

# Trimming Gene Deletion Strategies for Growth-Coupled Production in Constraint-Based Metabolic Networks: TrimGdel

Takeyuki Tamura 

**Abstract**—When simulating genome-scale metabolite production using constraint-based metabolic networks, it is often necessary to find gene deletion strategies which lead to growth-coupled production, which means that target metabolites are produced when cell growth is maximized. Existing methods are effective when the number of gene deletions is relatively small, but when the number of required gene deletions exceeds approximately 1% of whole genes, the time required for the calculation is often unfeasible. Therefore, a complementing algorithm that is effective even when the required number of gene deletions is approximately 1% to 5% of whole genes would be helpful because the number of deletable genes in a strain is increasing with advances in genetic engineering technology. In this study, the author developed an algorithm, TrimGdel, which first computes a strategy with many gene deletions that results in growth-coupled production and then gradually reduces the number of gene deletions while ensuring the original production rate and growth rate. The results of the computer experiments showed that TrimGdel can calculate stoichiometrically feasible gene deletion strategies, especially those whose sizes are 1 to 5% of whole genes, which lead to growth-coupled production of many target metabolites, which include useful vitamins such as biotin and pantothenate, for which existing methods could not.

**Index Terms**—Biology and genetics, chemistry, combinatorial algorithms, graphs and network, linear programming



## 1 INTRODUCTION

COMPUTATIONAL strain design plays an important role in the production of substances through microbial metabolism [1], [2], [3], [4], [5], [6], [7], [8]. Such strain designs are often evaluated through simulations with constraint-based models, which can perform genome-scale simulations efficiently by restricting the analysis to steady states [9]. In constraint-based models, (1) the sum of incoming reaction rates is equal to the sum of outgoing reaction rates for each metabolite, (2) the ratio of the produced and consumed metabolites in each reaction satisfies the ratio in the chemical reaction equation, and (3) each reaction rate is given upper and lower bounds.

Each constraint-based model includes a virtual reaction that represents cell growth. The cell growth reaction was designed to match the results of the biological experiments. The most standard simulation in constraint-based models maximizes the cell growth rate (GR) because strains with high growth rates are more likely to remain in the culture during passaging. In contrast, the reaction that produces the desired metabolite is called a production reaction, and its production rate is denoted as PR. Therefore, designed strains are evaluated by PR

at GR maximization in the simulations. The simultaneous occurrence of cell growth and target metabolite production is called *growth-coupled production* (See Fig. 1(A) and (B)). In this paper, we consider it growth-coupled production when the value of PR is 0.001 mmol/gDW/h or more at GR maximization.

Since only a limited number of metabolites are produced with growth coupling in the natural state of microorganisms, it is necessary to calculate gene deletion strategies for each target metabolite in such a case. However, the problem of calculating a gene deletion strategy that results in growth-coupled production is a computationally heavy task for genome-scale models, and many methods have been proposed to address this problem [10], [11], [12], [13], [14], [15], [16], [17].

One of the current best ideas for calculating gene deletion strategies for growth-coupled production is to find a minimal set of reactions, which is called a (constrained) elementary flux mode, where cell growth forces target metabolite production and then delete all reactions that are not included in that set. This idea was extended to calculate the minimal cut set (MCS), which is equivalent to gene deletion strategies because the MCS of the primary network is the elementary mode of the dual network [18], [19]. Examples of successful applications of the MCS-based method include the growth-coupled production of itaconic acid [20] and 2,3-butanediol [21] by *Escherichia coli* and indigoidine by *Pseudomonas putida* [22].

Because a small number of reaction deletions is desirable in current metabolic engineering technology, the number of reaction deletions required can be reduced by

- The author is with Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 6110011, Japan.  
E-mail: tamura@kuicr.kyoto-u.ac.jp.

Manuscript received 15 Nov. 2021; revised 27 Apr. 2022; accepted 19 June 2022. Date of publication 22 June 2022; date of current version 3 Apr. 2023. This work was supported in part by JSPS, KAKENHI under Grant #20H04242.

Digital Object Identifier no. 10.1109/TCBB.2022.3185221

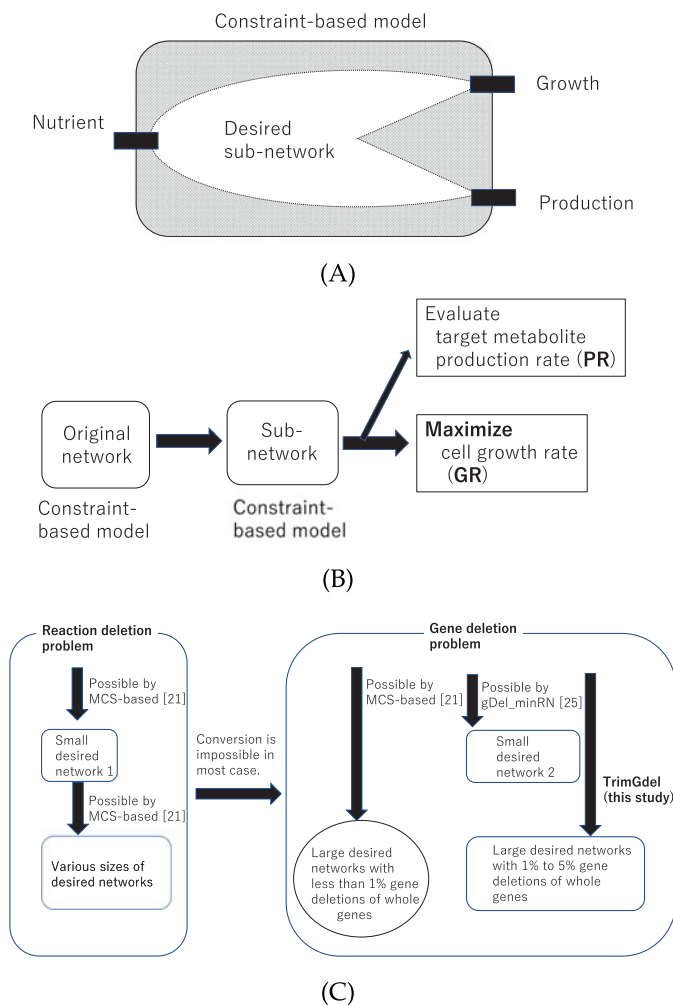


Fig. 1. (A) The main objective of this study is to calculate gene deletion strategies that simultaneously lead to cell growth and target metabolite production. (B) Each gene deletion strategy is evaluated by the least value of the target metabolite production rate (PR) when maximizing the cell growth rate (GR). (C) When compared to the calculation of reaction deletion strategies that consider the gene-protein-reaction (GPR) network is more complex. TrimGdel, developed in this study, enables the calculation of gene deletion strategies with a size of 1% to 5% of whole genes, which is difficult to calculate using existing methods for many target metabolites.

combining the resulting minimal flows to create larger flows. On the other hand, it is often impossible to compute a feasible gene deletion strategy from the obtained reaction deletion strategies because the gene-protein-reaction (GPR) networks described by Boolean functions need to be considered [23], [24].

Recently, Tamura *et al.* developed gDel\_minRN, which determines gene deletion strategies for the growth-coupled production of several vitamins in which the number of suppressed reactions is maximized [25]. Gene deletion strategies obtained by gDel\_minRN are suitable for biological analysis of the resulting metabolic flow because it leads to growth-coupled production with as few reactions as possible. However, it is not suitable for current metabolic engineering technology because gDel\_minRN deletes too many genes.

