Computational Biology and Chemistry 53 (2014) 175–183

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Identification of gene knockout strategies using a hybrid of an ant colony optimization algorithm and flux balance analysis to optimize microbial strains

Shi Jing Lu [a], Abdul Hakim Mohamed Salleh [a], Mohd Saberi Mohamad [a,*], Safaai Deris [a], Sigeru Omatu [b], Michifumi Yoshioka [c]

[a] Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[b] Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan
[c] Faculty of Engineering, Osaka Prefecture University, 1-1 Gakuencho Nakaku, Sakai, 599-8531Japan

ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form 14 September 2014
Accepted 23 September 2014
Available online 28 September 2014

Keywords:
Ant colony optimization algorithm
Flux balance analysis
Metabolic engineering
Optimization algorithm
Gene knockout strategy
Microbial strains

ABSTRACT

Reconstructions of genome-scale metabolic networks from different organisms have become popular in recent years. Metabolic engineering can simulate the reconstruction process to obtain desirable phenotypes. In previous studies, optimization algorithms have been implemented to identify the near-optimal sets of knockout genes for improving metabolite production. However, previous works contained premature convergence and the stop criteria were not clear for each case. Therefore, this study proposes an algorithm that is a hybrid of the ant colony optimization algorithm and flux balance analysis (ACOFBA) to predict near optimal sets of gene knockouts in an effort to maximize growth rates and the production of certain metabolites. Here, we present a case study that uses Baker's yeast, also known as *Saccharomyces cerevisiae*, as the model organism and target the rate of vanillin production for optimization. The results of this study are the growth rate of the model organism after gene deletion and a list of knockout genes. The ACOFBA algorithm was found to improve the yield of vanillin in terms of growth rate and production compared with the previous algorithms.

1. Introduction

Microbial strains are microorganisms that are widely used to produce biochemical products, antibiotics, drug targets, therapeutic proteins, food ingredients, vitamins, fuels, and other chemicals. In recent years, there has been increasing demand to replace chemical synthesis processes with more sustainable and environmentally friendly biotechnology methods based on microbial fermentation (Patil et al., 2005; Rocha et al., 2008). However, this means that microorganism metabolism needs to be retrofitted to comply with industrial purposes. This process is also known as genome-scale model reconstruction in general. These genome-scale models have been used to design strains that produce lactate (Fong et al., 2005), succinate (Wang et al., 2006), lycopene (Alper et al., 2005), L-threonine (Lee et al., 2007), L-valine (Park et al., 2007), and 1–3, propanediol (Wang et al., 2006). Traditionally, these genome-scale model reconstructions were made through the manipulation of genes by genetic engineering processes, such as random mutagenesis and screening. However, in recent years, this genetic engineering was coupled with metabolic engineering, which facilitated the simulation of a genome-scale model reconstruction *in silico*.

In metabolic engineering, a number of mathematical or computational tools have been developed and targeted to perform *in silico* gene manipulations that lead to the overproduction of desired compounds. There have been several successful attempts using metabolic engineering that are mainly based on qualitative and intuitive principles. However, many attempts have failed due to a lack of rational strategies that are based on predictive analysis tools that can precisely predict microorganism behavior. Nevertheless, the prediction of cellular metabolism is possible under steady state conditions, which enforces constraints throughout the reaction rates. One of the most widely used approaches is flux balance analysis (Kauffman et al., 2003), which is used to calculate the flux distribution within a metabolic network using linear programming. Several related works have also predicted and

* Corresponding author. Tel.: +60 7 553 3153; fax: +60 7 5565044.
E-mail addresses: lu_shi_jing@hotmail.com (S.J. Lu), abdhakim.utm@gmail.com (A.H.M. Salleh), saberi@utm.my (M.S. Mohamad), safaai@utm.my (S. Deris), omatu@rsh.oit.ac.jp (S. Omatu), yoshioka@cs.osakafu-u.ac.jp (M. Yoshioka).

http://dx.doi.org/10.1016/j.compbiolchem.2014.09.008
1476-9271/

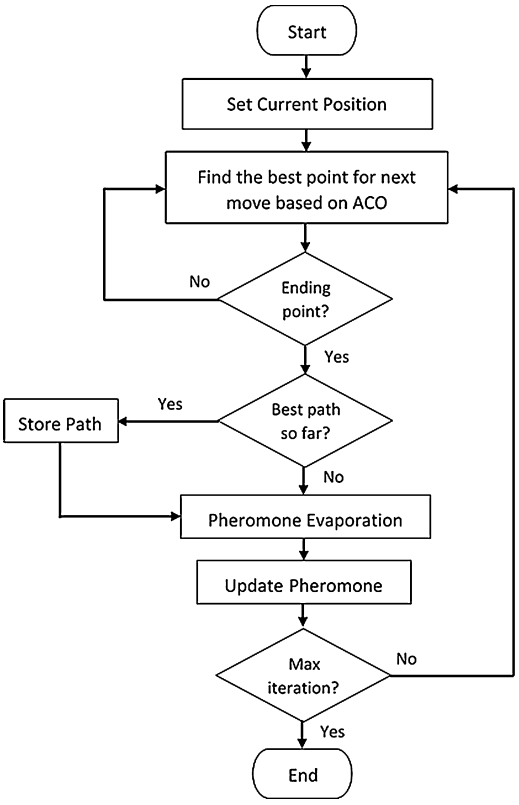

Fig. 1. A flow chart representing a conventional ACO algorithm.

analyzed the microorganism behavior, both in wild-type or mutant strains (Segre et al., 2002; Shlomi et al., 2005). All of these approaches provide a basis for analyzing microbial strain metabolic networks to predict possible phenotypes.

