

PAPER • OPEN ACCESS

Improved Metaheuristic Algorithms for Metabolic Network Optimization

To cite this article: K Mohd Daud *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012065

View the [article online](#) for updates and enhancements.

Improved Metaheuristic Algorithms for Metabolic Network Optimization

K Mohd Daud^{1*}, Z Zakaria¹, R Hassan², M S Mohamad^{3,4} and Z Ali Shah¹

¹ Artificial Intelligence and Bioinformatics Research Group, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

² Software Engineering Department, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

³ Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa, 16100, Kota Bharu, Kelantan, Malaysia

⁴ Faculty of Bioengineering and Technology, Universiti Malaysia Kelantan, Jeli Campus, Lock Bag 100, 17600, Jeli, Kelantan, Malaysia

Email: kautharmohdaud@yahoo.com

Abstract. Metaheuristic algorithms have been used in various domains to solve the optimization problem. In metabolic engineering, the problem of identifying near-optimal reactions knockout that can optimize the production rate of desired metabolites are hindered by the complexity of the metabolic networks. Through Flux Balance Analysis, different metaheuristics algorithms have been improved to optimize the desired phenotypes. In this paper, a comparative study of four metaheuristic algorithms have been proposed. Differential Search Algorithm (DSA), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) and Genetic Algorithm (GA) are considered. These algorithms are tested on succinic acid production in *Escherichia coli*. The comparative performances are measured based on production rate, growth rate, and computational time. Hence, from the results, the best metaheuristic algorithms to solve the metabolic network optimization is identified.

1. Introduction

The metabolic network is defined as a series of biochemical reactions involving the transformation and modification of substrates into different products in which the enzymes act as catalysis agent. The metabolic network is important in assessing the properties of biochemical and physiological of a cell. Recently, advancements in genome sequencing have brought upon many developments that allow the biological researchers to have deeper knowledge and information of an organism. One of the development is the establishment of metabolic engineering, which allows the researchers to probe in detail the organizations of an organism including the reactions, pathways, metabolites, and genes as well as exploit the organisms for strains optimization by suggesting genetic alteration strategies that can optimize the desired objectives, such as growth rate or production rate.

A metabolic model may consist thousands or hundreds of reactions, genes, and metabolites that are interconnected among each other. This resulted in the complexity of the data. The complexity of metabolic model has resulted in the dimensions of solutions space being too large and thus the computational time increases exponentially [1], [2]. In addition, it is difficult to obtain a near-optimal set of reactions knockout in order to optimize the desired phenotypes. Therefore, with the principle of mass balances, various techniques and methods have been developed for modelling and optimizing



metabolic network. These methods apply constraint-based modelling and swarm-metaheuristic algorithms to identify a set of reactions that optimize the desired objective function.

Constraint-based modelling (CBM) is an approach to investigate the optimality of an organism by predicting and describing the metabolic phenotypes [3]. In CBM, constraints are applied to the systems, thus creates feasible flux distributions space. There are four approaches under constraint-based modelling methods – flux balance analysis (FBA), flux variability analysis (FVA), minimization of metabolic adjustment (MoMA), and regulatory on/off minimization (ROOM). Considering that a metabolic network comprises of vast information, FBA is able to handle large data compared to the three approaches by predicting the final higher steady-state of biological objective functions [4]. Furthermore, FBA successfully predicts the final steady-state growth rates or production rates of desired metabolites, without considering the regulatory and kinetic information that is difficult to obtain.

However, FBA alone could not optimize the capabilities of organisms by redirecting more fluxes toward the desired metabolites. Furthermore, the problem of finding combinations of reaction for knockout is considered as a combinatorial optimization problem, thus it is not practicable to test all the combinations of reaction knockout using an exhaustive brute force approach. Therefore, the use of metaheuristic algorithms have inspired the development of optimization-modelling methods for optimizing the production rate of desired metabolites by means of reaction knockout.

The advantages of metaheuristic algorithms that computationally less expensive have driven the progress of myriad strain design metaheuristic optimization-modelling methods. Although metaheuristic algorithms do not guarantee for finding optimal solutions, these algorithms allow for optimization of individual biological objective function and engineering objective function. Furthermore, metaheuristic uses multiple agents/individuals/particles to search for near-optimal solutions, thus, the search spaces are larger than the other frameworks [5]. In other words, metaheuristic allows a large number of genetic perturbations.

The focus of this paper is to evaluate and compare four metaheuristics algorithms in maximizing the production rate of succinic acid in *Escherichia coli*. These algorithms were improved with FBA as fitness function evaluation. The paper is organized as follows: Section 2 briefly describes the selected four metaheuristics algorithms. The following Section 3 presents the results and discussion. Lastly, Section 4 concludes the paper..