Therefore, it would be desirable to develop a method to result in growth-coupled production with a small number of gene deletions, using the large-scale gene deletion strategy obtained in gDel\_minRN as a starting point. However, when the flows calculated by gDel\_minRN are combined, there is no corresponding feasible gene deletion strategy in most cases because of the conflict caused by GPR relations.

In this study, the author developed TrimGdel, which trims large-scale gene deletion strategies that lead to growth-coupled production while ensuring the original GR and PR. TrimGdel consists of three steps: (1) Step 1 employs gDel\_minRN to obtain the large-scale gene deletion strategies that lead to growth-coupled production of target metabolites. (2) Step 2 minimizes the number of gene deletions so that the repressed reactions do not become unrepressed. (3) Step 3 reduces the number of genes to be deleted one by one under the condition that GR and PR do not fall below the original values when GR is maximized.

To evaluate the performance of TrimGdel and compare with other methods, TrimGdel, GDLS [10] and optGene [26] were applied to iML1515 [27], iMM904 [28], and e\_coli\_core [29] in computational experiments: GDLS and optGene are ones of the most widely used software to derive gene deletion strategies and are available in COBRA Toolbox [30]; iML1515 and iMM904 are genome-scale constraint-based models of *E. coli* and *S. cerevisiae*, respectively; e\_coli\_core contains the only essential part of metabolism of *E. coli*.

When TrimGdel was applied to target metabolites in iML1515, iMM904, and e\_coli\_core, large-scale gene deletion strategies for growth-coupled production could be derived by gDel\_minRN [25] for 39.9%, 13.9%, and 84.6% of the target metabolites, respectively. For each obtained large-scale gene deletion strategy, trimming was conducted to reduce the number of gene deletions. On average, TrimGdel succeeded in reducing 86.1%, 91.1%, and 81.6% of the number of gene deletions obtained by gDel\_minRN while ensuring the original GR and PR. The size of the gene deletion strategies obtained by TrimGdel was less than 5% of whole genes in 86.9%, 91.7%, and 59.1%, and between 1% and 5% in 66.5%, 50.5%, and 56.8% of the gene deletion strategies. This includes important substances such as biotin and pantothenate, for which design methods for efficient production strains have not yet been established. Therefore, we conclude that TrimGdel effectively calculates gene deletion strategies with sizes between 1% and 5% of whole genes, which are difficult to calculate with existing methods but promising for designing production strains in the near future.

The remaining of this paper is organized as follows (See Fig. 1(C)): Section 2.1 describes the main problem of this study mathematically; Section 2.2 describes the developed algorithm TrimGdel mathematically; Section 2.3 illustrates the main problem and the application of TrimGdel to toy examples; Section 3.1 describes comprehensive computational experiments where TrimGdel was applied to iML1515, iMM904, and e\_coli\_core; Section 3.2 compares the performance of TrimGdel with

GDLS and optGene; Section 3.3 describes the changes in metabolic flows when biotin is the target metabolite, and Section 4 analyzes the results of the computer experiments, evaluates the performance of TrimGdel and other methods, and discusses future work. The developed software is available online.<sup>1</sup>

## 2 METHODS

### 2.1 Definition

Let  $C_1 = (M, R, S, L, U)$  be a *constraint-based metabolic network*.  $M = \{m_1, \dots, m_a\}$  and  $R = \{r_1, \dots, r_b\}$  are sets of metabolites and reactions, respectively.  $S$  is a stoichiometry matrix, where  $S_{ij} = k$  means that  $r_j$  produces  $k$  of  $m_i$  per unit time. If  $k$  is a negative number, then  $m_i$  is consumed. Let  $V = \{v_1, \dots, v_b\}$  be a set of reaction rates per unit time (flux) of  $R$ . Let  $L = \{l_1, \dots, l_b\}$  and  $U = \{u_1, \dots, u_b\}$  be the sets of the lower and upper bounds of the reaction rates (speed) for  $V$ , respectively.  $R$  always includes one special virtual reaction  $r_{growth}$  that represents cell growth, and the cell growth flux is represented by  $v_{growth}$ , which is called the *growth rate (GR)*. The rate of the reaction producing the target metabolite under the condition that cell growth is maximized is called the *production rate (PR)*. Let  $r_{production}$  be the target metabolite production and  $v_{production}$  be its rate, that is,  $v_{production} = PR$ .

For the gene deletion simulations,  $C_2 = (G, F, P)$ , which is the *GPR network part* of the constraint-based model, is also considered.  $G = \{g_1, \dots, g_c\}$  and  $F = \{f_1, \dots, f_b\}$  are sets of genes and Boolean functions, respectively.  $P = \{p_1, \dots, p_b\}$  is the set of the outputs of  $F$ . If  $p_j = 0$ , then  $l_j$  and  $u_j$  are forced to be zero and  $r_j$  is repressed. The constraint-based model  $C$  is defined as  $C = (C_1, C_2) = (M, R, S, L, U, G, F, P)$ .

In the standard setting of growth coupling simulation using FBA with  $C_1$  and  $C_2$ , PR is evaluated when GR is maximized by the following linear programming (LP):

**maximize**

$$v_{growth}$$

**suchthat**

$$\sum_j S_{ij} v_j = 0 \text{ for all } i$$

$$p_j = f_j(G) \text{ for all } j$$

$$\begin{cases} v_j = 0 & \text{when } p_j = 0, \\ l_j \leq v_j \leq u_j & \text{when } p_j = 1 \end{cases}$$

$$i = \{1, \dots, a\}, j = \{1, \dots, b\}$$

$$p, g \in \{0, 1\}$$

(evaluate the minimum PR =  $v_{production}$ )

Because LP does not always have a unique solution, the minimum PR at GR maximization is evaluated in this study.

### 2.2 Algorithm

TrimGdel aims to find the smallest gene deletion strategy for growth-coupled production by removing unnecessary gene deletions from the large-scale gene deletion strategy obtained by gDel\_minRN for a given target metabolite and constraint-based model. TrimGdel consists of three steps:

- 1) Step 1 employs gDel\_minRN to obtain a large-scale gene deletions strategy, derives which reactions are repressed, and determines the initial GR and PR. Initial GR and PR values will be the minimum requirements that TrimGdel must satisfy.
- 2) Step 2 minimizes the number of deleted genes while maintaining which reactions are repressed.
- 3) Step 3 trims unnecessary deleted genes while ensuring original GR and PR at the maximization of GR.

In the following, TrimGdel is mathematically explained while an explanation with a small toy example can be found in Section 2.3.

*Step 1:* Let  $C$  and  $m_{target}$  be the given constraint-based model and target metabolite. If  $C$  has a corresponding exchange reaction  $r$  to produce  $m_{target}$ ,  $r$  is treated as the target metabolite production reaction  $r_{target}$ . If not, TrimGdel adds a virtual exchange reaction  $r_{target}$  to produce  $m_{target}$  to evaluate gene deletion strategies. For  $C$  and  $r_{target}$ , TrimGdel employs gDel\_minRN to obtain a large-scale gene deletion strategy  $D_1$ , which results in the growth-coupled production of  $m_{target}$ . If gDel\_minRN cannot find  $D_1$ , TrimGdel stops and returns no solution.

