Gene deletion algorithms for minimum reaction network design by mixed-integer linear programming for metabolite production in constraint-based models: gDel_minRN

Takeyuki Tamura[1,*], Ai Muto-Fujita[2], Yukako Tohsato[3], Tomoyuki Kosaka[4,5]

1 Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan
2 RIKEN Center for Biosystems Dynamics Research, Suita, Osaka 565-0874, Japan
3 Faculty of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
4 Research Center for Thermotolerant Microbial Resources (RCTMR), Yamaguchi University, Yoshida, Yamaguchi 753-8515, Japan
5 Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yoshida, Yamaguchi 753-8515, Japan

Contact: tamura@kuicr.kyoto-u.ac.jp

Key words: Mixed integer linear programming, flux balance analysis, metabolic networks, gene deletions, growth-coupled production

A preprint whose DOI is https://doi.org/10.21203/rs.3.rs-611371/v1 is deposited on Research Square: https://www.researchsquare.com/article/rs-611371/v1

Abstract

Genome-scale constraint-based metabolic networks play an important role in the simulation of growth-coupled production, which means that cell growth and target metabolite production are simultaneously achieved. For growth-coupled production, a minimal reaction-network-based design is known to be effective. However, the obtained reaction networks often fail to be realized by gene deletions due to conflicts with gene-protein-reaction relations. Here, we developed gDel_minRN that determines gene deletion strategies using mixed-integer linear programming to achieve growth-coupled production by repressing the maximum number of reactions via gene-protein-reaction relations. The results of computational experiments showed that gDel_minRN could determine the core parts, which include only 30 to 55% of whole genes, for stoichiometrically feasible growth-coupled production for many target metabolites, which include useful vitamins such as biotin (vitamin B7), riboflavin (vitamin B2), and pantothenate (vitaminB5). Since gDel_minRN calculates a constraint-based model of the minimum number of gene-associated reactions without conflict with gene-protein-reaction relations, it helps biological analysis of the core parts essential for growth-coupled production for each target metabolite. The source codes, implemented in MATLAB using CPLEX and COBRA Toolbox, are available on {{https://github.com/MetNetComp/gDel-minRN}}

1 Introduction

Computational approaches are becoming increasingly important in the production of useful metabolites using microorganisms [1–7]. One of the most popular mathematical models in genome-scale metabolic engineering simulations is the constraint-based model. Constraint-based models mainly consist of metabolic networks and gene-protein-reaction (GPR) networks.

Metabolic networks represent the relationship between chemical reactions and compounds in cells. Many chemical reactions are catalyzed by enzymatic proteins encoded by genes. Therefore, metabolic networks can be controlled by gene deletions through reaction deletions. The relationships between reactions and genes are represented by GPR networks, in which the relationships between genes, proteins, and reactions are represented by Boolean functions.

In the metabolic network part of the constraint-based model, steady states are assumed in which each metabolic reaction speed (flux) is constant. Such an analysis is called flux balance analysis (FBA) [8]. In FBA, (1) for each compound, the sum of the producing fluxes is equal to the sum of the consuming fluxes; (2) in each reaction, the fluxes of substrates and products must satisfy the ratio in the chemical reaction equation, and (3) the upper and lower bounds are given for each flux.

The constraint-based model includes a virtual reaction that represents cell growth. The cell growth reaction of the constraint-based model was designed to match the results of the biological experiments. In the most standard FBA with constraint-based models, cell growth is maximized in the simulation because genotypes that result in higher cell growth are more likely to remain in the culture after repeated passaging. The cell growth reaction speed and the target metabolite-producing reaction speed are called the **growth rate (GR)** and **production rate (PR)**, respectively.

Therefore, in the simulation of useful metabolite production by FBA, we often evaluate PR when GR is maximized. When cell growth and the target metabolite production co-occur, we say that **growth-coupled production** is achieved. However, the number of metabolites for which growth-coupled production is achieved under natural conditions is limited. Therefore, it is often necessary to calculate the gene deletion strategy for the given constraint-based model and the target metabolite [9]. (See also Fig.1(A).)

Among the many existing methods [10–17], one of the most efficient methods for calculating reaction deletion strategies for growth-coupled production is the elementary flux vector-based method [18]. The elementary flux vector-based method determines a non-decomposable flux distribution that includes the cell growth reaction and the target metabolite production reaction and deletes the reactions that are not used by the flux distribution. In other words, this method selects a minimal number of reactions to be used in the flow where cell growth forces the production of the target metabolite and deletes reactions that are not used. It is also possible to combine such minimal reaction networks to obtain smaller reaction deletion strategies. It was shown that the elementary flux vector-based method could compute the reaction deletion strategies for growth-coupled production for most target metabolites for *Escherichia coli* and *Saccharomyces cerevisiae* under aerobic conditions by the combination of such core flows [18]. However, fewer than 10% of the reaction deletion strategies were feasible as gene deletion strategies because of the gene conflicts when the GPR network was considered [19, 20].

Therefore, it would be desirable if such a minimal reaction network detection-based method, in other words, core flow detection, for reaction deletion strategies could also detect the minimal

reaction network realized by gene deletion strategies. Such minimal reaction networks and gene deletion strategies help biological understanding of what is necessary for growth-coupled production. However, it is not straightforward to directly extend the calculation of reaction deletion strategies to the calculation of gene deletion strategies.

In this work, to achieve growth-coupled production by gene deletions, we have developed mixed-integer linear programming (MILP)-based algorithm, **gDel_minRN**, to calculate the gene deletion strategies that inactivate as many reactions as possible that are not essential for growth-coupled production. gDel_minRN calculates **g**ene **del**etion strategies that obtain the **min**imum **r**eaction **n**etwork for growth-coupled production.

In the computational experiments, gDel_minRN, GDLS [10] and optGene [21] were applied to iML1515 [22], iMM904 [23], and e_coli_core [24]: GDSL and optGene are some of the most widely used software to derive gene deletion strategies and are available in COBRA Toolbox [25]; 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 the metabolism of *E. coli*.

