction between enzymes and metabolites in genome level. Meanwhile, metabolic engineering is a process that modifies the metabolic network of a cell to increase the production of metabolites. However, the metabolic networks are too complex that cause problem in identifying near-optimal knockout genes/reactions for maximizing the metabolite’s production. Therefore, through constraint-based modelling, various metaheuristic algorithms have been improvised to optimize the desired phenotypes. In this paper, PSOMOMA was compared with CSMOMA and ABCMOMA for maximizing the production of succinic acid in *E. coli*. Furthermore, the results obtained from PSOMOMA were validated with results from the wet lab experiment.

**Keywords:** Metaheuristic algorithms, Minimization of Metabolic Adjustment, Metabolic Engineering, Bioinformatics, Artificial Intelligence

**DOI:** 10.1515/jib-2019-0073

**Received:** August 30, 2019; **Revised:** February 5, 2020; **Accepted:** January 18, 2020

**Introduction**

In previous years, petroleum is used as a primary component in transportation, mining, industrial and others. However, due to the restrained reserve, petroleum is unable to withstand the increased demand for global and consumer products. Furthermore, the use of fossil fuels rears serious implications to the environment such as greenhouse effects. Hence, this induced the use of biomass as an alternative source considering it is renewable and available locally. These biomass resources include agronomic residues such as sugarcane waste, wheat or rice straw, and paper waste. The bioprocess is known as biomass fermentation. Microorganisms such as *E. coli* and *Saccharomyces cerevisiae* are able to produce succinic acid and ethanol in anaerobic condition. However, the amount of succinic acid and ethanol produced are still below the threshold.

Metabolic network consists of reactions between enzymes and metabolites occur in an organism that may help the biologists and researchers to understand the genotypic and phenotypic characteristics of a cell. With the advancement in genome sequencing, a detail organization of an organism can be deciphered, thus exploit the organisms for strains optimization. However, metabolic network is too complex, which resulted in high dimensionality of solution space, thus increasing the computational time exponentially.

Therefore, metabolic engineering has been an important factor for improving the production of various chemical substances by altering organisms. Recently, metabolic engineering has been improved by incorporating systems biology known as systems metabolic engineering. Systems biology provides a more conceptual understanding of metabolic enzymes and pathways, thus accelerate the formation or modification of pathways...
with regard to optimize the production of industrial metabolites [1]. One of the modifications is gene knockout, whereby a set of genes is removed from the mutant and the phenotypic effect is analyzed. The purpose of gene knockout is to ensure the flux will go towards the production of desired metabolites [2]. However, it is difficult to obtain a near-optimal set of genes knockout.

Therefore, the development of constraint-based methods has become a great achievement in metabolic engineering as they help to predict, analyze and interpret all the biological functions in the metabolic networks [3]. The first constraint-based method is Flux Balance Analysis (FBA) that discovers the behaviors of a metabolic network using the mathematical computation [4]. Hence, a higher level of abstraction needs new mathematical approaches to illustrate these biological processes. Eventually, this brings to the development of Minimization of Metabolic Adjustment (MOMA) and Regulatory On/Off Minimization (ROOM) [5], [6].

Both MOMA and ROOM are used to predict the steady-state of the mutant’s metabolic network after gene knockouts. However, there is a possibility that the steady-state obtained by ROOM is hardly being found by the organism. In this research, MOMA is chosen as modeling algorithm considering that FBA assumes the mutant organism is having the same optimal metabolic state as a wild-type organism [7]. Furthermore, MOMA is more suitable to predict the suboptimal flux distribution in mutant organisms. Still, MOMA lacks the optimization algorithm that is used to identify knockout genes that can maximize ethanol production. Hence, MOMA is hybridized with an optimization algorithm to analyze and predict the effect of genes knockout towards the overproduction of ethanol.

