Journal of Bioscience and Bioengineering
VOL. 115 No. 2, 221–228, 2013

www.elsevier.com/locate/jbiosc

In silico screening of triple reaction knockout *Escherichia coli* strains for overproduction of useful metabolites

Satoshi Ohno,[1] Chikara Furusawa,[1,2] and Hiroshi Shimizu[1,*]

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan[1] *and Quantitative Biology Center (QBiC), RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan*[2]

Received 2 July 2012; accepted 5 September 2012
Available online 3 October 2012

For efficient production of industrially useful metabolites by microorganisms, it is important to design metabolic networks suitable for production. For this purpose, constraint-based metabolic flux simulation is a powerful tool to predict the effect of reaction knockouts on target productivity. In this study, using constraint-based metabolic simulation, *in silico* screening was performed to identify knockout candidate sets that increase the productivity of 1-butanol, 1-propanol, and 1,3-propanediol by engineered *Escherichia coli*. Metabolic changes caused by all possible sets of triple reaction knockouts were evaluated using a reduced metabolic model, which was constructed to significantly reduce the computational cost. The results demonstrated the strategies to improve target productivity by gene disruption; some of them could not be achieved by previous screening methods. Such knockout strategies can support further improvements of the target productivity.

[**Key words:** Flux balance analysis; *Escherichia coli*; 1-Butanol; 1-Propanol; 1,3-Propanediol]

Metabolic engineering has significantly facilitated the development of novel microbial strains that can be used for the production of useful metabolites. At present, an increasing number of chemical products, including pharmaceutical chemicals, commodity chemicals, fine chemicals, and diesels, are being produced using microbial strains as cell factories (1–5). To improve the productivity of these processes, metabolic networks within microbial cells have been modified through alteration of gene expression levels or through introduction of heterologous genes using molecular biology tools. However, identification of gene manipulations that result in a desired phenotype is still difficult because of the complexity of the intra-cellular metabolic network. Therefore, the development of an appropriate *in silico* platform to provide possible gene manipulations and appropriate flux distributions in order to improve target productivity is desired to facilitate the rational design of a metabolic network for industrial purposes (6,7).

Recently, based on the explosion of annotated genome sequence data, it has become possible to reconstruct genome-scale metabolic models (GSMs) of microbial strains *in silico*, which can be applied to constraint-based metabolic flux balance analysis (FBA) (8–10). A significant advantage of the constraint-based modeling compared with other method as kinetic modeling (11) is that it requires a smaller number of parameters. In the constraint-based modeling, metabolic fluxes can be quantitatively estimated without using kinetics parameters of metabolic reactions, by assuming a steady state of metabolic system and optimization of an objective function as biomass production. The maximization of biomass production flux has been widely used as the objective function, and it was demonstrated that the metabolic fluxes estimated by maximizing biomass production agreed well with the experimentally obtained ones (12,13). This agreement can be explained by growth maximization of microorganisms through evolutionary dynamics. Furthermore, for the mutant strains constructed in the laboratory, it was shown that cells can achieve the nearly optimal metabolic state calculated by FBA after long-term cultivation (14,15), which is also explained by the selection of faster growing cells. Thus, we can expect that constraint-based metabolic simulation will provide quantitative predictions of the flux changes caused by genetic modifications, which will be applicable for *in silico* screening of possible gene manipulations to improve the productivity of target products. This *in silico* screening of gene manipulations can drastically reduce the experimental costs for strain improvements; indeed, it has been applied for the production of several targets (16–18).

In this study, using constraint-based metabolic simulations, we performed *in silico* screening of effective gene knockout sets by using *Escherichia coli* as the host microorganism for the overproduction of three useful metabolites: 1-butanol, 1-propanol, and 1,3-propanediol (13PD). 1-Butanol is expected to be one of the important biofuels, since its energy content (27 MJ/L) is similar to that of gasoline (32 MJ/L). 1-Butanol can be replaced with gasoline in an existing gasoline engine. The vapor pressure of 1-butanol (4 mmHg at 20°C) is lower than that of ethanol (45 mmHg at 20°C).