2. Metaheuristic Algorithms

2.1. Differential Search Algorithm

Differential Search Algorithm (DSA) is a swarm based metaheuristic algorithm that mimics the behavior of a group of animals that simulates the migration of a superorganism that consists of organisms, from unfruitful area to a more fruitful area [6]. The exploration in DSA is done by *Brownian-random walk*, while a greedy algorithm carries exploitation.

2.2. Artificial Bee Colony

Artificial Bee Colony (ABC) is inspired by the behavior of honey bee colonies in food foraging using waggle dance [7]. In ABC, the forages can be categories into two, namely employed and unemployed. For each employed bees, a new food source is generated which represent as a feasible solution. The information reside in food source is improved by the unemployed bees which is onlooker bees. On a certain limit of time, an employed bee may become a scout bee, such that the onlooker bees could not improve the food source

2.3. Particle Swarm Optimization

Particle Swarm Optimization (PSO) simulates the movement behavior of a group of animals [8]. PSO consists of a swarm of particles with different velocities and fitness. For every iteration, each particle will benefit from its own experience and global best experience of the swarm. Therefore, the

“communication” among particles are occurred and eventually, allow a unity among the particles in moving together as a swarm.

2.4. Genetic Algorithm

The earliest biological based evolutionary algorithm is a Genetic algorithm (GA) which involves exact biological mechanisms, namely mutation, crossover and selection [9]. GA uses chromosomes to exploit and explore the solution space in determining and enhancing the value of near-optimal solutions. GA is a well-known algorithm that has been improved in many ways.

Table 1 shows the advantages and disadvantages of the said algorithms.

Table 1. Summary of metaheuristic algorithms.

Name	Advantage (s)	Disadvantage (s)	Ref (s)
DSA	- The control parameters are not sensitive towards the problem. - Fastest CPU computational time. - Fewer parameters.	- Too many random numbers involved in generating stopover site. - The solutions may go outside the boundaries.	[6], [10]– [12]
ABC	- Better in global search - Fast convergence	- Require more control parameters - The solutions may trapped in local optima.	[11], [13]
PSO	- Has less parameter tuning - Efficient in global search	- Sensitivity of control parameters - The loss of inertia weight over time may cause the swarm to lose its ability to search.	[14]– [16]
GA	- Easy to implement - Fast convergence	- High computational complexity - The solutions may trapped in local optima.	[17], [18]

3. Results and Discussion

The problem in identifying reactions knockout for maximizing the production of desired metabolites can be formulated mathematically by representing the metabolic model in the stoichiometry matrix, *S* that consists of reactions and metabolites of size *m*×*n*. The stoichiometry matrix describes the dynamic mass balance equation of the metabolic model by a differential equation between the flux rates of the reactions (*v*) and concentrations of metabolites (*c*). The mass balance must be at a steady-state level that signifies there is no internal and external changes in the concentration of metabolites.

$$\frac{dx}{dt} = S \cdot v = 0 \tag{1}$$

Whereby

$$S_i = \sum_{j=1}^r n_{ij} \tag{2}$$

$$v_{lb} \leq v \leq v_{ub} \tag{3}$$

Thus, from the equations above, it can be concludes to

$$\max/\min Z = \sum_i c_i v_i \tag{4}$$

Where *S* is the stoichiometric matrix with size of *r* reactions and *m* metabolites, *n* is the coefficients of stoichiometric with the size of *i*=1, 2...*m* and *j*=1, 2...*r*, *v* is the flux vector of size reactions, and *Z* is the objective function to be optimized that associated with a biological task such as product rate or growth rate. Herein, *Z* is refer to the production rate of succinic acid.

Table 2 shows the result comparison of different methods applied in this domain. The three methods were improved with FBA for fitness function evaluation, which is product rate. *E.coli* core model is tested and the target reaction to optimized is succinic acid.

Table 2. Result comparison of different evolutionary methods.

Method	Product rate (mmol/gDW/hr)	Growth Rate (hr-1)	Time (sec)	Knockout reactions
PSOFBA	15.4905	0.4769	25.8399	ACKr, GLUDy, AKGDH, ME1, ME2, PFL, PYK, RPE
GAFBA	15.5185	0.4621	45.4629	ACKr, GLUDy, GND, LDH_D, ME1, ME2, PYK, SUCOAS
ABCFBA	15.5185	0.4621	73.2244	ACALD, G6PDH2r, AKGDH, GLUDy, GND, ME1, ME2, PYK
DSAFBA	15.50	0.4836	11.9653	ACALD, G6PDH2r, GND, ME1, ME2, PFL, PYK, SUCOAS

Note: All the methods were developed and run again using the same dataset, with number of population is 100 and maximum iteration is 100.