According to  $D_1$  and  $p_j = f_j(G)$ , which represents the GPR network for  $r_j$ , TrimGdel determines the reactions that are repressed. Let  $P_0$  and  $P_1$  be the sets of repressed and non-repressed reactions, respectively. If  $p_j = 1$ , the original upper and lower bounds of the reaction rate are applied,  $l_j \leq v_j \leq u_j$ . Because the solution of linear programming (LP) may not be uniquely determined, there may be multiple PR values at the time of GR maximization. The values of GR (=  $v_{growth}$ ) and the minimum PR (=  $v_{target}$ ) at the maximization of GR are set to  $GR_{threshold}$  and  $PR_{threshold}$ , respectively, using the flux variability analysis (FVA)-based calculation. It is to be noted that FVA can be easily performed by two LPs.

*Step 2:* TrimGdel minimizes the number of deleted genes while maintaining  $P_0$  and  $P_1$ . Because the set of repressed reactions does not change, GR and PR at the maximization of GR =  $v_{growth}$  do not change either.  $D_2$ , which is the set of genes assigned to 0, is the set of deleted genes at this time point.

*Step 3:* TrimGdel picks up one deleted gene  $g_i$  in the ascending order of  $i$ . For the provisional set of deleted genes  $D_2 - g_i$ , TrimGdel maximizes  $v_{growth}$  and checks whether  $v_{growth} \geq GR_{threshold}$  and  $v_{production} \geq PR_{threshold}$  are satisfied. If the conditions are satisfied,  $g_i$  is deleted from  $D_2$  and proceeds to the next deleted gene. If  $i$  reaches  $n$ , it goes back to 1. If no deleted gene can be trimmed for  $1 \leq i \leq n$ , TrimGdel stops and returns  $D_2$ .

1. <https://github.com/MetNetComp/TrimGdel>

The pseudo-code of TrimGdel is as follows.

```

Procedure TrimGdel( $C, m_{target}$ )
/* Step1 */
Run gDel.minRN[25] to obtain  $D_1$  for given  $C$  and  $m_{target}$ .
maximize
 $v_{growth}$ 
such that
 $\sum_j S_{ij} v_j = 0$  for all  $i$ 
 $\begin{cases} v_j = 0 & \text{if } p_j = 0 \\ l_j \leq v_j \leq u_j, & \text{otherwise} \end{cases}$ 
 $p_j = f_j(G)$ 
 $\begin{cases} g = 0, & \text{if } g \in D_1 \\ g = 1, & \text{otherwise} \end{cases}$ 
 $GR_{threshold} = v_{growth}, PR_{threshold} = \min v_{target}$ 
 $P_0 = \{p_j | p_j = 0\}, P_1 = \{p_j | p_j = 1\}$ 
 $p, g \in \{0, 1\}$ 
 $i = \{1, \dots, a\}, j = \{1, \dots, b\}$ 
/* Step2 */
maximizes
 $\sum_{g_k}$ 
such that
 $\sum_j S_{ij} v_j = 0$  for all  $i$ 
 $\begin{cases} v_j = 0 & \text{if } p_j = 0 \\ l_j \leq v_j \leq u_j, & \text{otherwise} \end{cases}$ 
 $\begin{cases} p_j = 1, & \text{if } p_j \in P_1 \\ p_j = 0, & \text{if } p_j \in P_0 \end{cases}$ 
 $p_j = f_j(G)$ 
 $p, g \in \{0, 1\}$ 
 $i = \{1, \dots, a\}, j = \{1, \dots, b\}$ 
 $D_2 = \{g_k | g_k = 0\}$ 
/* Step3 */
do
 $flag = 0$ 
for  $i = 1$  to  $|G|$ 
if  $g_i \in D_2$ 
 $D_2 = D_2 - g_i$ 
maximize
 $v_{growth}$ 
such that
 $\sum_j S_{ij} v_j = 0$  for all  $i$ 
 $\begin{cases} v_j = 0 & \text{if } p_j = 0 \\ l_j \leq v_j \leq u_j, & \text{otherwise} \end{cases}$ 
 $p_j = f_j(G)$ 
 $\begin{cases} g = 0, & \text{if } g \in D_2 \\ g = 1, & \text{otherwise} \end{cases}$ 
if  $v_{growth} \geq GR_{threshold}$ 
and  $(\min v_{production}) \geq PR_{threshold}$ 
 $flag = 1$ 
else
 $D_2 = D_2 \cup g_i$ 
end
end
end
while  $flag = 1$ 
return  $D_2$ 

```

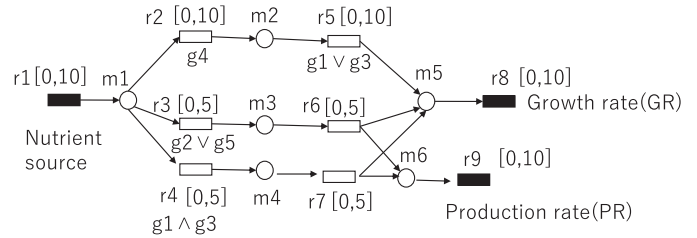


Fig. 2. Section 2.3 illustrates the main problem of this study and the idea and behavior of the developed algorithm TrimGdel using this toy example of the constraint-based model.

## 2.3 Example

A toy example of the constraint-based model  $C = (C_1, C_2) = (M, R, S, L, U, G, F, P)$  is shown in Fig. 2 and explained as follows.

**Metabolic Network Part.**  $M = \{m_1, \dots, m_6\}$  and  $R = \{r_1, \dots, r_9\}$  are sets of metabolites and reactions, respectively. The target metabolite is  $m_6$ . The growth and target metabolite production reactions are  $r_8$  and  $r_9$ , respectively. The stoichiometry matrix is

$$S = \begin{pmatrix} 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix}.$$

$[x, y]$  denotes the lower and upper bounds for each reaction rate, that is,  $L = (0, 0, 0, 0, 0, 0, 0, 0, 0)$  and  $U = (10, 10, 5, 5, 10, 5, 5, 10, 10)$ .

**GPR Network Part.** The set of genes, Boolean functions, and outputs of the Boolean functions are  $G = \{g_1, \dots, g_5\}$ ,  $F = \{f_1, \dots, f_9\}$ , and  $P = \{p_1, \dots, p_9\}$ , respectively, where

$$\begin{aligned} f_1 &: p_1 = 1, \\ f_2 &: p_2 = g_4, \\ f_3 &: p_3 = g_2 \vee g_5, \\ f_4 &: p_4 = g_1 \wedge g_3, \\ f_5 &: p_5 = g_1 \vee g_3, \\ f_6 &: p_6 = 1, \\ f_7 &: p_7 = 1, \\ f_8 &: p_8 = 1, \\ f_9 &: p_9 = 1. \end{aligned}$$

For example, because the GPR relation for  $r_4$  is given as  $f_4 : p_4 = g_1 \wedge g_3$ , the reaction rate of  $r_4$ , denoted as  $v_4$ , is forced to be 0 if one of  $g_1$  and  $g_3$  is 0, while  $0 \leq v_4 \leq 5$  is held if both  $g_1$  and  $g_3$  are 1. However,  $v_5$  is forced to be 0 if both  $g_1$  and  $g_3$  are 0, while  $0 \leq v_5 \leq 10$  is held if at least one of  $g_1$  and  $g_3$  is 1. Since  $p_1 = p_6 = p_7 = p_8 = p_9 = 1$  always holds, none of  $v_1, v_6, v_7, v_8, v_9$  are forced to be 0 by genes.