* Corresponding author. Tel./fax: +81 6 6879 7446.
E-mail address: shimizu@ist.osaka-u.ac.jp (H. Shimizu).

1389-1723/$ – see front matter
http://dx.doi.org/10.1016/j.jbiosc.2012.09.004

As such, the production of 1-butanol by microorganisms has attracted much attention (19). 1-Propanol is generally used as a solvent, food additive, or chemical intermediate such as *n*-propylamine (NCBI, 1-propanol − PubChem). 1-Propanol converged from biomass can be used for sustainable fuel in a similar way to 1-butanol. 13PD is a building block of polymers such as polytrimethylene terephthalate. 13PD has been produced commercially via a bioconversion process by DuPont Tate & Lyle Bioproducts since 2007.

For screening of effective gene knockouts, we added the heterologous metabolic pathways to produce the above metabolites in to a GSM of *E. coli*. Then we estimated the change of metabolic fluxes caused by all possible combinations of triple reaction knockouts by assuming the maximization of biomass production to obtain the sets of triple reaction knockouts for increased production yields of the above metabolites. Although similar *in silico* screenings of multiple knockout candidates have been reported (17,20), these studies relied on the iterative screening of knockout candidates. For example, in the first iteration, the effects of all possible single reaction knockouts were evaluated, and the reactions whose knockout resulted in the highest score (e.g., highest production yields) were selected. In the second iteration, in addition to the reaction knockouts selected in the first iteration, the effects of additional single reaction knockouts were screened. Although such iterative screening can reduce the computational costs, and screening of multiple knockouts by this method is relatively easy, this approach does not identify the combination of reaction knockouts that show a desired phenotype only when they are simultaneously disrupted. In another study, a bi-level programming framework for screening knockout targets, namely OptKnock, was used to obtain knockout candidates (21). This method utilizes a nested optimization framework to identify the combination of knockout targets that optimize the production yield of target metabolites. The program for this algorithm is open for public use and has been extended to predict other metabolic strategies such as gene amplifications or reaction additions (22,23). However, a disadvantage of these algorithms is that the output is only one set of knockouts for optimal productivity, although other knockout targets with sub-optimal productivities can provide valuable information on the strategy of strain improvement. To overcome these problems, in this study, we first constructed a reduced metabolic model that can provide flux estimations that are completely identical to the original GSM and can significantly decrease the computational cost. Then, using this reduced metabolic model, we calculated the effect of all possible sets of triple reaction knockouts on the productivities of target metabolites. This comprehensive screening provided optimal and sub-optimal knockout combinations for strain improvement aimed at the production of 1-butanol, 1-propanol, and 13PD, which included non-intuitive mechanisms for overproduction.

MATERIALS AND METHODS

Genome-scale metabolic model of *E. coli* As an original metabolic model, we used a GSM of *E. coli* K-12 MG1655 named iAF1260 (24), which contains 1260 ORFs from the latest genome annotation and over 2000 transport and intracellular reactions. To evaluate the production yields of target metabolites, we added the production pathways of targets into the original GSM as presented in the Results section.

Since the original GSM contained a large number of metabolic reactions, the calculation of all possible combinations of several reaction knockouts was difficult because of the computational cost. To overcome this problem, a reduced metabolic model was constructed by the following procedure. First, using FBA with maximum biomass production and flux variability analysis (25), we identified metabolic reactions whose maximum and minimum fluxes were estimated to be zero in the environmental conditions used by us; these reactions were removed since the deletion of these reactions has no effect on the flux profile. Second, adjacent metabolic reactions without branching were combined into a single reaction, since the knockout of adjacent reactions results in an identical effect on the flux changes. Finally, the following two sets of reactions were combined into single reactions. First, the three reactions of NADH dehydrogenase (NADH dehydrogenase of ubiquinone, menaquinone, and demethylmenaquinone) were merged into a single reaction because these reactions are encoded by genes in the same *nuo* operon. Second, the two reactions of aldehyde dehydrogenase and alcohol dehydrogenase, which catalyze the conversion of acetyl-CoA to acetaldehyde, and then to ethanol, were combined into a single reaction, because these reactions were primarily activated by the same gene *adhE*. After these adjustments, the number of reactions in the reduced metabolic model of *E. coli* decreased to about 1000 reactions (Table S1); the original model contained more than 2000 reactions. We confirmed that the changes in metabolic flux caused by single reaction knockouts were identical in the original and the reduced model.