The success ratio of gDel_minRN for iML1515, iMM904, and e_coli_core were 40.6%, 13.6%, and 97.9%, which were substantially better than GDLS and optGene. The number of remaining genes by gDel_minRN was 35 to 38% for iML1515, 30 to 35% for iMM904, and 48 to 54% for e_coli_core of whole genes. Each gene deletion strategy was considered as successful when the minimum GR and PR were 0.001 or more at GR maximization.

The gene deletion strategies obtained by gDel_minRN do not contradict the GPR network and allow us to design metabolic networks that achieve growth-coupled production by repressing the maximum number of reactions. Therefore, if we analyze the gene deletion strategies obtained by gDel_minRN, we may be able to clarify the biological significance of the core part required for growth-coupled production for the target compounds without contradicting the gene-protein-reaction relationships. We conducted a biological analysis of the obtained gene deletion strategy for biotin growth-coupled production.

The remaining of this paper is organized as follows: Section 2.1 describes the main problem of this study mathematically; Section 2.2 illustrates the main problem with small examples; Section 2.3 illustrates the developed algorithm gDel_minRN with small examples and gives the pseudo-code; Section 2.4 illustrates the MILP formalization of gDel_minRN using small examples; Section 3.1 describes the performance comparison of gDel_minRN, GDLS, and optGene for iML1515, iMM904, and e_coli_core; Section 3.2 illustrates case study for vitamin production; Section 4 analyzes the results of the computational experiments and the case study; Section 5 gives a conclusion.

2 Method

2.1 Definition

Let $C = (M, R, S, L, U, G, F, P)$ be a constraint-based model, where $M = \{m_1, \ldots, m_a\}$, $R = \{r_1, \ldots, r_b\}$, $G = (g_1, \ldots, g_c)$, $F = (f_1, \ldots, f_b)$, and $P = (p_1, \ldots, p_b)$ are sets of metabolites, reactions, genes, Boolean functions, and the outputs of F, respectively. R always includes one special virtual reaction r_{growth} that represents cell growth, and the cell growth flux is represented by v_{growth}. 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 $V = \{v_1, \ldots, v_b\}$ be a set of reaction speeds per unit time (flux) of R. Let $L = \{l_1, \ldots, l_b\}$ and $U = \{u_1, \ldots, u_b\}$ be the sets of the lower and upper bounds for V, respectively.

$C_1 = (M, R, S, L, U)$ is called the **metabolic network part** of C. v_{growth} is called the **growth rate** (**GR**). In FBA using C_1, GR is maximized by the following linear programming (LP):

maximize

v_{growth}

such that

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

$l_j \leq v_j \leq u_j$ for all j

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

If the i-th column of S has only one non-zero element; in other words, r_i connects to only one metabolite, then r_i is called an **external reaction** (also called an exchange reaction), and is considered to be connected to the external environment. Reactions that are not external reactions are called **internal reactions**. The flux of the external reaction producing the target metabolite is called the **production rate** (**PR**). In this study, we evaluate PR at GR maximization. When PR is not uniquely determined at GR maximization, the minimum PR at GR maximization is evaluated.

In contrast, $C_2 = (G, F, P)$ is called the **GPR network part** of C, and

$$p_j = f_i(G), \quad \text{where} \quad p, g \in \{0, 1\}.$$

If $p_j = 0$, then l_j and u_j are forced to be 0. In other words,

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

hold.

The main problem of this study is formalized as follows.

Given

$C, r_{target}, PR_{threshold}, GR_{threshold}$

Find

$D \subset G$ that results in $v_{target} \geq PR_{threshold}, v_{growth} \geq GR_{threshold}$

such that minimize

v_{target}

such that maximize

v_{growth}

such that

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

$$\begin{cases} v_j = 0 \text{ if } p_j = 0 \\ l_j \leq v_j \leq u_j, \text{ otherwise} \end{cases}$$

$p_j = f_j(G)$

$$\begin{cases} g = 0 \text{ if } g \in D \\ g = 1, \text{otherwise} \end{cases}$$

$p, g \in \{0, 1\}$
$i = \{1, \ldots, a\}$, $j = \{1, \ldots, b\}$

2.2 Example for problem setting

Fig. 2 shows a small toy example of the constraint-based model, where $M = \{m_1, \ldots, m_4\}$, $R = \{r_1, \ldots, r_7\}$, and $S = \begin{pmatrix} 1 & -1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix}$.
Because $[\alpha, \beta]$ attached to r_j means that $\alpha \leq v_j \leq \beta$, $L = \{l_1, \ldots, l_7\}$ and $U = \{u_1, \ldots, u_7\}$ are as follows; $l_1, \ldots, l_7 = 0$, $u_1, \ldots, u_3 = 10$, $u_4, u_5 = 5$, $u_6, u_7 = 10$. For C_2, it is given that $G = \{g_1, \ldots, g_5\}$, $F = \{f_1, \ldots, f_7\}$ and

$p_1 = f_1 = \phi$,
$p_2 = f_2 = g_1 \wedge g_2 \wedge g_3$,
$p_3 = f_3 = g_1$,
$p_4 = f_4 = g_2 \wedge g_5$,
$p_5 = f_5 = (g_3 \vee g_4) \wedge g_5$,
$p_6 = f_6 = \phi$,
$p_7 = f_7 = \phi$.

Note that f represents a Boolean function, whereas p takes either 0 or 1. $p_j = f_j = \phi$ means that r_j cannot be repressed via gene deletions.

In the original state, when GR (v_6) is maximized, $v_6 = 10$ is obtained. To achieve $v_6 = 10$, there are two options: using only r_2 or using r_3, r_4 in addition to r_2. The latter corresponds to the optimistic case for PR (v_7): $v_7 = 5$ is obtained by $v_2 = v_3 = v_4 = 5$ as shown in ID 1 of Fig.2(B). The former corresponds to the pessimistic case for PR: all fluxes from r_1 flow through r_2 to r_6; $v_1 = v_2 = v_6 = 10$ and $v_3 = v_4 = v_5 = v_7 = 0$ are obtained as shown in ID 2 of Fig.2(B).