Metaheuristic algorithms have been proposed to improve the production of ethanol in E. coli [8]. Different metaheuristic algorithms have been applied to identify near-optimal genes knockout as metaheuristic algorithms are computationally less expensive. The first method that applies the metaheuristic algorithm is OptGene. OptGene applies Genetic Algorithm (GA) for searching and identifying a set of genes knockout that is evaluated by FBA [9]. Furthermore, OptGene introduced a new fitness function, which is Biomass-Product Couple Yield (BPCY). Following that, Simulated annealing (SA) and Set-based Evolutionary Algorithm (SEA) have been proposed to identify a set of genetic manipulations that resulted in increased desired phenotypes [10]. However, these methods produce over-optimistic solutions, solutions trapped in local optima and high computation time [9], [11], [12].

Several major advances in in silico metabolic engineering take different approaches. One of the development is multiobjective optimization that produces a set of non-dominated solutions between two competing objectives such as production rate and growth rate [13], [14], [15], [16]. Several methods have been developed to solve the issues of competing objectives, including Linear Physical Programming based Flux Balance Analysis (LPPFBA), Noninferior Set Estimation (NISE) with FBA, Genetic Design through Multi-objective Optimisation (GDMO) and others [17], [18], [19]. The advantages of these methods are the decision-makers, which are industrialists or biologists, may have various solutions instead of one single solution. Furthermore, the suggested knockout genes may produce mutant with higher growth rate as well as higher production rate.

In this paper, a comparative study of PSOMOMA, ABCMOMA and CSMOMA are shown in terms of production rate and growth rate of succinic in mutant E. coli. These algorithms were improved with MOMA as fitness function evaluation. The paper is organized as follows: Section 2 describes the metaheuristic algorithms, Section 3 provides the results and discussion and lastly is the conclusion of the paper.

2 Swarm Intelligence

Swarm intelligence was inspired by the foraging behavior of animals such as bees, ants, birds and fishes. The discipline focuses on the behaviors of animal interaction with one another and with the environment under a decentralized control system. At a high level, a swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve the same goal [20]. Foraging behaviors describe the movement of animals around their food resources or movement when finding their nest and mates. Besides, swarm intelligence provides a global optimization method that helps to solve complex problems in real life.

2.1 Particle Swarm Optimization (PSO)

PSO is an algorithm used to solve discrete and continuous optimization problems in a population. Traditional PSO was inspired by the social-psychology, such as bird flocking and fish schooling which is introduced by [21]. PSO involves the use of simple concepts and mathematical operators. Besides, PSO is similar to the genetic algorithm (GA), whereby the algorithm is initialized randomly. The only difference between GA and PSO is that PSO has particles, which is agents that move across the problem space. The population in PSO is known as “swarm”. Each particle has its own velocity and position at a certain instance. The different location of the
particle in the problem space indicates different possible solutions for a given optimization problem. Every particle will look for the best location in problem space by changing their velocity towards the best solution.

2.2 Artificial Bee Colony (ABC)

Artificial Bee Colony (ABC) algorithm was inspired by the foraging behavior of honeybee colony [22]. ABC consists of two modes of behavior, which are recruitment to a nectar source and the abandonment of a source. It consists of three main components: employed foragers, onlookers and scouts.

1. Employed foragers: are associated with the food sources that are currently exploited. They share the information such as distance and direction of the food sources with other bees waiting in the hive.

2. Onlookers: acquire the information from employed foragers and chooses the food source with higher nectar amounts.

3. Scouts: randomly search for new food sources (solutions) that are abandoned by the employed bees.

2.3 Cuckoo Search (CS)

Cuckoo search is based on the parasitic behavior of cuckoos in nature [23]. It incorporates a Levy flight strategy in finding the best solution. There are three rules in CS:

a. Only one egg can be laid in a nest at once

b. The nest with higher fitness will survive for the next generations

c. The probability of replacing and discovered by the host is between [0,1].