**Behavior of the Constraint-Based Model.** In the original state, the maximum value of GR is  $v_8 = 10$ . However, there are three paths to reach from the nutrient reaction  $r_1$  to the growth reaction  $r_8$ ; that is,  $(r_1, r_2, r_5, r_8)$ ,  $(r_1, r_3, r_6, r_8)$ ,  $(r_1, r_4, r_7, r_8)$ .

TABLE 1

Flux (Reaction Rate) Distributions Corresponding to the Optimistic and Pessimistic Production Rate (PR) At Growth Rate (GR) Maximization for Gene Deletion Strategies

ID	Gene KO		$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$
1	none	best	10	0	5	5	0	5	5	10	10
		worst	10	10	0	0	10	0	0	10	0
3	g1	best	10	5	5	-	5	5	0	10	5
		worst	10	10	0	-	10	0	0	10	0
5	g4	best	10	-	5	5	0	5	5	10	10
		worst	10	-	5	5	0	5	5	10	10
7	g1,g3,g4,g5	best	10	-	5	-	-	5	0	5	5
		worst	10	-	5	-	-	5	0	5	5
9	g1,g3,g4	best	10	-	5	-	-	5	0	5	5
		worst	10	-	5	-	-	5	0	5	5
11	g3,g4	best	10	-	5	-	-	5	0	5	5
		worst	10	-	5	-	-	5	0	5	5

If the first path is not used, GR=10 and PR=10 are obtained, as shown in ID 1 of Table 1. This is the most optimistic case regarding the value of PR. However, if only the first path is used, GR=10 and PR=0 are obtained as shown in ID 2 of Table 1. This is the most pessimistic case regarding the value of PR. In this study, we evaluate the most pessimistic value of PR when GR is maximized. Therefore, GR=10 and PR=0 are obtained in the original state. The behaviors of this constraint-based model are described for the cases in which the gene deletion strategies are  $\phi$ ,  $\{g_1\}$ ,  $\{g_4\}$ ,  $\{g_1, g_3, g_4, g_5\}$ ,  $\{g_1, g_3, g_4\}$ , and  $\{g_3, g_4\}$  in Table 1. The other cases are omitted because they are not used in the following explanation. IDs 6, 8, 10, and 12 represent the cases that result in the growth-coupled production of  $m_6$ . (GR,PR)=(10,10) is held for ID 6 and (GR,PR)=(5,5) is held for IDs 8, 10, and 12.

*Behavior of TrimGdel.* First TrimGdel employs gDel\_minRN [25]. Suppose that gDel\_minRN outputs a large-scale gene deletion strategy  $\{g_1, g_3, g_4, g_5\}$ . Because GR=PR=5 is obtained,  $GR_{threshold}$  and  $PR_{threshold}$  are set to 5. Therefore, TrimGdel must find a smaller gene deletion strategy that satisfies  $GR \geq 5$  and  $PR \geq 5$ .

*Step 1:* When  $\{g_1, g_3, g_4, g_5\}$  is deleted,

$$f_2 : p_2 = g_4 = 0,$$

$$f_3 : p_3 = g_2 \vee g_5 = 1,$$

$$f_4 : p_4 = g_1 \wedge g_3 = 0,$$

$$f_5 : p_5 = g_1 \vee g_3 = 0,$$

are obtained since  $(g_1, g_2, g_3, g_4, g_5) = (0, 1, 0, 0, 0)$ . Since  $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9) = (1, 0, 1, 0, 0, 1, 1, 1, 1)$  holds,  $\{r_2, r_4, r_5\}$  is repressed and then  $v_2 = v_4 = v_5 = 0$  holds.  $P_0$  is  $\{p_2, p_4, p_5\}$  and  $P_1$  is  $\{p_1, p_3, p_6, p_7, p_8, p_9\}$ .

*Step 2:* The number of deleted genes is minimized while  $P_0$  and  $P_1$  are maintained using integer linear programming. Since trimming  $g_5$  does not affect  $P_0$  and  $P_1$ ,  $(g_1, g_2, g_3, g_4, g_5) = (0, 1, 0, 0, 1)$  is obtained. It is to be noted that GR and PR remain at 5.

*Step 3:* TrimGdel checks the effect of trimming  $g_1$  from  $\{g_1, g_3, g_4\}$ . When the gene deletion strategy is  $\{g_3, g_4\}$ , GR=PR=5 is still obtained. Next, the effect of trimming  $\{g_3\}$  is

TABLE 2

The Constraint-Based Models That Were Used in the Computational Experiments

Model	#genes	#reactions	#metabolites
iML1515	1515	2712	1877
iMM904	904	1577	1226
e_coli_core	137	95	72

checked. When the gene deletion strategy is  $\{g_4\}$ , GR=PR=10 is obtained. It is to be noted that GR and PR are allowed to be larger than  $GR_{threshold}$  and  $PR_{threshold}$ , respectively. However, when  $g_4$  is trimmed from the gene deletion strategy, GR=10 and PR=0 are obtained since no gene is deleted. Then, TrimGdel stops and outputs  $\{g_4\}$ .

### 3 COMPUTATIONAL EXPERIMENTS

First, to analyze the performance of the developed method, TrimGdel was applied to iML1515, iMM904, and e\_coli\_core. The results are described in Section 3.1 The number of genes, reactions, metabolites included in each model is described in Table 2.

Second, to compare the performance, GDLS and optGene were applied to iML1515, iMM904, and e\_coli\_core with the same problem setting as TrimGdel. Comparative analysis was conducted in Section 3.2.

All procedures in the computational experiments were implemented on a CentOS 7 machine with an AMD Ryzen Processor with 2.90 GHz 64 cores/128 threads, 128 GB memory, and 1 TB SSD. This workstation had CPLEX 12.10, COBRA Toolbox v3.0 and MATLAB R2021a installed and used for these analyses.

An auxiliary exchange reaction was temporarily added to the model to simulate target metabolite production if the target metabolite did not have a production reaction.

#### 3.1 Comprehensive Experiments

*For iML1515.*

The model contains 1877 metabolites, but by solving a linear programming problem that maximizes PR, we can confirm that 785 of these metabolites cannot satisfy  $PR \geq 0.001$ , which means that it is theoretically impossible to obtain the gene deletion strategies for growth-coupled production. In other words, for the remaining 1092 metabolites, there may be a gene deletion strategy that leads to growth-coupled production. See also Table 3(A).

Because gDel\_minRN succeeded in calculating large-scale gene deletion strategies for 436 of these metabolites, trimming was conducted to these 436 large-scale gene deletion strategies. Table 3(B) shows the number of deleted genes after Step 1 (*Case 1*), after Step 2 (*Case 2*), and after Step 3 (*Case 3*). The average number of deleted genes in the gene deletion strategies for the 436 target metabolites was 963.74 for Case 1, 898.75 for Case 2, and 134.81 for Case 3. On the other hand, the maximum number of deleted genes was 976 in Case 1, 905 in Case 2, and 903 in Case 3, while the minimum number of deleted genes was 945 for Case 1, 881 for Case 2, and 1 for Case 3.