If g_1 is deleted, then $p_2 = g_1 \wedge g_2 \wedge g_3 = 0$ since $g_1 = 0$. Therefore, r_2 does not work and v_2 is forced to be zero. Similarly, r_3 does not work and v_3 is forced to be zero because $p_3 = g_1 = 0$ holds. Therefore, when GR is maximized, fluxes from r_1 cannot reach r_6, and GR becomes 0. In the optimistic case for PR, $v_1 = v_5 = v_7 = 5$ is obtained, but no flux flows in the pessimistic case, as shown in IDs 3 and 4 of Fig.2(B), respectively. To ensure growth-coupled production, we need to evaluate the pessimistic case for PR, and the maximized GR must exceed the minimum required value. Therefore, we consider that growth-coupled production cannot be achieved by deleting g_1. When g_2 is deleted, similar results are obtained because neither r_2 nor r_4 works as shown in IDs 5 and 6.

If g_3 is deleted, r_2 does not work, but the other reactions can work. Therefore, the maximum GR is five because $0 \leq v_4 \leq 5$. In the optimistic case, the flux from r_1 flows to r_7 via r_5 in addition to via r_3 and r_4. In this case, GR=5 and PR=10 are obtained. However, in the pessimistic case, GR=PR=5 is obtained as shown in IDs 7 and 8 of Fig.2(B), respectively.

If g_4 is deleted, $p_j = 1$ for all j: the situation is the same as the original state. If g_5 is deleted, neither r_4 nor r_5 works since $p_4 = p_5 = 0$: $v_1 = v_2 = v_6 = 10$ and $v_3 = v_4 = v_5 = v_7$ are obtained

as shown in ID 11. If g_1 and g_2 are deleted, the situation is the same as when only g_1 is deleted.

Suppose that GRLB=PRLB=1; that is, the minimum required GR and PR are 1. Then, deleting g_3 achieves growth-coupled production because GR=PR=5 is obtained even for the pessimistic case, and GR≥GRLB and PR≥PRLB are satisfied. Thus, in this example, growth-coupled production can be achieved by deleting one gene g_3. However, in practice, it may be necessary to examine all genes on and off, which results in a combinatorial explosion.

2.3 Algorithm

The developed algorithm gDel_minRN first searches, using MILP, the flux and corresponding gene deletions that satisfy

(1) GR and PR are above the given thresholds, GRLB and PRLB, respectively.
(2) The number of reactions repressed by gene deletions is maximum.
(3) GR is maximized where (2) has a higher priority than (3).

It should be noted that the GR and PR obtained above are not always realized when GR is maximized without PRLB. Therefore, gDel_minRN tests whether the obtained gene deletion strategy achieves growth-coupled production under the condition that GR is maximized without PRLB. In particular, gDel_minRN checks the lowest PR value when GR is maximized. If the obtained gene deletion strategy does not achieve growth-coupled production in this pessimistic case, then the gene deletion strategy is added to the prohibited list, and another gene deletion strategy is searched in the same way by MILP.

For example, suppose that GRLB=PRLB=1 in Fig.2(A). When GR is maximized under the conditions of GR≥1 and PR≥1, the flux distribution for each gene deletion strategy for the pessimistic case of PR is summarized in Table 1(A). Because deleting g_1, g_2, or g_5 cannot satisfy GR≥GRLB or PR≥PRLB, the gene deletion strategy candidates that can satisfy (1) are limited to $\{g_3\}$, $\{g_4\}$ and $\{g_3, g_4\}$. The number of repressed reactions by deleting $\{g_3\}$, $\{g_4\}$ and $\{g_3, g_4\}$ are 1, 0 and 2, respectively, as shown in Table 1(A).

Therefore, gene deletion strategies are applied in the order of $\{g_3, g_4\}$, $\{g_3\}$, $\{g4\}$. When $\{g_3, g_4\}$ is deleted and GR is maximized without PRLB, GR=PR=5 is obtained and growth-coupled production is achieved as as shown in Table 1(B). If the first candidate fails to achieve growth-coupled production, the second candidate is applied. When the gene deletion strategy is not uniquely determined under the condition that the number of repressed reactions is maximized, gDel_minRN selects the gene deletion strategy whose GR is maximum among them. The flux distributions for $\{g_3\}$ and $\{g_4\}$ for the pessimistic case of PR at GR maximization are also shown in Table 1(B).

The pseudo-code of gDel_minRN is as follows.

Procedure **gDel_minRN**$(C, v_{target}, maxloop, \alpha, \beta)$
$TMPR =$**max** v_{target} /*theoretical maximum PR*/
s.t. $\Sigma_j \ S_{i,j} \cdot v_j = 0$ for all $1 \le i \le a$
$LB_j \le v_j \le UB_j$ for all $1 \le j \le b$
$PRLB = \alpha \cdot TMPR$

$TMGR$ =**max** v_{growth} /*thoeretical maximum GR*/
s.t. $\Sigma_j \ S_{i,j} \cdot v_j = 0$ for all $1 \leq i \leq a$
$LB_j \leq v_j \leq UB_j$ for all $1 \leq j \leq b$
$GRLB = \beta \cdot TMGR$
* Finding a gene deletion strategy candidate.*/
$prohibited\_list = \phi, loop = 1$
while $loop \leq maxloop$
/*maximize #repressed reactions first and GR second.*/
max $TMGR \cdot KO + v_{growth}$
/*KO: the number of repressed reactions.*/
s.t. $\Sigma_j S_{i,j} \cdot v_j = 0$ for all $1 \leq i \leq a$
$\begin{cases} v_j = 0 \ if \ p_j = 0 \\ l_j \leq v_j \leq u_j, \ otherwise \end{cases}$
$p_j = f_j(G)$ /*by the methods of Table 4.*/
$\begin{cases} g = 0 \ \text{ if } \ g \in D \ /*D \text{ is the set of deleted genes.}*/ \\ g = 1, otherwise \end{cases}$
$D \notin prohibited\_list$
GRLB $\leq v_{growth}$
PRLB $\leq v_{target}$
$D_{candidate} = D$
/*Check whether growth-coupled production is achieved by $D_{candidate}$.*/
min
v_{target}
such that max
v_{growth}
such that
$\Sigma_j S_{ij} v_j = 0$ for all i
$\begin{cases} v_j = 0 \ if \ p_j = 0 \\ l_j \leq v_j \leq u_j, \ otherwise \end{cases}$
$p_j = f_j(G)$
$\begin{cases} g = 0 \ \text{ if } \ g \in D_{candidate} \\ g = 1, otherwise \end{cases}$
if v_{target} ≥PRLB and v_{growth} ≥GRLB **then**
return $D_{candidate}, v_{target}, v_{growth}$
else
$prohibited\_list = prohibited\_list \cup D_{candidate}$
$loop = loop + 1$

2.4 Example of MILP formalization

In this section, the MILP formalization by which gDel_minRN determines $D_{candidate}$ is illustrated using the example of Fig.2(A).