The rules above are used in searching operations of CS where the selection process is operated by Levy flight while the exploitation process is operated by applying the probability of \( p \in [0,1] \). The advantage of CS is the incorporation of Levy flight, which allows the new solutions to be generated far from the current best solution [23]. Due to this, there are fewer chances of solutions trapped in local optima. Therefore, a fraction of probability is imposed on the cuckoo egg. These metaheuristic algorithms have been compared and the advantages and disadvantages are represented in Table 1.

| Algorithm | Advantages | Disadvantages | Ref. |
|-----------|------------|---------------|------|
| PSO       | – Easy implement \n– No overlapping mutation calculation | – Easily suffers from the partial optimism | [24], [25], [26], [27] |
| ABC       | – Strong robustness \n– Fast convergence \n– High flexibility | – Premature convergence in the later search period \n– Accuracy of the optimal value may not meet the requirements | [7], [28], [29] |
| CS        | – Dynamic applicable (adapt to changes) \n– Easy to implement | – Easily trapped in local optima \n– Levy flight affects the convergence rate | [22], [30], [31] |

To improve metabolite production, the problem can be described as follows: MOMA is similar to FBA. Thus, metabolic network is represented in a stoichiometric matrix \( S \) of a size \( m \times n \), whereby \( m \) is the metabolites and \( n \) is the reactions. The matrix, \( S \) shows the relationship between reactions \( v \) of length \( n \) and concentrations \( x \) of length \( m \). FBA is used to evaluate the fitness, which is fluxes as shown in the equation below:

\[
\text{Flux vector, } v = (v_1, v_2, ..., v_n)^T
\]
Metabolites vector, $x = (x_1, x_2, ..., x_n)$

The fluxes are evaluated to time,

Mass balance equation, $\frac{dx}{dt} = S \times v$

where $T$ means transposed. FBA is used to calculate the flux distribution of wild-type and mutant, while MOMA is used to minimize the Euclidean distance between wild-type fluxes and mutant fluxes. Therefore, using linear programming, the objective of FBA is optimized as follows:

$$\max Z = c^T x$$

Subject to $S \times v$

where $v$ is flux vector and $c$ is a vector weight of coefficient reactions to be optimized. After FBA computation, by using quadratic programming, MOMA is used to minimize the distance between wild-type and mutant. The objective of MOMA is shown as follows:

$$\min \| v_{wt} - v_{mt} \|^2$$

$$\min \frac{1}{2} v_{mt}^T I v_{mt} + (-v_{wt}) \cdot v_{mt}$$

where $v_{wt}$ and $v_{mt}$ are flux distribution of wild-type and mutant, respectively. $I$ is the identity matrix of size $n \times n$ with length $v_{mt}$.

## 3 Materials and Methods

In this paper, PSOMOMA, ABCMOMA and CSMOMA have been validated with *E. coli* for maximizing the production of succinic acid. The glucose is used as the sole carbon and its uptake rate is set to 10 mmol gDW$^{-1}$ h$^{-1}$. MATLAB R2013b is used to implement these algorithms. Meanwhile, Constraints Based Reconstruction Analysis (COBRA) toolbox is used to model and analyse the metabolic model by MOMA. SBML Toolbox is used to read the file in SBML format. Table 2 shows the model used.

### Table 2: Numbers of reactions and metabolites involved before and after the model pre-processing.

| Model                | Number of reactions | Number of metabolites |
|----------------------|---------------------|-----------------------|
| Raw model            | 2583                | 1805                  |
| Pre-processed model  | 2342                | 1585                  |

## 4 Results and Discussion

The experimental result obtained from the hybrid of PSOMOMA algorithm is compared with the previous algorithms in enhancing the succinic production and the growth rate of *E. coli*. This section compares the growth rate and succinic production of *E. coli* from PSOMOMA with the previous results obtained for CSMOMA, ABCMOMA and also results from the wet laboratory [7], [11].