Table 3(C) shows the distribution of the number of deleted genes for Case 3. The number of gene deletions was 53 or less

**TABLE 3**  
The Performance of TrimGdel for iML1515

#metabolites in iML1515	1877
#metabolites with TMPR $\geq$ 0.001	1092
#cases where gene deletion strategies were determined by TrimGdel	436

(A)

	Case 1 after Step 1	Case 2 after Step 2	Case 3 after Step 3
avg. gDel	963.74	898.75	134.81
max gDel	976	905	903
min gDel	945	881	1

(B)

#gene deletions	#target metabolites
0 ~ 7 ( $\leq$ 0.5%)	79
8 ~ 15 ( $\leq$ 1%)	10
16 ~ 22 ( $\leq$ 1.5%)	164
23 ~ 30 ( $\leq$ 2%)	82
31 ~ 37 ( $\leq$ 2.5%)	34
38 ~ 45 ( $\leq$ 3%)	9
46 ~ 53 ( $\leq$ 3.5%)	1
54 ~ 886 ( $\leq$ 58.5%)	0
887 ~ 909 ( $\leq$ 60%)	57
total	436

(C)

	Case 1 after Step 1	Case 2 after Step 2	Case 3 after Step 3	elapsed time
Biotin	978	890	32	18m54s
Pantothenate	954	890	34	17m33s
Riboflavin	952	890	3	16m13s

(D)

(A) The number of target metabolites is 1092, for which the theoretical maximum production rate (TMPR) is 0.001 or more. (B) The Average, maximum, and minimum numbers of gene deletions for cases 1, 2, and 3 for the 436 successful cases. (C) Distribution of the number of gene deletions for the 436 target metabolites after step 3 of trimGdel was applied.

for 86.9% of the target metabolites. For the remaining 13.1% of the target metabolites, the number of gene deletions ranged from 887 to 909. There were no target metabolites for which the number of gene deletions was between 54 and 886. The distribution of the number of gene deletions was divided into those 53 or less and 887 and more. In particular, the number of gene deletions was between 1% and 3.5% of whole genes for 66.5% of the target metabolites.

Among the gene deletion strategies obtained by TrimGdel, those for the three vitamins shown in Table 3(D) represent typical cases. For biotin, the number of deleted genes was 978, 890, and 32 in Cases 1, 2, and 3, respectively. The total elapsed time was 18m54 s. Similarly, for pantothenate, the number of deleted genes was 954, 890, and 34 in Cases 1, 2, and 3, respectively. The total elapsed time was 17m33 s. For biotin and pantothenate, the final number of deleted genes was around 30 for both. However, for riboflavin, the number of deleted genes was 952, 890, and 3 in Cases 1, 2, and 3, respectively. The total elapsed time was 16m13 s.

For iMM904.

The same framework of experiments as iML1515 was applied to iMM904. The results are summarized in Table 4.

As shown in Table 4(A), gDel\_minRN succeeded in calculating large-scale gene deletion strategies for 109 of the 782 target metabolites. The average number of deleted genes for the 109 target metabolites was 613.39 for Case 1, 534.60 for Case 2, and 54.39 for Case 3. The maximum number of deleted genes was 630 in Case 1 and 547 in Cases 2 and 3, while the minimum number of deleted genes was 591 for Case 1, 518 for Case 2, and 1 for Case 3. The number of gene deletions was 36 or less for 91.7% of the target

**TABLE 4**  
The Performance of TrimGdel for iMM904

#metabolites in iMM904	1226
#metabolites with TMPR $\geq$ 0.001	782
#cases where gene deletion strategies were determined by TrimGdel	109

(A)

	Case 1 after Step 1	Case 2 after Step 2	Case 3 after Step 3
avg. gDel	613.39	534.60	54.39
max gDel	630	547	547
min gDel	591	518	1

(B)

#gene deletions	#target metabolites
0 ~ 4 ( $\leq$ 0.5%)	30
5 ~ 9 ( $\leq$ 1%)	15
10 ~ 13 ( $\leq$ 1.5%)	24
14 ~ 18 ( $\leq$ 2%)	16
19 ~ 22 ( $\leq$ 2.5%)	4
23 ~ 27 ( $\leq$ 3%)	5
28 ~ 31 ( $\leq$ 3.5%)	5
32 ~ 36 ( $\leq$ 4%)	1
37 ~ 529 ( $\leq$ 58.5%)	0
530 ~ 547 ( $\leq$ 61%)	9
total	109

(C)

(A) The number of target metabolites is 782 in iMM904, for which the theoretical maximum production rate (TMPR) is 0.001 or more. (B) The average, maximum, and minimum numbers of gene deletions for cases 1, 2, and 3 for the 109 successful cases. (C) Distribution of the number of gene deletions for the 109 target metabolites after step 3 of trimGdel was applied.

metabolites. For the remaining 8.3% of the target metabolites, the number of gene deletions ranged from 530 to 547. There were no target metabolites for which the number of gene deletions was between 37 and 529. The distribution of the number of gene deletions was divided into those 36 or less and 530 or more. The number of gene deletions was between 1% and 4% of whole genes for 50.5% of the target metabolites.

For *e\_Coli\_Core*.

The same framework of experiments as iML1515 and iMM904 was applied to *e\_coli\_core* as well. The results are summarized in Table 5.

As shown in Table 5(A), gDel\_minRN succeeded in calculating large-scale gene deletion strategies for 44 of the 52 target metabolites. The average number of deleted genes for the 44 target metabolites was 67.41 for Case 1, 48.95 for Case 2, and 12.39 for Case 3. The maximum number of deleted genes was 71 in Case 1, 55 in Case 2, and 49 in Case 3, while the minimum number of deleted genes was 64 for Case 1, 44 for Case 2, and 1 for Case 3. The number of gene deletions was 13 or less for 84.1% of the target metabolites. For the remaining 15.9% of the target metabolites, the number of gene deletions ranged from 45 to 49. There were no target metabolites for which the number of gene deletions was between 14 and 44. The distribution of the number of gene deletions was divided into those 13 or less and 45 or more. The number of gene deletions was between 1% and 5% of whole genes for 56.8% of the target metabolites.

### 3.2 Comparison With Other Methods

GDLS and optGene were applied to iML1515, iMM904, and *e\_coli\_core* to compare the performance with TrimGdel. In the GDLS experiments, the minimum GR was set to 0.001. The time limit was set to 20 m for all methods. The results are summarized in Table 6.

TABLE 5  
The Performance of TrimGdel for E\_coli\_core

#metabolites in e_coli_core	72
#metabolites with TMPR $\geq$ 0.001	52
#cases where gene deletion strategies were determined by TrimGdel	44

(A)

	Case 1 after Step 1	Case 2 after Step 2	Case 3 after Step 3
avg. gDel	67.41	48.95	12.39
max gDel	71	55	49
min gDel	64	44	1

(B)

#gene deletions	#target metabolites
0 ~ 1 ( $\leq$ 1%)	1
2 ~ 6 ( $\leq$ 5%)	25
7 ~ 13 ( $\leq$ 10%)	11
14 ~ 44 ( $\leq$ 32%)	0
45 ~ 49 ( $\leq$ 36%)	7
total	44

(C)

(A) The number of target metabolites is 52 in e\_coli\_core, for which the theoretical maximum production rate (TMPR) is 0.001 or more. (B) The average, maximum, and minimum numbers of gene deletions for cases 1, 2, and 3 for the 44 successful cases. (C) Distribution of the number of gene deletions for the 44 target metabolites after step 3 of trimGdel was applied.