The first rational modelling algorithm developed to perform *in silico* gene manipulations, which was scarcely based on the gene knockout strategy to optimize microbial strains, is the OptKnock algorithm introduced by Burgard et al. 2003; this uses mixed integer linear programming (MILP) to reach the optimum solution. Several extensions to OptKnock have been described in the literature. The goal of OptStrain (Pharkya et al., 2004), which was proposed in 2005, was to perform pathway modification, through pathway additions and deletions from the microbial strain, leading to overproduction of the desired compounds. OptReg (Pharkya and Maranas, 2005) was then proposed in 2006 to perform reaction modifications by considering overexpression and down-regulation to overproduce a target metabolite in microbial strains. OptGene (Patil et al., 2005) is an extension of OptKnock, which formulated problems using the genetic algorithm (GA) to search for a global optimal solution. It is the first algorithm that has employed an evolutionary search to solve combinatorial problems in the gene knockout strategy. However, it also has the tendency to fail in local optimum solutions because it does not exhaustively search for the complete solution. Additionally, it does not show clear stop criteria for every problem and may cause premature convergence.

Since GA does not solve problems efficiently, this study proposes an algorithm that is a hybrid of the ant colony optimization algorithm and flux balance analysis (ACOFBA) to address these limitations. In addition, the *Saccharomyces cerevisiae* model was used to predict the production of vanillin as a case study in this paper. The ACOFBA algorithm was applied to predict sets of gene knockouts for biomass and production optimization. The ACO algorithm was chosen to solve the limitations of previous works because it has been proven to be efficient for solving combinatorial optimization problems (Sivanandam and Deepa, 2008; Xue et al., 2010; Yang and Zhuang, 2010; Tavares Neto and GodinhoFilho, 2011). Additionally, it also promotes the exhaustive search for a complete solution since every possible solution is evaluated, and the search process stops when a near optimal solution is found. Thus, it guarantees the generation of an optimal solution. Furthermore, the ant colony algorithm avoids premature convergence of the solution search because it performs global and local searches simultaneously. An ACO-based technique was first introduced to solve the problems associated with gene knockout strategies. An ACO-based gene knockout algorithm was developed by modification of the original ACO algorithm. This paper is organized as follows: Section 2 briefly describes the conventional version of the ACO and ACOFBA algorithms; Section 3 describes the experimental setup and results; and Section 4 summarizes this paper by providing the main conclusions and addresses future developments.

2. Methods

2.1. Conventional ant colony optimization algorithm

The ACO algorithm is a meta-heuristic optimization approach that was introduced in the early 1990s (Colorni et al., 1991; Mo et al., 2009). The search process used by the ACO algorithm was inspired by real ant intelligent behavior, in which they explore the shortest path from a food source to the nest. In the ACO algorithm, a number of artificial ants build a solution to the corresponding optimization problem and exchange the quality of these solutions by using a pheromone that allows artificial ants to communicate. The original ACO algorithm is known as the Ant System (Colorni et al., 1991) and was implemented to solve the travelling salesman problem (TSP). It performs a parallel search over constructive computational threads, based on local problem data and on a dynamic memory structure, which contain solution quality information. The goal is to minimize the total travel distance. In this method, a set of artificial ants is placed on a graph that represents a set of cities and they are forced to move. The ACO algorithm generates solutions by using a construction mechanism, in which the selection of the solution component to be added at each step is probabilistically influenced by deposited pheromones and heuristic information. This means that the construction process probabilistically builds the problem solutions step by step, and the probabilistic model has feedback for its modification based on the solutions found. The steps of the ACO algorithm were designed as proposed by Colorni et al. (1991) and are shown in Fig. 1.

2.2. Ant colony optimization for use as a gene knockout strategy

A gene knockout strategy is a combinatorial problem regarding a large solution space (a large number of reactions taking place in cellular metabolism) of a genome-scale model. In particular, the size of the problem, which is defined by the number of deletions and number of enzymes, and the corresponding search space

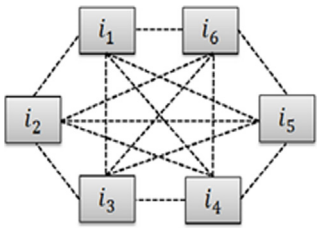

Fig. 2. ACO problem representation for individuals.