Table 2 summarizes variables used in the MILP formalization. The real number variables x_1 to x_7 represent v_1 to v_7, which are reaction rates of r_1 to r_7. The binary variables x_8 to x_{12} represent g_1 to g_5, which represent whether genes are deleted. The binary variable x_{13} represent $(g3 \vee g4)$, which is the internal term of the GPR rule for r_5. The binary variables x_{14} to x_{17} represent whether r_2 to r_5 is repressed.

Table 3 shows linear constraints derived from Boolean functions in GPR rules. KO are variables that represent reactions are repressed: x_{14} to x_{17}. For example, $KO3$ represents whether r_3 is repressed. Since the value of g_1 and $KO3$ is represented by x_8 and x_{15}, respectively, $-x_8 + x_{15}$ holds. AND and OR functions can be converted into linear constraints by the methods represented by Table 4. Since g_3 and g_4 are represented by x_{10} and x_{11}, the Boolean function $x_{13} = g_3 \vee g_4$ can be converted into $x_{10} + x_{11} - 2x_{13} \leq 0$ and $-x_{10} - x_{11} + x_{13} \leq 0$.

Suppose that MILP is formalized by the following notations:

minimize

fv

such that

$A_{eq}v = b_{eq}$

$Av \leq b$

$lb \leq v \leq ub$

$type$ /*that describes which variables are integers.*/

Let nr, ng, nt, and nko be the number of reactions, genes, internal functions, and repressible reactions, respectively. Suppose that the role of each variable for the general case of gDel_minRN is represented as Table 5.

Then, Aeq, beq, A, and b can be represented as Fig. 3 (A): Beq, B and Bb, that are used in Aeq, A and b, can be represented as Fig. 3 (B).

3 Computational experiments

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 1TB SSD. This workstation had CPLEX 12.10, COBRA Toolbox v3.0 [25], and MATLAB R2021a installed and used for these analyses. When a target metabolite does not have an external (exchange) reaction, an auxiliary exchange reaction was temporarily added to the model to simulate the secretion. The unit of every reaction rate is mmol/gDW/h, which will be omitted hereafter for simplicity of notation.

3.1 Performance comparison by comprehensive experiments

In the computational experiments, we applied gDel_minRN to iML1515, iMM904, and e_coli_core to compare the performance of existing methods, GDLS and optGene. The number of genes, reactions, and metabolites for each model are summarized in Table 6. When a gene deletion makes the maximum GR zero, we call it an essential gene.

As some metabolites were always produced in the original states or their production were proved to be stoichiometrically impossible for any gene deletion strategies, they were excluded

from the target metabolites. Such metabolites were determined by flux variability analysis and calculating the theoretical maximum production rate.

The number of target metabolites was 1085, 773, and 48 for iML1515, iMM904, and e_coli_core, respectively: Seven, nine, and four metabolites were producible in the original state for each model, even in the worse case; The theoretical maximum PR was zero for 785, 444, and 20 metabolites for each model.

Table 7(A) summarizes the performance of gDel_minRN. The minimum PR and GR at GR maximization were 0.001 or more by gDel_minRN strategies for 441, 105, and 47 target metabolites for iML1515, iMM904, and e_coli_core, respectively: The success ratios were 40.6%, 13.6%, and 97.9%; The average number of remaining genes was 555.2, 291.85, and 69.47; The maximum number of remaining genes was 571, 314, and 73; The minimum number of remaining genes was 539, 275, and 66. The number of remaining genes ranged from 35 to 38%, 30 to 35%, and 48 to 54% of whole genes for iML1515, iMM904, and e_coli_core, respectively.

Table 7(B) summarizes the performance of GDLS. The minimum PR and GR were 0.001 or more at GR maximization by GDLS strategies for 0, 0, and 5 target metabolites for iML1515, iMM904, and e_coli_core, respectively: The success ratios were 0%, 0%, and 10.4%. In the GDLS strategies, the maximum PR at GR maximization was 0.001 or more for many target metabolites: 1085, 128, 13 for iML1515, iMM904, and e_coli_core, respectively. However, GR or the minimum PR at GR maximization was zero for 1085, 773, and 43 target metabolites. Therefore, only five gene deletion strategies for e_coli_core resulted in growth-coupled production when the worst PR was evaluated at GR maximization. The average, maximum and minimum number of remaining genes for the successful strategies was 68.4, 71, and 66 for e_coli_core, respectively; The number of remaining genes ranged from 48 to 52% of whole genes.

Table 7(C) summarizes the performance of optGene. The minimum PR and GR were 0.001 or more at GR maximization by optGene strategies for 0, 30, and 22 target metabolites for iML1515, iMM904, and e_coli_core, respectively: The success ratios were 0%, 3.9%, and 45.8%. For the successful strategies, the average number of remaining genes was 897.4 and 130.3 for iMM904 and e_coli_core; The maximum number of remaining genes was 895 and 136; The minimum number of remaining genes was 901 and 127. The number of remaining genes ranged 98 to 100%, 92 to 100% for iMM904 and e_coli_core, respectively.

