Journal of the Taiwan Institute of Chemical Engineers 54 (2015) 57–63

Contents lists available at ScienceDirect

Journal of the Taiwan Institute of Chemical Engineers

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

Optimal design of growth-coupled production strains using nested hybrid differential evolution

Feng-Sheng Wang*, Wu-Hsiung Wu

Department of Chemical Engineering, National Chung Cheng University, Chiayi 62102, Taiwan

ARTICLE INFO

Article history:
Received 10 December 2014
Revised 13 March 2015
Accepted 15 March 2015
Available online 1 April 2015

Keywords:
Growth-coupled production strain
Hybrid differential evolution
Bi-level optimization problem
Flux balance analysis
Flux variability analysis

ABSTRACT

Various traditional optimization approaches have been applied to identify optimal manipulation strategies for metabolic networks of microorganism leading to maximization of desired products. However, because of the transient effects of traditional strategies on production rate, the design of growth-coupled production strains is essential for metabolic engineering. Most current approaches for optimal strain design problems apply a two-stage procedure to identify a growth-coupled strain. This study reformulated the optimal strain design problem as a decision making problem with a guarantee of identifying growth-coupled production strains, and a nested hybrid differential evolution (HDE) algorithm that combined the two-stage procedure into one stage was introduced to solve this problem. The performance of the proposed algorithm was demonstrated by using it to design several growth-coupled production strains for a genome-scale metabolic model of *Escherichia coli* iAF1260. The nested HDE was able to control the magnitude of the association between cell growth rate and chemical production rate and can outperform state-of-the-art algorithms for the design of growth-coupled strains.

1. Introduction

The goal of metabolic engineering is to obtain optimal manipulation strategies for metabolic networks of bacteria leading to maximization of desired products [1]. Various traditional strategies have been employed to achieve this goal, including elimination of competing pathways, over expression of genes to increase fluxes of production pathways, and activation of inactive pathways to maximize the production rate of target compounds at the expense of growth rate [2,3]. Experiments based on traditional strategies have observed transient effects on production rate, i.e., mutants always increase their growth rate and decrease production rate after adaptive evolution [4]. A promising and innovative method to solve this problem is the design of growth-coupled production strains that increase the production of target compounds as they evolve to higher growth rates.

Mathematical modeling is a crucial tool for the design of growth-coupled production strains; modeling and analysis typically implement dynamic and static approaches [5–7]. Dynamic approaches with kinetic models use prior knowledge to make specific molecular predictions and work most effectively with pathways where components and connectivity are relatively firmly established [5,6]. Static approaches, such as constraint-based reconstruction and analysis (COBRA), consist of stoichiometric relationships of metabolic networks that can be applied to the study of genome-scale metabolism [7,8]. Both approaches have also been used to investigate the effects of environmental and genetic perturbations on the metabolic capabilities of an organism [9], and numerous algorithms have been developed to solve constraint-based problems [10–18]. Flux balance analysis (FBA) is the most popular approach for determining optimal flux distributions of wild-type microorganisms and for identifying the optimal metabolic state of mutant strains by using linear programming (LP) [8]. FBA predicts cellular behavior based on the assumption of maximized biomass growth rate, which can predict various metabolic phenotypes [11]. By contrast, the minimization of metabolic adjustment (MOMA) identifies optimal flux solutions of mutants based on the minimization of Euclidean norm of flux differences between wild type and gene knockout strains using quadratic programming (QP) [12]. Regulatory on/off minimization (ROOM) is another constraint-based flux analysis technique that minimizes the number of significant flux changes by using mixed integer linear programming (MILP) [13].

