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*

Pooi San Chua, Abdul Hakim Mohamed Salleh, Mohd Saberi Mohamad, Safaai Deris, Sigeru Omatu, and Michifumi Yoshioka

Abstract  The current problem for metabolic engineering is how to identify a suitable set of genes for knockout that can improve the production of certain metabolites and sustain the growth rate from the thousands of metabolic networks which are complex and combinatorial. Some approaches, such as OptKnock and OptGene, are developed to enhance the production of desired metabolites. However, the performances of these approaches are suboptimal and the obtained results are unsatisfactory because of computational limitations such as local minima. In this paper, we propose a hybrid of Bat Algorithm and Flux Balance Analysis (BATFBA) to enhance succinate and lactate production by identifying a set of genes for knock out. The Bat Algorithm is an optimisation algorithm, whereas Flux Balance Analysis (FBA) is a mathematical approach to analyse the flow of metabolites through a metabolic network. The *Escherichia coli* iJR904 dataset was used to determine optimal knockout genes, production rate, and growth rate. By applying this hybrid method to the iJR904 dataset, we found that BATFBA yielded better results than existing methods, such as OptKnock and a hybrid of Artificial Bee Colony algorithms and Flux Balance Analysis (ABCFBA), at predicting succinate and lactate production.

Keywords: bat algorithm, flux balance analysis, succinate, lactate, gene knockout strategy, *Escherichia coli*, bioinformatics, artificial intelligence, metabolic engineering

1. Introduction

Succinate and its derivatives are commonly used in the food and pharmaceutical industries, whereas lactate and its derivatives are extensively employed in industries such as food processing, leather making, and textiles. Traditionally, succinate and lactate are produced through either fermentation or chemical synthesis. However, these traditional approaches are expensive and time-consuming. Therefore, metabolic engineering has been introduced as an alternative method to improve metabolite production.

The goal of metabolic engineering is to manipulate microorganisms towards a desired phenotype. Microorganisms’ manipulation can be performed by genetic modification to achieve optimal production of the desired products. Gene knockout is a metabolic engineering technique that modifies genes to improve the production of metabolites. The advancement of metabolic engineering together with the vast collection of existing literature have enabled the reconstruction of genome-scale models of metabolic networks that contain information on network components and topology [1]. These genome-scale organismal models are used in Flux Balance Analysis (FBA) [2] for various purposes, including the optimisation of cellular objectives, such as metabolite production [3,4] and biomass formation [5]. However, the complexities of the metabolic networks
result in data ambiguity and cause the effects of genetic modification on the desirable phenotypes difficult to predict. Furthermore, the vast number of reactions in cellular metabolism results in combinatorial problem to obtain a suitable set of genes for knockout.

The first rational modelling framework for a gene knockout strategy was OptKnock [6]. OptKnock is based on a mixed linear programming method and the results aid in the overproduction of the desired metabolites. This framework was extended to OptGene [7], which used a Genetic Algorithm (GA) [8] to identify targeted genes for knock out. Unfortunately, these two approaches face the local minima problem in their multimodal functions, thereby limiting the performance [9].

The ease of genetically manipulating *Escherichia coli* (*E. coli*) has successfully aided in optimising succinate and lactate production [10]. We used *E. coli* as the model due to its high growth rate under aerobic and anaerobic conditions and its ability to undergo mixed-acid fermentation of glucose, the principal product of which produces lactate and succinate [11]. Currently, genome-scale models and some constraint-based methods such as FBA and Minimisation of Metabolic Adjustment (MOMA) [12] are used in industry for bioprocess engineering. These models and methods aid in accelerating the bioprocess development cycle, particularly within the early stage of optimal host selection to improve biochemical production [13-16]. As stated in Choon et al. (2014), a genome-scale metabolic model is usually large in size. Hence, another problem arises, because computation time increases exponentially along with the size of the model [17].

This paper proposes a hybrid of the Bat Algorithm and Flux Balance Analysis (BATFBA) to identify a set of genes to be knocked out for enhancing succinate and lactate production. The Bat Algorithm is an optimisation algorithm inspired by a microbat process. Microbats use this process to find their prey and it combines the advantages of existing swarm-intelligence algorithms. The balance between exploration and exploitation is similar to the Particle Swarm Optimisation algorithm (PSO) [18,19]. The Bat Algorithm is dynamic, simple to implement, and has the potential to solve combinatorial optimisation problems. FBA is used to evaluate the fitness of each prey and predict the growth rate of an organism and the production of a significant metabolite. FBA is computationally quick and can be used to predict which reactions should be knocked out. Unlike MOMA, FBA enables large modification in single fluxes [20]. The proposed BATFBA model utilises stoichiometric information from genome-scale metabolic models. This property allows the method to choose any metabolite of interest within the genome-scale model of an organism. The proposed method can also be used to identify a suitable set of genes to be knocked out to optimise the production of a particular metabolite in a genome-scale metabolic model.

This paper is organised as follows: In Section 2, we briefly discuss the conventional version of the Bat Algorithm, Flux Balance Analysis, and the methodology, datasets, and experimental setup of the proposed algorithm BATFBA. We explain the model pre-processing steps, parameters, hardware, and software in Section 2.2. Section 3 describes the results and analysis from the BATFBA experiments. Section 4 summarises the study through a discussion of the results and addresses possible future experiments.

## 2. Materials and Methods

### 2.1. Methods

#### 2.1.1. The bat algorithm

The Bat Algorithm is based on the echolocation behaviour of bats, which enables bats to find their prey even in complete darkness. Echolocation works by transmitting sound pulses that bounce back, allowing bats to detect the distance and direction of a target and even the type of prey [21-23]. Bats fly randomly to find their prey (or their solution) by using their pulse rates and loudness of sound. The pulse rate and loudness are changed when a new solution is better. This change indicates that the bats are approaching an optimal solution (catching the prey). The balance between exploration and exploitation can be controlled by tuning algorithm-dependent parameters in the Bat Algorithm. The quality of the solution is characterised by the volume and pulse rate, which are related to the fitness of the solution towards the global optimal solution. This property ensures that the algorithm can explore search space globally and efficiently [21]. The action of this algorithm combines the advantages of existing swarm intelligence algorithms. The balance between exploration and exploitation is similar to the PSO. The Bat Algorithm is dynamic, simple to implement, and has the potential to solve combinatorial optimisation problems.

#### 2.1.2. Flux balance analysis

FBA is a mathematical approach to analyse the flow of metabolites through a metabolic network and forecast the growth rate of an organism or the production of a targeted metabolite. FBA does not require kinetic parameters and thus can quickly execute computation even for a large network. At the same time, FBA can predict missing reactions by comparison with *in silico* growth simulations [20]. FBA is