3.2 Case study using vitamins

Table 8 represents a case study for gDel_minRN strategies for three vitamins, pantothenate (vitamin B5), biotin (vitamin B7), and riboflavin (vitamin B2). These three metabolites are highly valuable, but no effective biosynthesis methods have been established except riboflavin. For pantothenate, the number of remaining genes by gDel_minRN was 555. PR and GR were 0.7444 and 0.2485, respectively. As TMPR and TMGR were 5.511 and 0.877, PR/TMPR and GR/TMGR were 0.135 and 0.283. The elapsed time was 4m24s. For biotin, the number of remaining genes by gDel_minRN was 540. PR and GR were 0.1313 and 0.1493, respectively. As TMPR and TMGR were 1.307 and 0.877, PR/TMPR and GR/TMGR were 0.100 and 0.170. The elapsed time was 5m8s. For riboflavin, the number of remaining genes by gDel_minRN was 544. PR and GR were 0.1198 and 0.1212, respectively. As TMPR and TMGR were 2.739 and 0.877, PR/TMPR and GR/TMGR were 0.044 and 0.138. The elapsed time was 3m34s.

The obtained biotin production pathway was analyzed biologically using Escher [26] and KEGG Mapper [27] as follows. In the obtained pathway for biotin production by gDel_minRN, it was observed that the pathways from acetyl-CoA to acetate were removed from the map. The acetyl-CoA obtained in glycolysis was consumed in the TCA circuit or converted to acetate and was also used to generate malonyl-CoA. Since malonyl-CoA is located at the beginning of the biotin-generating pathway, we hypothesized that by inhibiting the conversion of acetyl-CoA to acetate, acetyl-CoA that was not fully consumed by the TCA cycle was used for biotin generation via malonyl-CoA.

To test this hypothesis, we revived all eight deleted genes (b0871, b2296, b0968, b2297, b2458, b4069, b3588, b1241) located on the pathways from acetyl-CoA to acetate. As a result, GR = 0.3341 and PR = 0 were obtained. This reinforces the hypothesis that by removing the pathways from the acetyl-CoA to the acetate, the substrate used for cell growth was replaced by biotin production via malonyl-CoA.

4 Discussion

Performance comparison of gDel_minRN

As shown in Table 7, gDel_minRN, optGene, and GDLS are in descending order of the success ratio for every experiment: the success ratio by gDel_minRN, GDLS, and optGene were 40.6%, 0%, and 0% for iML1515; 13.6%, 0%, and 3.9% for iMM904; 97.9%, 10.4%, and 45.8% for e_coli_core.

The number of remaining genes by gDel_minRN and GDLS were less than that by optGene: 35 to 54% for gDel_minRN, 48 to 52% for GDLS, and 92 to 100% for optGene. As the success ratio by GDLS was 0% for iML1515 and iMM904 and the success ratio by optGene was 0% for iML1515, the three methods were directly comparable only for e_coli_core.

As shown in Table 8, gDel_minRN successfully derived gene deletion strategies for growth-coupled production of pantothenate, biotin, and riboflavin. The elapsed time was 3 to 6m, which was acceptable for practical use.

In metabolite production using microorganisms, it has been necessary to minimize the number of genes to be deleted in terms of cost and accuracy [20, 28]. However, gDel_minRN maximizes the number of reactions that are repressed to obtain the core part necessary for growth-coupled production, so it would rather delete as many genes as possible. Therefore, the obtained gene deletion strategies are quite different from those obtained using existing methods. Such a gene deletion strategy is helpful for biological analysis of which part of the constraint-based model is necessary for growth-coupled production but may not be practical for metabolite production with current metabolic engineering technology. However, it could be useful if zero-based DNA synthesis for metabolite production is possible in the future [29, 30].

The problem of finding gene deletion strategies for growth-coupled production is NP- hard since even the reaction deletion problem is NP-hard [31, 32]. For genome-scale models, it is practically impossible to consider the solution space consisting of all 2^n gene deletions strategies, where n is the number of genes. Therefore, each method for genome-scale models needs to narrow the search space to find gene deletion strategies within reasonable computation time. To this end, many methods limit the size of gene deletions to a small constant k: the number of candidates for

gene deletion strategies is narrowed down to $O(n^k)$. However, such methods cannot find large deletion strategies for genome-scale models. If such k is not applied, GDLS can find large gene deletion strategies for small networks but cannot for genome-scale models as the search space is too large.

Each constraint-based model has essential genes whose deletion results in non-growth. Gene deletion strategies for growth-coupled production never include essential genes in the simulations because GR is always maximized in this study. The core networks necessary for growth-coupled production are induced by essential genes and non-essential genes that are determined by gDel_minRN. The reason why gDel_minRN can effectively determine gene deletion strategies is that the number of non-essential genes necessary for growth-coupled production is not significant: the solution space is effectively narrowed.

On the other hand, for reaction deletions, the idea of finding a core network for growth-coupled production has been studied using elementary flux vector-based methods [18]. However, because the deletion strategies obtained by the existing methods often conflict with GPR networks, it was difficult to extend their reaction deletion strategies to gene deletion strategies [20].

A number of mixed-integer linear programming (MILP)-based methods have been proposed for calculating gene or reaction deletion strategies that result in growth-coupled production [1, 2, 33, 34]. Solving MILP is an NP-complete problem and requires computation time proportional to the exponential function of the number of reactions and genes. Methods for reducing the size of the constraint-based models were also proposed [35]. Many methods that are not limited to MILP have been proposed to speed up the computation time by avoiding the optimization of PR [10–17]. However, to the best of our knowledge, there was no method for calculating the gene deletion strategy that resulted in a minimal network for growth-coupled production.

Case study using vitamins

Vitamins have been industrially produced by chemical synthesis and biosynthesis. Considering sustainability, biosynthesis is more promising than chemical synthesis, which produces pollutants, and improvement of vitamin biosynthesis is still needed because high productivity and cost savings are important factors [36]. The reduction of metabolic pathways leads to efficient vitamin production by reducing the amount of protein required. When gDel_minRN was applied, gene deletion strategies for growth-coupled production were successfully obtained for these three vitamins.

One of the motivations for developing gDel_minRN was to calculate the core parts required for growth-coupled production and to biologically elucidate which features are necessary for growth-coupled production and which are not. Among the three gene deletion strategies obtained by gDel_minRN, most genes were deleted in the case of biotin.