Table 3 below shows the result obtained for PSOMOMA, CSMOMA and ABCMOMA. The results showed that PSOMOMA achieves the highest growth rate compared to ABCMOMA and CSMOMA. Meanwhile, CSMOMA is able to found a mutant with the highest production rate of succinic acid. PSOMOMA able to found the highest production rate with 4 suggested genes knockout, CSMOMA with 5 genes knockout and ABCMOMA with 2 genes knockout.
Table 3: Result comparison on succinate production for PSOMOMA, CSMOMA and ABCMOMA.

| Method      | Gene knockouts         | Succinic production (mmol gDW\(^{-1}\) h\(^{-1}\)) | Growth rate (h\(^{-1}\)) |
|-------------|------------------------|--------------------------------------------------|--------------------------|
| PSOMOMA     | ackA, pta, ghrA\(^*\), dctA\(^*\) | 15.27                                            | 0.7967                   |
| CSMOMA [7]  | asnA, ghrA\(^*\), pykA, putP, dctA\(^*\) | 16.58                                            | 0.50898                  |
| ABCMOMA [11]| fum, zwf                | 6.69                                             | 0.44                     |

The suggested genes knockout by PSOMOMA are: [ackA, pta, fum and lpd]. The inactivation of pta-ackA genes has been proved to improve the production of succinic acid [32]. According to the authors, the removal of these genes affects the fluxes towards ethanol formation. The inactivation of these genes indirectly will affect the production of ethanol, which is encoded by adhE gene. Therefore, the mutant strain will increase the production of succinate and D-lactate. The inactivation of ghrA responsible for glycoxylate reductase will affect the metabolism of glycine and serine [33]. Meanwhile, dctA gene is required for transport of dicarboxylate [34]. The removal of these genes will reduce the competition for carbon sources, which is glucose. Moreover, PSOMOMA can find 2 similar gene knockout as CSMOMA.

Although CSMOMA found the highest production rate, however, it involves knocking out five genes compared to PSOMOMA and ABCMOMA, which only knocked out four and two genes, respectively. Furthermore, the suggested knockout genes obtained by PSOMOMA generates viable mutant with the highest growth rate. Nevertheless, the suggested knockout genes obtained by these algorithms are restricted to the computer simulation. In a wet-lab experiment, various other factors need to identify and considered, as it is difficult to apply and identify a single gene. Overall, PSOMOMA can find a set of genes knockout with the highest growth rate in *E. coli* compared to the other methods.

5 **Wet Laboratory**

In this section, the production of ethanol in *E. coli* obtained by PSOMOMA is compared with results from the wet laboratory. The results of ethanol production by PSOMOMA has been published in [35]. According to [36], three mutant strains of *E. coli* were created for maximizing the ethanol production, which are SY03, SY04 and MG1655. The results of iJO1366 are compared with MG1655 mutant strains considering that iJO1366 was constructed from this strain. Table 4 shows the ethanol production obtained from both PSOMOMA algorithm and wet laboratory test.

Table 4: Result comparison on ethanol production for PSOMOMA and Wet Laboratory Test.

| Method      | Knockouts/environment condition | Gene knockouts | Ethanol Production (mmol gDW\(^{-1}\) h\(^{-1}\)) |
|-------------|---------------------------------|----------------|-----------------------------------------------|
| PSOMOMA     | 2                               | ACKr, PPS      | 17.2029                                       |
|             | 3                               | pflA,frdB,ldhA | 17.2270                                       |
|             | 4                               | ACKr, ldhA, FUMt2, fdhF | 16.4891                                       |
|             | 5                               | ACKr, fumB, PPS, GND, GLUDy | 16.4501                                       |
| Wet Laboratory [36] | pH 7.5                       | MG1655 (pZSBlank) | 7.8400                                        |
|             | pH 7.5                          | MG1655 (pZSKLMgLdA) | 8.7000                                        |
|             | pH 6.3                          | MG1655 (pZSKLMgLdA) | 11.1400                                       |

As shown in Table 4 below, all different numbers of genes knockout in PSOMOMA result in higher ethanol production than the wet laboratory test. The highest ethanol by mutant MG1655 in wet laboratory is only 11.14 mmol gDW\(^{-1}\) h\(^{-1}\) whereas the highest production by PSOMOMA is 17.227 mmol gDW\(^{-1}\) h\(^{-1}\), which is a significant difference of 6.087 mmol gDW\(^{-1}\) h\(^{-1}\). Although PSOMOMA provides an overly optimistic result of ethanol production, however, the suggested knockout genes obtained are restricted to the computational simulation. It is advisable, thus, to test the suggested knockout genes by PSOMOMA in wet-lab experiments.
6 Conclusion