For iML1515, the success ratio of GDLS and optGene were 0.2% and 0.64%, respectively. In the GDLS strategies, the maximum PR was 0.001 or more for all 1092 target metabolites. However, GR and the minimum PR at GR maximization were zero for 1078 and 1085 target metabolites, respectively. Therefore, only two gene deletion strategies resulted in growth-coupled production when the worst PR was evaluated at GR maximization. In the optGene strategies, the maximum GR was 0.001 or more for 900 cases. However, the worst PR at GR maximization was 0,001 or more only for 12 target metabolites. The average numbers of gene deletions in the successful cases were 1.342 and 4.556 for GDLS and optGene, respectively. The average elapsed times of TrimGdel, GDLS, and optGene for the successful cases were 17m42 s, 38 s, and 20m4s, respectively.

For iMM904, the success ratio of GDLS and optGene were 0.12% and 4.6%, respectively. In the GDLS strategies, the maximum GR and PR at GR maximization were 0.001 or more for 125 and 670 target metabolites, respectively. However, the minimum PR was 0.001 or more only for one target metabolite. Therefore, only one gene deletion strategy resulted in growth-coupled production when the worst PR was evaluated at GR maximization. In the optGene strategies, the maximum GR was 0.001 or more for 664 cases. However, the worst PR at GR maximization was 0,001 or more only for 36 target metabolites. The average numbers of gene deletions in the successful cases were 1 and 6.69 for GDLS and optGene, respectively. The average elapsed times of TrimGdel, GDLS, and optGene for the successful cases were 2m34 s, 0.81 s, and 20m19 s, respectively.

For e\_coli\_core, the success ratio of GDLS and optGene were 13.5% and 48.1%, respectively. In the GDLS strategies, GR, the minimum PR, and the maximum PR at GR maximization were 0.001 or more for 52, 9, and 17 target metabolites, respectively. In the optGene strategies, the maximum GR was 0.001 or more for 46 cases. The worst PR at GR maximization was 0,001 or more for 39 target metabolites. The average numbers of gene deletions in the successful cases were 3.22 and

TABLE 6  
Performance Comparison of TrimGdel, GDLS, and Optgene for (A)iML1515, (B)iMM904, and (C)e\_coli\_core

	TrimGdel	GDLS	optGene
Success ratio	39.9%	0.2%	0.64%
Avg. deletion size	134.81	1.342	4.556
Avg. elapsed time	17m42s	38s	20m4s

(A) iML1515

	TrimGdel	GDLS	optGene
Success ratio	13.9%	0.12%	4.6%
Avg. deletion size	54.39	1	6.69
Avg. elapsed time	2m34s	0.81s	20m19s

(B) iMM904

	TrimGdel	GDLS	optGene
Success ratio	84.6%	13.5%	48.1%
Avg. deletion size	12.39	3.22	7.12
Avg. elapsed time	1.305s	0.797s	20m5s

(C) e\_coli\_core

Each gene deletion strategy was considered as successful when the minimum GR and PR were 0.001 or more at GR maximization.

7.12 for GDLS and optGene, respectively. The average elapsed times of TrimGdel, GDLS, and optGene for the successful cases were 1.305 s, 0.797 s, and 20m5s, respectively.

### 3.3 Gene Deletion Strategies for Biotin Production

The gene deletion strategies for biotin in iML1515 were analyzed to compare metabolic flows for different cases. The larger gene deletion strategy after Step 1 deleted 978 genes, while 32 genes were deleted after Step 3 for growth-coupled biotin production.

Fig. 3 shows an overview of the area around the biotin production reaction in the main flux flows obtained in the computational experiments. There is almost a single path from L-homocysteine to biotin production. There are two paths to reach l-homocysteine: from malonyl CoA and from l-cysteine.

Table 7 shows the reaction rates of the growth reaction, biotin synthase, Malonyl-CoA methyltransferase (MAL-COAMT), 0-succinylhomoserine lyase (SHSL1) and ATP synthase for the natural state (Case 0), Cases 1 and 3. It should be noted that the reaction rates are the same for Cases 1 and 2. The sums of the absolute values of all reactions are also shown.

In Case 0, the growth reaction rate was 0.8770, but the biotin synthase reaction rate was  $1.7 \times 10^{-6}$ , which is lower than the required lower bound 0.001. After deleting 978 genes determined by gDel\_minRN (Case 1), GR decreased to 0.1493, but the biotin synthase rate increased to 0.1313. In Case 3, TrimGdel trimmed 946 deleted genes, and the strategy with 32 gene deletions increased both GR and biotin synthase to 0.5272 and 0.2629, respectively. The MAL-COAMT rates were the same as the biotin synthase rates in all cases. The SHSL1 rates may be linked to the biotin synthase rates since its value increases in the application of Steps 1 and 2 and the subsequent application of Step 3. The ATP synthase rate was almost the same in Case 0 and Case3, but it decreased only in Case 1. The sum of the absolute values of all reaction rates was the largest in Case 0, decreased significantly in Case 1, and took an intermediate value in Case 3.

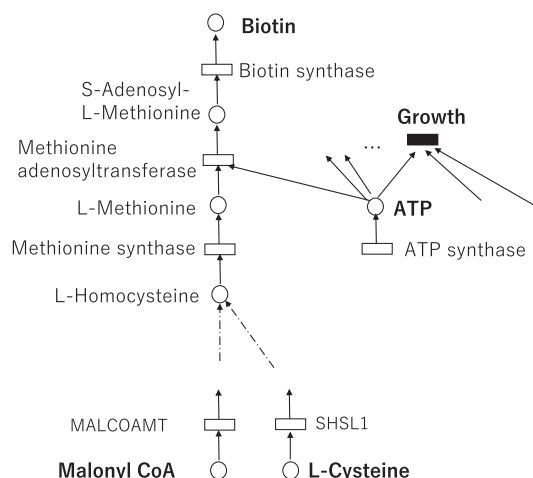


Fig. 3. An overview of the pathways around biotin in iML1515.

## 4 DISCUSSION

### Comprehensive Experiments.

As shown in Tables 3(A), 4(A), and 5(A) for 41.8% (785 = 1877-1092), 36.2% (444 = 1226-782), 27.8% (20 = 72-52) of the metabolites in iML1515, iMM904, e\_coli\_core, there is no gene deletion strategy for growth-coupled production, respectively. Therefore, it is reasonable to use the remaining 58.2% (1092), 63.8% (782), 72.2% (52) as the target metabolites to evaluate the performance of TrimGdel and other methods.

TrimGdel could calculate large-scale gene deletion strategies in iML1515, iMM904, e\_coli\_core for 39.9% (436), 13.9% (109), and 84.6% (44) of the target metabolites, respectively. However, these values depend on the performance of gDel\_minRN because TrimGdel employs gDel\_minRN in Step 1. If a method cannot obtain a large-scale gene deletion strategy, Steps 2 and 3 cannot be applied. Therefore, improving the performance of gDel\_minRN is a necessary future work for the effective utilization of TrimGdel.