Since the existing basic strategy for improving biotin productivity using bacterial cells is the overexpression of rate-limiting enzymes, removal of negative regulators and addition of intermediates or precursors [37], complete optimization of the metabolic pathways by altering the whole genomic network has not been extensively tested. The constructed pathway for biotin synthesis from iML1515, a recent solid computational model for *E. coli* metabolism, with the lowest number of reactions by gDel_minRN in this study, showed new possibilities for the *E. coli* metabolic pathway that can be changed from the original genome.

Although the constructed pathway is stoichiometrically reasonable because iML1515 has an almost complete metabolic network [22], it is not clear whether it can be created in *E. coli* real cells. Therefore, we considered this pathway from a biological point of view. The constructed pathway from glucose to biotin can be separated into two phases, from glucose to malonyl acyl-carrier-protein (ACP) and malonyl-ACP to biotin, respectively (Fig. 4A). For biotin production, S-adenosylmethionine (SAM) and L-alanine are required to synthesize and adjust the production ratio in the upper pathway to drive the lower pathway (Fig. 4A). The reactions in the lower pathway were not so unique because almost one connected pathway from malonyl-ACP to biotin in *E. coli* [36]. On the other hand, the biological consideration of the upper pathway, glucose to malonyl-ACP, revealed notable characteristics.

One of the interesting characteristics of the constructed pathway is the requirement for aerobic metabolism. In these reactions, a high amount of NADPH is produced from glucose to ribose 5-phosphate pathway, and oxidation is performed in dihydroxyacetone phosphate to glycerol 3-phosphate by glycerol-3-phosphate dehydrogenase, and NADP could then be produced (Fig. 4B). Countering, in the opposite direction from glycerol 3-phosphate to dihydroxyacetone phosphate utilizing ubiquinone-8 (UQ8) as an electron acceptor to produce $UQ8H_2$ (Fig. 4B). In addition, the reactions for pyruvate to lactate and succinate to fumarate generate $UQ8H_2$. These reactions cause a high accumulation of $UQ8H_2$; oxidation is required to proceed with the metabolic reaction accomplished by using oxygen as an electron acceptor on the respiratory chain, which also causes adenosine triphosphate (ATP) production in respiratory chains. High ATP production also requires not only the lower pathway but also nucleotide and amino acid synthesis. Therefore, this pathway requires oxygen or a respiratory oxidative substrate. We did not investigate the effect of the presence of oxygen on the constructed pathway. Therefore, future experiments should consider how substrate and culture conditions affect this pathway.

The second characteristic is that the intermediates of this pathway do not consider cytotoxicity. The upper pathway utilizes methylglyoxal as the intermediate from dihydroxyacetone phosphate to lactate (Fig. 4B). The methylglyoxal utilizing pathway is known in 1,2-propanediol-producing bacteria, but it shows that high cytotoxicity [38]. This pathway is possible but has problems. Several microorganisms for 1,2-propanediol production consider the pathway not to be exchanged because of the reduction in growth or production by the pathway [39]. This suggests that if we try to resolve more cell-suitable pathways, we need some tricks to avoid using the pathways from the literature to produce more realistic computational minimum pathway predictions for production.

Finally, several problematic points for the construction or reproduction of this pathway in *E. coli* were found, but the constructed pathway was almost biologically possible in our consideration. Interestingly, when using a short and small number of reactions for some material production, cells can reduce the protein amount, which finally guides more efficient material production by the cell. The biological consideration of this pathway is only a knowledge base, and an experimental demonstration of this pathway on a cell should be performed in the future.

5 Conclusion

In this study, we developed gDel_minRN to calculate gene deletion strategies that repress as many reactions as possible to achieve growth-coupled production. Computer experiments using two

genome-scale models and a core model showed that gDel_minRN could find strategies that deleted 45 to 70% of all genes. Unlike existing biosynthetic methods, the strategy obtained by gDel_minRN is based on a fundamental modification of the metabolic pathway. Existing computational methods aimed to delete a small number of genes or computed core networks by deleting reactions, and their purpose was fundamentally different from that of gDel_minRN, which calculates core networks by gene deletion. Analyzing gene deletion strategies obtained by gDel_minRN is helpful for biological analysis of which parts are necessary for growth-coupled production.

Acknowledgements

The authors appreciate their families and colleagues.

Authorship confirmation/contribution

TT designed this work, developed the algorithm, implemented the software, and conducted the computational experiments. AMF performed the analysis using the database. YT performed the visualization analysis. TK provided the biological interpretation of the experimental results. TT and TK wrote the manuscript. All authors have read the manuscript and approved it.

Conflict of interest

Not applicable.

Funding

TT, AMF, YT, and TK were partially supported by grants from JSPS, KAKENHI #20H04242. No funding body played any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

References

1. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering. 2003;84(6):647-57.

2. Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. Genome research. 2004;14(11):2367-76.

3. Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metabolic engineering. 2006;8(1):1-13.

4. Patil KR, Rocha I, Förster J, Nielsen J. Evolutionary programming as a platform for in silico metabolic engineering. BMC bioinformatics. 2005;6(1):308.

5. Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol. 2010;6(4):e1000744.

6. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: an open-source software platform for in silico metabolic engineering. BMC systems biology. 2010;4(1):1-12.

7. Toya Y, Shimizu H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. Biotechnology advances. 2013;31(6):818-26.

8. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature biotechnology. 2010;28(3):245-8.

9. Vieira V, Maia P, Rocha M, Rocha I. Comparison of pathway analysis and constraint-based methods for cell factory design. BMC bioinformatics. 2019;20(1):1-15.

10. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, et al. Large-scale identification of genetic design strategies using local search. molecular systems biology. 2009;5(1):296.

11. Rockwell G, Guido NJ, Church GM. Redirector: designing cell factories by reconstructing the metabolic objective. PLoS Comput Biol. 2013;9(1):e1002882.

12. Yang L, Cluett WR, Mahadevan R. EMILiO: a fast algorithm for genome-scale strain design. Metabolic engineering. 2011;13(3):272-81.

13. Egen D, Lun DS. Truncated branch and bound achieves efficient constraint-based genetic design. Bioinformatics. 2012;28(12):1619-23.

14. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Molecular systems biology. 2010;6(1):390.

15. Gu D, Zhang C, Zhou S, Wei L, Hua Q. IdealKnock: a framework for efficiently identifying knockout strategies leading to targeted overproduction. Computational biology and chemistry. 2016;61:229-37.

16. Ohno S, Shimizu H, Furusawa C. FastPros: screening of reaction knockout strategies for metabolic engineering. Bioinformatics. 2014;30(7):981-7.

17. Tamura T. Grid-based computational methods for the design of constraint-based parsimonious chemical reaction networks to simulate metabolite production: GridProd. BMC bioinformatics. 2018;19(1):325.

18. von Kamp A, Klamt S. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. Nature communications. 2017;8:15956.

19. Machado D, Herrgård MJ, Rocha I. Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. PLoS computational biology. 2016;12(10):e1005140.

20. Razaghi-Moghadam Z, Nikoloski Z. GeneReg: A constraint-based approach for design of feasible metabolic engineering strategies at the gene level. Bioinformatics. 2020.

21. Rocha I, Maia P, Rocha M, Ferreira EC. OptGene: a framework for in silico metabolic engineering. 2008.

22. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. Nature biotechnology. 2017;35(10):904-8.

23. Mo ML, Palsson BØ, Herrgård MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC systems biology. 2009;3(1):37.

24. Orth JD, Fleming RM, Palsson BO. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. EcoSal plus. 2010.

25. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. Nature protocols. 2019;14(3):639-702.

26. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLoS Comput Biol. 2015;11(8):e1004321.

27. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Science. 2020;29(1):28-35.

28. Apaolaza I, Valcarcel LV, Planes FJ. gMCS: fast computation of genetic minimal cut sets in large networks. Bioinformatics. 2019;35(3):535-7.

29. Mori Y, Noda S, Shirai T, Kondo A. Direct 1, 3-butadiene biosynthesis in Escherichia coli via a tailored ferulic acid decarboxylase mutant. Nature communications. 2021;12(1):1-12.

30. Shimizu Y, Kuruma Y, Kanamori T, Ueda T. The PURE system for protein production. In: Cell-Free Protein Synthesis. Springer; 2014. p. 275-84.

31. Yousofshahi M, Orshansky M, Lee K, Hassoun S. Probabilistic strain optimization under constraint uncertainty. BMC systems biology. 2013;7(1):1-13.

32. Deng X. Complexity issues in bilevel linear programming. In: Multilevel optimization: Algorithms and applications. Springer; 1998. p. 149-64.

33. Kim J, Reed JL, Maravelias CT. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One. 2011;6(9):e24162.

34. Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 2010;26(4):536-43.

35. Röhl A, Bockmayr A. A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks. BMC bioinformatics. 2017;18(1):1-10.

36. Acevedo-Rocha C, Gronenberg L, Mack M, Commichau F, Genee H. Microbial cell factories for the sustainable manufacturing of B vitamins. Curr Opin Biotechnol. 2019;56:18-29.

37. Xiao F, Wang H, Shi Z, Huang Q, Huang L, Lian J, et al. Multi-level metabolic engineering of *Pseudomonas mutabilis* ATCC31014 for efficient production of biotin. Metab Eng. 2020;61:406-15.

38. Booth I, Ferguson G, Miller S, Li C, Gunasekera B, Kinghorn S. Bacterial production of methylglyoxal: a survival strategy or death by misadventure. Biochem Soc Trans. 2003;31(Pt 6):1406-8.

39. Niu W, Kramer L, Mueller J, Liu K, Guo J. Metabolic engineering of *Escherichia coli* for the de novo stereospecific biosynthesis of 1,2-propanediol through lactic acid. Metab Eng Commun. 2019;8:e00082.

Table 1. (A) The flux distribution for each gene deletion strategy when GR is maximized under the condition with GR≥1 and PR≥1. (B) The priority of each gene deletion strategy candidate and the resulting flux distribution for the pessimistic case of PR at GR maximization.

KO	v_1	v_2	v_3	v_4	v_5	v_6	v_7	reactions
g_1	-	-	-	-	-	-	-	cannot satisfy GR≥1
g_2	-	-	-	-	-	-	-	cannot satisfy GR≥1
g_3	5	0	0	0	0	5	5	r_2
g_4	10	9	1	1	0	9	1	ϕ
g_5	-	-	-	-	-	-	-	cannot satisfy PR≥1
g_3, g_4	5	0	5	5	0	5	5	r_2, r_5

(A)

KO	priority	v_1	v_2	v_3	v_4	v_5	v_6	v_7
g_3, g_4	1	5	0	5	5	0	5	5
g_3	2	5	0	5	5	0	5	5
g_4	3	10	10	0	0	0	10	0

(B)

Table 2. Variables used in the MILP formalization in gDel_minRN for the example of Fig.2(A).

Variables	Type	Object
x_1 to x_7		reactions r_1 to r_7
x_8 to x_{12}	binary	genes g_1 to g_5
x_{13}	binary	internal term(s) ($g_3 \vee g_4$)
x_{14} to x_{17}	binary	whether repressed or not for r_2, r_3, r_4, r_5

Table 3. The linear constraints for the GPR rules in Fig.2.

Boolean functions		Linear constraints
$KO3 = g_1$	$\longrightarrow$	$-x_8 + x_{15} = 0$
$x_{13} = g_3 \vee g_4$	$\longrightarrow$	$x_{10} + x_{11} - 2x_{13} \leq 0$
		$-x_{10} - x_{11} + x_{13} \leq 0$
$KO2 = g_1 \wedge g_2 \wedge g_3$	$\longrightarrow$	$-x_8 - x_9 - x_{10} + 3x_{14} \leq 0$
		$x_8 + x_9 + x_{10} - x_{14} \leq 2$
$KO4 = g_2 \wedge g_5$	$\longrightarrow$	$-x_9 - x_{12} + 2x_{16} \leq 0$
		$x_9 + x_{12} - x_{16} \leq 1$
$KO5 = x_{13} \wedge g_5$	$\longrightarrow$	$-x_{12} - x_{13} + 2x_{17} \leq 0$
		$x_{12} + x_{13} - x_{17} \leq 1$

Table 4. The methods for representing Boolean functions by linear constraints.