This paper focuses on a comparison of metaheuristic algorithms to solve the identification of near-optimal genes knockout to optimize the production of succinic acid. Of the three tested algorithms, PSO performs better in terms of growth rates while CS performs better in finding mutant with a higher production rate. Although CSMOMA produces the highest production rate for 5 suggested genes knockout, however, the growth rate is lesser than PSOMOMA. In future works, multiobjective optimization can be included for optimization of two competing objectives.

Acknowledgements

We would like to thank the Ministry of Education Malaysia for supporting this re-search by the Fundamental Research Grant Scheme – Malaysia’s Research Star Award -FRGS-MRSA (grant number: R/FRGS/A0800/01655A/003/2020/00720) and Fundamental Research Grant Schemes (grant number: RDU190113 and R.J130000.7828.4F720).

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal’s publication ethics and publication malpractice statement available at the journal’s website and hereby confirm that they comply with all its parts applicable to the present scientific work.

References

[1] Tomar N, De RK. Comparing methods for metabolic network analysis and an application to metabolic engineering. Gene 2013;521:1–14.
[2] Choon YW, Mohamad MS, Deris S, Illias RM, Chong CK, Chai LE, et al. Differential bees flux balance analysis with OptKnock for in silico microbial strains optimization. PLoS One 2014;9:1–13.
[3] Chen PW, Theisen MK, Liao JC. Metabolic systems modeling for cell factories improvement. Curr Opin Biotechnol 2017;46:114–9.
[4] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8.
[5] Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 2002;99:15112–7.
[6] Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux. Proc Natl Acad Sci USA 2005;102:7695–700.
[7] Anf MA, Mohamad MS, Abd Latif MS, Deris S, Remli MA, Mohd Daud K, et al. A hybrid of Cuckoo Search and Minimization of Metabolic Adjustment to optimize metabolites production in genome-scale models. Comput Biol Med 2018;102:112–9.
[8] Mutturi S. FOCuS: a metaheuristic algorithm for computing knockouts from genome-scale models for strain optimization. Mol Biosyst 2017;13:1355–63.
[9] Patil KR, Rocha I, Forster J, Nielsen J. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 2006;7:308.
[10] Rocha M, Maia P, Mendes R, Pinto JP, Ferreira EC, Nielsen J, et al. Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinformatics 2009;9:499.
[11] Tang PW, Choon YW, Mohamad MS, Deris S, Napis S. Optimising the production of succinate and lactate in Escherichia coli using a hybrid of artificial bee colony algorithm and minimisation of metabolic adjustment. J Biosci Bioeng 2015;119:363–8.
[12] Klamt S, Müller S, Regensburger G, Zhang J, Zanghellini J. A mathematical framework for yield (vs. rate) optimization in constraint-based modelling and applications in metabolic engineering. Metab Eng 2018;47:153–69.
[13] Kim TY, Park JM, Kim HY, Cho KM, Lee SY. Design of homo-organic acid producing strains using multi-objective optimization. Metab Eng 2015;28:63–73.
[14] Villaverde AF, Bongard S, Mauch K, Balsa-Canto E, Banga JR. Metabolic engineering with multi-objective optimization of kinetic models. J Biotechnol 2016;222:1–8.
[15] Patané A, Jansen G, Conca P, Carapezza G, Costanza J, Nicosia G. Multi-objective optimization of genome-scale metabolic models: the case of ethanol production. Ann Oper Res 2018;261:1–17.
[16] Mohd Daud K, Mohamad MS, Zakaria Z, Hassan R, Ali Shah Z, Deris S, et al. A non-dominated sorting Differential Search Algorithm for in silico multiobjective optimization in identifying reactions knockout. Comput Biol Med 2019;113:103390.
[17] Nagrath D, Avila-Echiver MM, Berthiaume FF, Tilless AW, Messac A, Yarmush ML, et al. Soft constraints-based multiobjective framework for flux balance analysis. Metab Eng 2010;12:429–45.
[18] Oh YG, Lee D, Yoon H, Lee SY, Park S. Multi-product trade-off analysis of E-coli by multiobjective flux balance analysis. Eur Symp Comput Process Eng 2004;18:1099–104.
[19] Costanza J, Carapezza G, Angione C, Lio P, Nicosia G. Multi-objective optimisation, sensitivity and robustness analysis in FBA modelling. In: International Conference on Computational Methods in Systems Biology. Springer Berlin Heidelberg; 2012:127–47.
[20] Piotrowski AP, Napiorkowski MJ, Napiorkowski JJ, Rowinski PM. Swarm Intelligence and Evolutionary Algorithms: Performance versus speed. Inf Sci 2017;384:34–85.