As shown in Table 3(B), the average number of deleted genes in Cases 2 and 3 was 93.3% (898.75) and 14.0% (134.81) of Case 1 (963.74), respectively, for iML1515. In other words, the number of gene deletions could only be reduced by 6.74% on average when whether each reaction was suppressed or not was maintained. On the other hand, when the constraint was only that GR and PR were not reduced, the number of gene deletions was reduced by 86.0% compared to Case 1. These trimming rates were 12.9% and 91.1% for iMM904, and 27.4% and 81.6% for e\_coli\_core as shown in Tables 4(B) and 5(B). In all three models, Step 3 trimmed much more genes than Step 2.

In Cases 1 and 2, the differences between the maximum and minimum numbers of gene deletions for iML1515 were small, 31 and 24, respectively; however, in Case 3, it was 902. These values were 39, 29, and 546 for iMM904, and 7, 11, and 48 for e\_coli\_core. This may be due to the fact that the number of genes that can be trimmed does not differ greatly until Case 2 but varies greatly in Case 3.

Table 3(C) summarizes the distribution of the number of gene deletions in Case 3 for iML1515. 86.9% were 53 or less, 13.1% were 887 or more, and 0% were between 54 and 886.

The number of gene deletions could be reduced to 53 or less except for the 57 target metabolites for which the number of gene deletions was 887 or more. Although this number may be related to the distribution of the number of gene deletions in Case 2, the reason is unclear and remains to be clarified in future studies. As shown in Tables 4(C) and 5(C), similar discussion can be made for iMM904 and e\_coli\_core. 91.7% were 36 or less for iMM904, and 84.1% were 13 or less for e\_coli\_core.

The number of gene deletions was between 1% to 5% of whole genes for 66.5%, 50.5%, and 56.8% of the TrimGdel methods for iML1515, iMM904, and e\_coli\_core, respectively. Because existing methods focus on gene deletion strategies with sizes less than 1% of whole genes, TrimGdel can be considered an effective complement. However, the number of gene deletions obtained by TrimGdel is a local optimum, for which a smaller gene deletion strategy may exist. In particular, Step 3 of TrimGdel is very simple, and further improvement may lead to a smaller gene deletion strategy.

As shown in Table 3(D), biotin and pantothenate need around 30 gene deletions, while riboflavin needs only three gene deletions in the TrimGdel method for iML1515. It should be noted that although there are many known production strain strategies for riboflavin [31], there is no established design strategy for biotin or pantothenate. The elapsed time was 18m54 s, 17m33ss, and 16m13ss, respectively. Considering the number of combinations in the solution space, gene deletion strategies with a size of 1% or less of whole genes were already computable by existing methods. However, TrimGdel made it possible for the first time to compute gene deletion strategies for many target metabolites that require a gene deletion number greater than 1% of whole genes for genome-scale models.

Steps 1 and 2 of TrimGdel are exponential time algorithms since they solve mixed-integer linear programming, which is known to be NP-complete [32]. Step 3 is solvable in polynomial time since it employs linear programming at most  $O(c^2)$  times and linear programming is solvable in polynomial time [32], where  $c$  is the number of genes.

To evaluate the applicability of TrimGdel on larger-scale models, TrimGdel was applied to Recon3D [33], which is a three-dimensional genome-scale model of human. Recon3D has 2248 genes, 5835 metabolites, and 10600 reactions. As gDel\_minRN could derive gene deletion strategies for only 20 target metabolites, TrimGdel was also applied to the 20 target metabolites. The average number of deleted genes was 2131.9 for Cases 1, 1208.4 for Case 2, and 187.3 for Case 3. The average number of deleted genes in Cases 2 and 3 was 56.7% and 8.8% of Case 1, respectively. The average elapsed time was 1h47m55 s. It can be considered that Step 1 (gDel\_minRN) was not effective, but Steps 2 and 3 were effective for Recon3D. The reason for the low success ratio of gDel\_minRN may be that Recon3D is more complex as it considers three-dimension. Steps 2 and 3 of TrimGdel can be applied to any large gene deletion strategies derived by any method if growth-coupled production is achieved. However, it seems that such a method is not available other than gDel\_minRN at the moment.

### Comparison With Other Methods.

As shown in Table 6, TrimGdel had the highest success ratio among TrimGdel, GDSL, and optGene, for iML1515,



TABLE 7

Comparison of the Key Reaction Rates Between Cases 0, 1, 2, and 3 for Growth-Coupled Production of Biotin in iML1515

	Case 0	Cases 1 & 2	Case 3
	Natural state	after Step 1	after Step 3
#gene deletions	0	978	32
Growth	0.8770	0.1493	0.5272
Biotin synthase	$1.7 \times 10^{-6}$	0.1313	0.2629
MALCOAMT	$1.7 \times 10^{-6}$	0.1313	0.2629
SHSL1	0.1350	0.1543	0.3441
ATP synthase	70.43	57.37	72.33
Total flux	35447	922	20832

iMM904, and e\_coli\_core. GDLS had the minimum elapsed time, but its success ratio was low. As described in Section 3.2, many GDLS strategies resulted in target metabolite production without growth or growth-coupled production when the best PR was evaluated. However, when the worst PR at GR maximization was evaluated, growth-coupled production was rarely achieved. As optGene returns the best solution at that point when the specified computation time runs out, designating a longer computation time may increase the success ratio.

TrimGdel, optGene, and GDLS are in descending order of the average number of deleted genes, which is consistent with the descending order of the success ratio. One reason for the high success ratio of TrimGdel is the success in searching for larger gene deletion strategies. The elapsed time of TrimGdel, which was less than 20 m for every case, even for genome-scale models, is acceptable for actual use. The average numbers of deleted genes for TrimGdel were 8.90%, 6.02%, and 9.04% of whole genes, but they were less than 5% for more than a half of the gene deletion strategies.

For TrimGdel, the number of gene deletions was between 1% and 5% of whole genes for 66.5%, 50.5%, and 56.8% of the successful gene deletion strategies for iML1515, iMM904, and e\_coli\_core, respectively. On the other hand, for GDLS, the number of gene deletions was 1% or less of whole genes for 100%, 100%, 33.3% of the successful gene deletion strategies for iML1515, iMM904, and e\_coli\_core, respectively. These numbers were 100%, 80.1%, and 4% for optGene, respectively. It was seen that GDLS and optGene could derive gene deletion strategies whose sizes are 1% or more in a small model. However, we can conclude that TrimGdel is much more suitable for deriving gene deletion strategies whose sizes are between 1% and 5% of whole genes for genome-scale models.

#### Biotin Production.

In all cases described in Table 7, the values of biotin synthase and MALCOAMT are identical, which may imply that malonyl CoA should be produced in abundance in biotin production. The sum of the absolute reaction rates of all reactions (total flux) is correlated with the number of gene deletions and the growth rate but not with the biotin production rate. Because more gene deletions lead to more repressed reactions, it results in a lower total flux and GR. However, there seem to be other factors that determine whether growth-coupled production of biotin occurs.

It is difficult to derive a clear conclusion from Table 7. However, one possible hypothesis is that a lower GR and a relatively high ATP synthase rate lead to biotin production. In genome-scale constraint-based models, it is not easy to manually and visually analyze how each gene deletion strategy modifies the metabolic flow distribution and leads to growth-coupled production. An important future task is to develop algorithms and software to automatically analyze and extract the key points of metabolic distribution changes.