In the present study, we used the names of representative genes in the combined reactions to represent the knockout candidate reactions. For example, phosphate acetyltransferase, which catalyzes the reaction between acetyl-CoA and acetyl phosphate encoded by the *pta* and *eutD* genes, was represented by the *pta* gene. The list of genes and reactions are presented in Table S2.

Flux balance analysis The constraint-based FBA was performed following previous literature procedures (12,13,26). Briefly, a pseudo-steady state of the metabolic profile was assumed, i.e., the net sum of all production and consumption metabolic fluxes for each internal metabolite was set to zero. This assumption resulted in a feasible space that was a convex set in the *N*-dimensional space of metabolic fluxes (where *N* stands for the total number of fluxes). In FBA, a particular objective function, written as a linear combination of fluxes, can be used to calculate the optimal solution at one corner in the feasible flux space. Using the matrix notation, this problem can be stated as follows:

$$\begin{aligned} &\text{maximize}: \ \mathbf{c}^{\mathrm{T}} \cdot \mathbf{v} \\ &\text{subject to}: \ \mathbf{S} \cdot \mathbf{v} = 0 \\ &\qquad \mathbf{v}_{\min} \le \mathbf{v} \le \mathbf{v}_{\max} \end{aligned}$$

where **S** is the stoichiometric matrix representing the stoichiometry of metabolic reactions in the network and **v** is the vector of all metabolic fluxes. $\mathbf{v}_{\min}$ and $\mathbf{v}_{\max}$ indicate the minimum and maximum constraints on the fluxes, respectively, and were used to define the constraints for irreversibility of the reaction, constant uptake from the environment, and reaction knockouts; **c** is the vector of weights indicating how much each reaction contributes to the objective function. In this study, we used the maximization of biomass production flux as the objective function. After obtaining the maximal biomass production flux, we further maximized the production flux of target metabolites under the fixed biomass production flux on the maximal value to avoid undetermined production fluxes.

For all simulations, glucose or glycerol was chosen as the sole carbon source, and the substrate uptake rate was set to 10 or 20 mmol/gDW/h for glucose or glycerol, respectively, where gDW was the dry weight of the biomass in grams. The oxygen uptake rate was set to 5 mmol/gDW/h, which corresponds to a micro-aerobic condition.

We chose the low oxygen uptake rate because of high production yields of target metabolites in this condition, in comparison with a higher oxygen uptake rate in which carbon was used to produce biomass. In facts, we confirmed that the target production rates of mutant strains under the oxygen uptake rate of 5 mmol/gDW/h were always higher than those under the oxygen uptake rate of 10 mmol/gDW/h. Other external metabolites such as CO_2 and NH_3 were allowed to freely transport through the cell membrane, following a previous study (24). All calculations, including the linear programming problems, were performed using the commercially available software GLPK (GNU Linear Programming Kit) and Matlab (Mathworks Inc., Natick, MA, USA).

RESULTS AND DISCUSSION

Screening of all possible triple reaction knockout strains for overproduction of target metabolites To identify knockout targets for overproduction of 1-butanol, 1-propanol, and 13PD by *E. coli*, we evaluated metabolic changes caused by all possible combinations of triple reaction knockouts after adding heterologous pathways for target production. Two carbon sources (glucose and glycerol) were adopted for the production of target metabolites. We used the reduced metabolic model as presented in Materials and methods, which can significantly reduce the computational time for the knockout screening. For the all possible triple reaction knockout, more than 1 year is necessary when we used the original metabolic model iAF1260, while it takes 4.6 min using our reduced model (2.93 GHz Intel Xeon 12 core CPU). Throughout this study, the maximization of biomass production was used as the objective function of the FBA simulations.