As shown in **Error! Reference source not found.**, DSAFBA is compared with PSOFBA, GAFBA, and ABCFBA in terms of the production rate of succinic acid, growth rate after perturbations, and time. From the result comparison, GAFBA and ABCFBA are able to find the highest amount of product rate compared to DSAFBA and PSOFBA. Nevertheless, DSAFBA is able to identify the mutant with the highest growth rate. In addition, by looking at the result of GAFBA, ABCFBA, and DSAFBA; the difference in growth rate between DSAFBA and GAFBA-ABCFBA is 0.0215, whilst the difference in product rate between DSAFBA and GAFBA-ABCFBA is 0.0185. Furthermore, DSAFBA is able to find near-optimal combinations of reactions that optimize the production of succinic acid in a short amount of time, compare to PSOFBA, GAFBA and ABCFBA.

In terms of knockout reactions identified by the methods, DSAFBA and ABCFBA are able to find reactions that were predicted by other methods as well. From the total of 8 reactions being knockout, the first 2 of the suggested reactions identified by DSAFBA are found by ABCFBA, 3 of them are found by PSOFBA, GAFBA, and ABCFBA. Meanwhile, reactions PFL and PYK are found by other methods as well, whereas reaction SUCOAS is suggested by GAFBA. Therefore, from the results shown in **Error! Reference source not found.**, the searching strategy that uses *Brownian-random walk* movement and mutation structure, DSA are able to find the combination of reactions that improve the production rate of target metabolites in a short amount of time, while maintaining the viability of the mutant.

4. Conclusions

This paper focusing on modifying metaheuristic algorithms to solve the identification of near-optimal reactions knockout. From the four tested algorithms, DSA outperforms other metaheuristic algorithms in terms of CPU computational times, and performance. The discovery of the suggested reactions knockout are limited to the computational validation only. Thus, in future, these reactions can be serve as prior knowledge for in vivo implementation.

Acknowledgements

This work was financially supported by FRGS, Vot No: 4F973 from Ministry of Education (MOHE), and also RUG, Vot No: 16H73 from Universiti Teknologi Malaysia.

References

- [1] Hansen A S L, Lennen R M, Sonnenschein N and Herrgård M J 2017 Systems biology solutions for biochemical production challenges *Curr. Opin. Biotechnol.* **45** 85–91
- [2] Wang N, Zhao W J, Wu N and Wu D 2017 Multi-objective optimization: a method for selecting the optimal solution from Pareto non-inferior solutions *Expert Syst. Appl.* **74** 96–104
- [3] Klamt S, Muller S, Regensburger G and Zanghellini J 2018 A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering

- Metab. Eng.* **47** 153–169
- [4] Bonarius H P J, Schmid G and Tramper J 1997 Flux analysis of underdetermined metabolic networks: The quest for the missing constraints *Trends Biotechnol.* **15** 308–314
 - [5] Maia P, Rocha M and Rocha I 2016 In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories *Microbiol. Mol. Biol. Rev.* **80** 45–67
 - [6] Civicioglu P 2012 Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm *Comput. Geosci.* **46** 229–247
 - [7] Karaboga D 2005 An idea based on honey bee swarm for numerical optimization *Technical Report -TR06* **46** 185–189
 - [8] Kennedy J and Eberhart R 2011 Particle Swarm Optimization *Nat. Comput. Series*, Springer Boston, MA 760–766
 - [9] Mitchell M 1995 Genetic algorithms: An overview *Complexity* **1** 31–39
 - [10] Atay Y, Koc I, Babaoglu I and Kodaz H 2017 Community detection from biological and social networks : A comparative analysis of metaheuristic algorithms *Appl. Soft Comput. J.* **50** 194–211
 - [11] Kurban T, Civicioglu P, Kurban R and Besdok E 2014 Comparison of evolutionary and swarm based computational techniques for multilevel color image thresholding *Appl. Soft Comput. J.* **23** 128–143
 - [12] Liu J, Wu C, Cao J, Wang X and Teo K L 2016 A Binary differential search algorithm for the 0 – 1 multidimensional knapsack problem *Appl. Math. Model.* **40** 9788–9805
 - [13] Waghole V and Tiwari R 2014 Optimization of needle roller bearing design using novel hybrid methods *Mech. Mach. Theory.* **72** 71–85
 - [14] Yuce B, Packianather M S, Mastrocinque E, Pham D T and Lambiase A 2013 Honey bees inspired optimization method: The bees algorithm *Insects.* **4** 646–662
 - [15] Mora-Melià D, Martínez-Solano F J, Iglesias-Rey P L and Gutiérrez-Bahamondes J H 2017 Population Size Influence on the Efficiency of Evolutionary Algorithms to Design Water Networks *Procedia Eng.* **186** 341–348
 - [16] Kachitvichyanukul V 2012 Comparison of Three Evolutionary Algorithms : GA, PSO, and DE *Ind. Eng. Manag. Syst.* **11** 215–223
 - [17] Alam M N, Das B and Pant V 2015 A comparative study of metaheuristic optimization approaches for directional overcurrent relays coordination *Electr. Power Syst. Res.* **128** 39–52
 - [18] Goldberg D E and Holland J H 1998 Genetic algorithms and machine learning *Machine Learning* **3** 95–99