## REFERENCES

- [1] W. B. Copeland *et al.*, "Computational tools for metabolic engineering," *Metabolic Eng.*, vol. 14, no. 3, pp. 270–280, 2012.
- [2] A. P. Burgard, P. Pharkya, and C. D. Maranas, "OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnol. Bioeng.*, vol. 84, no. 6, pp. 647–657, 2003.
- [3] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: A computational framework for redesign of microbial production systems," *Genome Res.*, vol. 14, no. 11, pp. 2367–2376, 2004.
- [4] P. Pharkya and C. D. Maranas, "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems," *Metabolic Eng.*, vol. 8, no. 1, pp. 1–13, 2006.
- [5] K. R. Patil, I. Rocha, J. Förster, and J. Nielsen, "Evolutionary programming as a platform for in silico metabolic engineering," *BMC Bioinf.*, vol. 6, no. 1, 2005, Art. no. 308.
- [6] S. Ranganathan, P. F. Suthers, and C. D. Maranas, "OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions," *PLoS One Comput. Biol.*, vol. 6, no. 4, 2010, Art. no. e1000744.
- [7] I. Rocha *et al.*, "OptFlux: An open-source software platform for in silico metabolic engineering," *BMC Syst. Biol.*, vol. 4, no. 1, pp. 1–12, 2010.
- [8] Y. Toya and H. Shimizu, "Flux analysis and metabolomics for systematic metabolic engineering of microorganisms," *Biotechnol. Adv.*, vol. 31, no. 6, pp. 818–826, 2013.
- [9] J. D. Orth, I. Thiele, and B. Palsson, "What is flux balance analysis?," *Nature Biotechnol.*, vol. 28, no. 3, pp. 245–248, 2010.
- [10] D. S. Lun *et al.*, "Large-scale identification of genetic design strategies using local search," *Mol. Syst. Biol.*, vol. 5, no. 1, 2009, Art. no. 296.
- [11] G. Rockwell, N. J. Guido, and G. M. Church, "Redirector: Designing cell factories by reconstructing the metabolic objective," *PLoS One Comput. Biol.*, vol. 9, no. 1, 2013, Art. no. e1002882.
- [12] L. Yang, W. R. Cluett, and R. Mahadevan, "EMILiO: A fast algorithm for genome-scale strain design," *Metabolic Eng.*, vol. 13, no. 3, pp. 272–281, 2011.
- [13] D. Egen and D. S. Lun, "Truncated branch and bound achieves efficient constraint-based genetic design," *Bioinformatics*, vol. 28, no. 12, pp. 1619–1623, 2012.
- [14] N. E. Lewis *et al.*, "Omic data from evolved E coli are consistent with computed optimal growth from genome-scale models," *Mol. Syst. Biol.*, vol. 6, no. 1, 2010, Art. no. 390.
- [15] D. Gu, C. Zhang, S. Zhou, L. Wei, and Q. Hua, "IdealKnock: A framework for efficiently identifying knockout strategies leading to targeted overproduction," *Comput. Biol. Chem.*, vol. 61, pp. 229–237, 2016.
- [16] S. Ohno, H. Shimizu, and C. Furusawa, "FastPros: Screening of reaction knockout strategies for metabolic engineering," *Bioinformatics*, vol. 30, no. 7, pp. 981–987, 2014.
- [17] T. Tamura, "Grid-based computational methods for the design of constraint-based parsimonious chemical reaction networks to simulate metabolite production: GridProd," *BMC Bioinf.*, vol. 19, no. 1, 2018, Art. no. 325.
- [18] A. von Kamp and S. Klamt, "Enumeration of smallest intervention strategies in genome-scale metabolic networks," *PLoS One Comput. Biol.*, vol. 10, no. 1, 2014, Art. no. e1003378.
- [19] O. Hädicke and S. Klamt, "Computing complex metabolic intervention strategies using constrained minimal cut sets," *Metabolic Eng.*, vol. 13, no. 2, pp. 204–213, 2011.
- [20] B. J. Harder, K. Bettenbrock, and S. Klamt, "Model-based metabolic engineering enables high yield itaconic acid production by escherichia coli," *Metabolic Eng.*, vol. 38, pp. 29–37, 2016.

- [21] P. Schneider, A. von Kamp, and S. Klamt, "An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets," *PLoS One Comput. Biol.*, vol. 16, no. 7, 2020, Art. no. e1008110.
- [22] D. Banerjee *et al.*, "Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale," *Nature Commun.*, vol. 1, no. 1, pp. 1–11, 2020.
- [23] D. Machado and M. J. Herrgård, and I. Rocha, "Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction," *PLoS One Comput. Biol.*, vol. 12, no. 10, 2016, Art. no. e1005140.
- [24] Z. Razaghi-Moghadam and Z. Nikoloski, "GeneReg: A constraint-based approach for design of feasible metabolic engineering strategies at the gene level," *Bioinformatics*, vol. 37, no. 12, pp. 1717–1723, 2021.
- [25] T. Tamura, A. Muto-Fujita, Y. Tohsato, and T. Kosaka, "Gene deletion algorithms for minimum reaction network design by mixed-integer linear programming for metabolite production in constraint-based models: gDel\_minRN," 2021, doi: [10.21203/rs.3.rs-611371/v1](https://doi.org/10.21203/rs.3.rs-611371/v1).
- [26] I. Rocha, P. Maia, M. Rocha, and F. C. Ferreira, "OptGene: A framework for in silico metabolic engineering," in *Proc. 10th Int. Chem. Biol. Eng. Conf.*, 2008, pp. 218–219.
- [27] J. M. Monk *et al.*, "iML1515, a knowledgebase that computes *Escherichia coli* traits," *Nature Biotechnol.*, vol. 35, no. 10, pp. 904–908, 2017.
- [28] M. L. Mo, B. Palsson, and M. J. Herrgård, "Connecting extracellular metabolomic measurements to intracellular flux states in yeast," *BMC Syst. Biol.*, vol. 3, no. 1, 2009, Art. no. 37.
- [29] J. D. Orth, R. M. Fleming, and B. O. Palsson, "Reconstruction and use of microbial metabolic networks: The core *Escherichia coli* metabolic model as an educational guide," *EcoSal Plus*, vol. 4, pp. 1–48, 2010.
- [30] L. Heirendt *et al.*, "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0," *Nature Protoc.*, vol. 14, no. 3, pp. 639–702, 2019.
- [31] S. Liu, P. Kang, Z. Cui, Z. Wang, and T. Chen, "Increased riboflavin production by knockout of 6-phosphofructokinase I and blocking the entner–doudoroff pathway in *escherichia coli*," *Biotechnol. Lett.*, vol. 38, no. 8, pp. 1307–1314, 2016.
- [32] A. Schrijver, *Theory of Linear and Integer Programming*. Hoboken, NJ, USA: Wiley, 1998.
- [33] E. Brunk *et al.*, "Recon3D enables a three-dimensional view of gene variation in human metabolism," *Nature Biotechnol.*, vol. 36, no. 3, pp. 272–281, 2018.



**Takeyuki Tamura** (Member, IEEE) received the BE, ME, and the PhD degrees in informatics from Kyoto University, Japan, in 2001, 2003, and 2006, respectively. He joined Bioinformatics Center, Institute for Chemical Research, Kyoto University as a postdoctoral fellow, in 2006. He worked as an assistant professor from Dec. 2007 to Sep. 2017, and works as an associate professor from Oct. 2017. His research interests include bioinformatics and the theory of combinatorial optimization for graphs and networks.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/csdl](http://www.computer.org/csdl).