Recently, the growth-coupled strain design problem was formulated as a bi-level optimization problem (BLOP) that consists of a bioengineering optimization problem with integer variables and a cellular optimization problem. The BLOP is a special case of multiobjective optimization problems and numerous solutions for solving it have been developed, including OptKnock [14], OptStrain [15], OptReg [16], OptForce [17], OptORF [18], EMILiO [19], and ReacKnock

* Corresponding author at: Faculty of Chemical Engineering, National Chung Cheng University, No.168, Sec.1, University Rd., Min-Hsiung Township, Chia-yi County 621, Taiwan (R.O.C.). Tel.: +886 5 2720 411x33404; fax: +886 5 2721 206.
E-mail address: ccuchmfsw@gmail.com, chmfsw@ccu.edu.tw (F.-S. Wang).

http://dx.doi.org/10.1016/j.jtice.2015.03.015
1876-1070/

[20]. These methods transfer the BLOP into a single-level mixed-integer linear programming (MILP) problem by applying duality theory or Karush–Kuhn–Tucker method to the optimization problem in the inner level. However, such a duality transformation can increase computational time exponentially when the problem dimension increases. Although the OptKnock algorithm requires a long CPU time (up to one week) to predict a five-reaction knockout design using the *E. coli* iAF1260 model, it was the first constraint-based method used to predict strain designs for various substrates and products [21]. To avoid non-uniquely growth-coupled strains, RobustKnock extended the OptKnock method to yield guaranteed production rates by accounting for the presence of competing pathways in the network model iJR904 [2]. Evolutionary algorithms have been applied to identify the modulated genes of strain-design problems with a user-defined objective function in which complicated nonlinear objective functions can be used [21–27].

Most methods solving the BLOPs apply a two-stage process to identify a growth-coupled strain. A candidate set of target reactions is obtained by solving the BLOP in the first stage. However, the solution of BLOP may not be a growth-coupled production strain. A posterior decision-making procedure is conducted to identify a growth-coupled strain from the candidate set in the second stage. This study merged the two-stage process into a one-stage multilevel optimization problem supporting decision making. Two optimization problems for flux variability analysis (FVA) in the inner level of the multilevel optimization problem guarantee that the optimal solution is a growth-coupled strain. We had developed a hybrid differential evolution (HDE) algorithm previously for static and dynamic optimization of fedbatch fermentation processes [28]. This algorithm has been applied to many optimization problems, such as parameter estimation [29], chemical plant design [30], and modeling of biological systems [31]. In this study, the HDE algorithm was extended to a nested procedure to solve the genome-scale multilevel optimization problem for the determination of optimal growth-coupled production strains through the introduction of fuzzy membership functions.

2. Problem formulation and methods

2.1. Cellular flux analysis

The FBA problem is an *in silico* flux-based optimization problem for the prediction of metabolic flux distributions in genome-scale metabolic networks. Such optimization problem includes a cellular objective function v_{cellular} (*e.g.*, the maximization of cell growth rate) governing cellular behavior, and can be expressed as follows:

$$\begin{cases} \text{FBA problem:} \\ \max\limits_{\mathbf{v}} v_{\text{cellular}} \\ \text{subject to} \\ \quad \mathbf{Nv} = \mathbf{0} \\ \quad v_i = 0, i \in \Omega_{\text{KO}} \\ \quad v_j^{\text{LB}} \le v_j \le v_j^{\text{UB}}, j \notin \Omega_{\text{KO}} \end{cases} \tag{1}$$

where $\mathbf{v}$ is an n-dimensional vector of intracellular fluxes, $\mathbf{N}$ is an $m \times n$ stoichiometric matrix where m is the number of metabolites and n is the number of reactions, Ω_{KO} is the set of knockout reactions, and v_j^{LB} and v_j^{UB} are the lower and upper bounds of flux of reaction j that do not belong to Ω_{KO}, respectively. The flux value was set to zero if the corresponding reaction was selected to be knocked out. Although FBA assessed the theoretical fluxes solely based on the information regarding the stoichiometry of metabolic networks and fundamental physicochemical constraints in the absence of kinetic information on the dynamics and regulation of metabolic reactions, it can incorporate additional available information by adding equations that impose bounds on the fluxes as the inequality constraints in Eq. (1) [8]. Some known systemic constraints on the fluxes of exchange reactions under different growth conditions have been validated experimentally [32] and were included in the genome-scale model of *E. coli*, iAF1260. By imposing these systemic capacity constraints on exchange reactions, the flux distribution of whole metabolic network obtained by FBA based on the maximization of cellular growth and the study of feasible metabolic behavior within these systemic constraints will be more sensible.