[21] Kennedy J, Eberhart R. Particle swarm optimization. In: Natural Computing Series. Springer Boston, MA, 2011:760–6.

[22] Nabaei A, Hamian M, Parsaei MR, Saifdari R, Samad-Soltani T, Zarrabi H, et al. Topologies and performance of intelligent algorithms: a comprehensive review. Artif Intell Rev 2018;49:1–25.

[23] Yang XS, Deb S. Multiobjective cuckoo search for design optimization. Comput Oper Res 2013;40:1616–24.

[24] Mahesh K, Nallagowdend P, Elamvazuthi I. Advanced pareto front non-dominated sorting multi-objective particle swarm optimization for optimal placement and sizing of distributed generation. Energies 2016;9:982.

[25] Nair G, Jungreuthmayer C, Zanghellini J. Optimal knockout strategies in genome-scale metabolic networks using particle swarm optimization. BMC Bioinformatics 2017;18:1–9.

[26] Li L, Liu F, Long G, Guo P, Bie X. Modified particle swarm optimization for BMDS interceptor resource planning. Appl Intell 2015;44:471–88.

[27] Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 2003;84:647–57.

[28] Karagoz S, Yildiz AR. A comparison of recent metaheuristic algorithms for crashworthiness optimisation of vehicle thin-walled tubes considering sheet metal forming effects. Int J Veh Des 2017;73:179.

[29] Akhmedova S, Semenkin E. Co-operation of biology related algorithms for multi-objective binary optimization. In: 2015 IIAI 4th International Congress on Advanced Applied Informatics; 2015:580–5.

[30] Chua PS, Salleh AHM, Mohamad MS, Deris S, Omatu S, Yoshioka M. 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. Biotechnol Bioprocess Eng 2015;20:349–57.

[31] Mohd Daud K, Zakaria Z, Hassan R, Mohamad MS, Shah ZA. Improved Metaheuristic algorithms for metabolic network optimization. In: IOP Conference Series: Materials Science and Engineering 2019;351:1–5.

[32] Yun NR, San KY, Bennett CN. Enhancement of lactate and succinate formation in adhE or pta-ackA mutants of NADH dehydrogenase-deficient Escherichia coli. J Appl Microbiol 2005;99:1404–12.

[33] Iida A, Olnishi Y, Horinouchi S. Identification and characterization of target genes of the GinI/GinR quorum-sensing system in Gluconacetobacter intermedius. Microbiology 2009;155:3021–32.

[34] Thakker C, Martínez I, San K-Y, Bennett CN. Succinate production in Escherichia coli. Biotechnol J 2012;7:213–24.

[35] Lee MK, Mohamad MS, Choon YW, Daud KM, Nasarudin NA, Ismail MA, et al. A Hybrid of particle swarm optimization and minimization of metabolic adjustment for ethanol production of Escherichia Coli. In: International Conference on Practical Applications of Computational Biology & Bioinformatics. 2019:36–44.

[36] Shams S, Gonzalez R. Engineering Escherichia coli for the efficient conversion of glycerol to ethanol and co-products. Metab Eng 2008;10:340–51.