The FBA problem can have multiple solutions with a same cellular optimal objective value. Flux variability analysis determines the flux range of each reaction based on the maximal cellular objective by solving one minimization optimization problem and one maximization optimization problem [33], and is formulated as follows:

$$\begin{cases} \text{FVA problem:} \\ \max\limits_{\mathbf{v}} / \min\limits_{\mathbf{v}} v_{\text{bioeng}} \\ \text{subject to} \\ \quad \mathbf{Nv} = \mathbf{0} \\ \quad v_{\text{cellular}} \ge v_{\text{cellular}}^{\text{FBA}} \\ \quad v_i = 0, i \in \Omega_{\text{KO}} \\ \quad v_j^{\text{LB}} \le v_j \le v_j^{\text{UB}}, j \notin \Omega_{\text{KO}} \end{cases} \tag{2}$$

where v_{bioeng} is the bioengineering objective and $v_{\text{cellular}}^{\text{FBA}}$ is the optimal solution to the FBA problem. FVA was used not only to determine alternative optima in the FBA problem, but also to determine if a mutant is a unique growth-coupled strain. This study embedded FVA as a constraint in the optimal strain design problem to determine a growth-coupled production strain.

2.2. Optimal strain design problem

The design of growth-coupled production strains intends to identify mutants employed the smallest number of knockout reactions to achieve certain specifications (*e.g.*, maximal product yield or maximal substrate specific productivity) [21], and can be formulated as a multiobjective decision-making problem as follows:

$$\begin{cases} \max\limits_{\mathbf{z}} f_1 \equiv v_{\text{bioeng}}^{\text{minFVA}} \\ \max\limits_{\mathbf{z}} f_2 \equiv v_{\text{cellular}}^{\text{FBA}} \\ \min\limits_{\mathbf{z}} f_3 \equiv \sum\limits_j z_j \\ \text{subject to} \\ \quad \text{obj}\left(\begin{array}{l} \min\limits_{\mathbf{v}} v_{\text{bioeng}} \\ \text{subject to} \\ \quad \mathbf{Nv} = \mathbf{0} \\ \quad v_{\text{cellular}} \ge v_{\text{cellular}}^{\text{FBA}} \\ \quad v_i = 0, i \in \Omega_{\text{KO}} \\ \quad v_j^{\text{LB}} \le v_j \le v_j^{\text{UB}}, \\ \quad j \notin \Omega_{\text{KO}} \end{array} \right) = \text{obj}\left(\begin{array}{l} \max\limits_{\mathbf{v}} v_{\text{bioeng}} \\ \text{subject to} \\ \quad \mathbf{Nv} = \mathbf{0} \\ \quad v_{\text{cellular}} \ge v_{\text{cellular}}^{\text{FBA}} \\ \quad v_i = 0, i \in \Omega_{\text{KO}} \\ \quad v_j^{\text{LB}} \le v_j \le v_j^{\text{UB}}, \\ \quad j \notin \Omega_{\text{KO}} \end{array} \right) \\ \quad \begin{cases} \max\limits_{\mathbf{v}} v_{\text{cellular}} \\ \text{subject to} \\ \quad \mathbf{Nv} = \mathbf{0} \\ \quad v_i = 0, i \in \Omega_{\text{KO}} \\ \quad v_j^{\text{LB}} \le v_j \le v_j^{\text{UB}}, j \notin \Omega_{\text{KO}} \end{cases} \end{cases} \tag{3}$$

where $\text{obj}(\min v_{\text{bioeng}}, \ldots) = v_{\text{bioeng}}^{\text{minFVA}}$ and $\text{obj}(\max v_{\text{bioeng}}, \ldots) = v_{\text{bioeng}}^{\text{maxFVA}}$ stand for the optimal objective values of the minimization and maximization optimization problems, respectively, and $\mathbf{z}$ is a binary vector indicating the deletion status of reactions. The outer level consists of three objective functions to make a determination toward a satisfactory strain and one equality constraint,