Boolean functions		Linear constraints
$y = x_1 \wedge x_2 \wedge \cdots \wedge x_k$	$\longrightarrow$	$-x_1 - \ldots - x_k + ky \leq 0$
		$x_1 + \cdots + x_k - y \leq k - 1$
$y = x_1 \vee x_2 \vee \cdots \vee x_k$	$\longrightarrow$	$x_1 + \cdots + x_k - ky \leq 0$
		$-x_1 - \cdots - x_k + y \leq 0$

Table 5. The purpose and type of variables used in MILP for the general case of gDel_minRN.

Variables	Type	For
x_1 to x_{nr}	real	reaction fluxes
x_{nr+1} to x_{nr+ng}	binary	genes
$x_{nr+ng+1}$ to $x_{nr+ng+nt}$	binary	internal terms
$x_{nr+ng+nt+1}$ to $x_{nr+ng+nt+nko}$	binary	reaction repressions

Table 6. The constraint-based models that were used in the computational experiments.

Model	iML1515	iMM904	e_coli_core
#genes	1516	905	137
#reactions	2712	1577	95
#metabolites	1877	1226	72
#target metabolites	1085	773	48
#essential genes	196	110	7

Table 7. The performance comparison between gDel_minRN, GDLS, and optGene. Each gene deletion strategy was considered as successful when the minimum GR and PR were 0.001 or more at GR maximization.

Model	iML1515	iMM904	e_coli_core
#success	441	105	47
#success ratio	40.6%	13.6%	97.9%
Avg. #genes	555.2	291.85	69.47
Max #genes	571	314	73
Min #genes	539	275	66
Range	35-38%	30-35%	48-54%
Time	7m35s	49s	0.40s

(A) gDel_minRN

Model	iML1515	iMM904	e_coli_core
#success	0	0	5
#success ratio	0%	0%	10.4%
Avg. #genes	-	-	68.4
Max #genes	-	-	71
Min #genes	-	-	66
Range	-	-	48-52%
Time	39s	0.83s	0.812s

(B) GDLS

Model	iML1515	iMM904	e_coli_core
#success	0	30	22
#success ratio	0%	3.9%	45.8%
Avg. #genes	-	897.4	130.3
Max #genes	-	895	136
Min #genes	-	901	127
Range	-	98-100%	92-100%
Avg. time	20m3s	20m20s	20m6s

(C) optGene

Table 8. Case study of gDel_minRN performance for three vitamins.

Target	#used genes	PR	GR	time
Pantothenate	555	0.7444	0.2485	4m24s
Biotin	540	0.1313	0.1493	5m8s
Riboflavin	544	0.1198	0.1212	3m34s

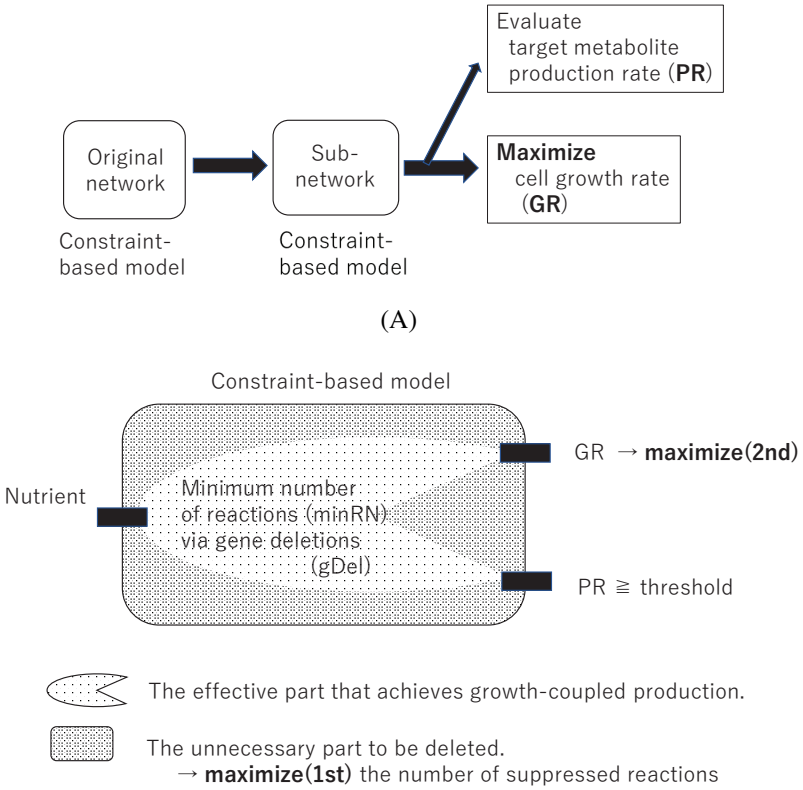
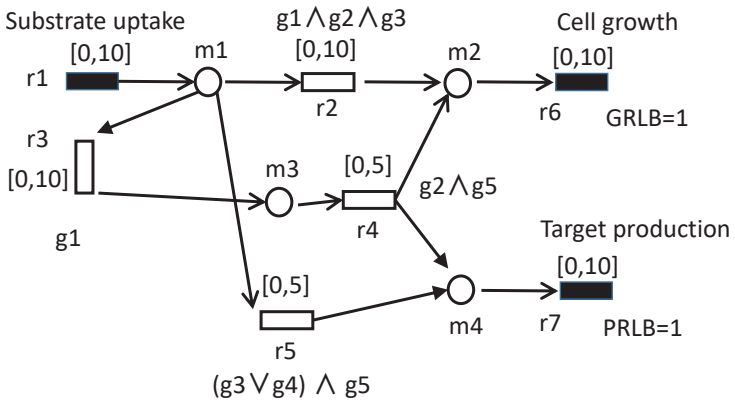


Figure 1. (A) Problem setting of this study. The minimum PR of the target metabolite is evaluated when the GR is maximized. (B) The idea of gDel_minRN algorithm. The maximum number of reactions are repressed via gene deletions for growth-coupled production.

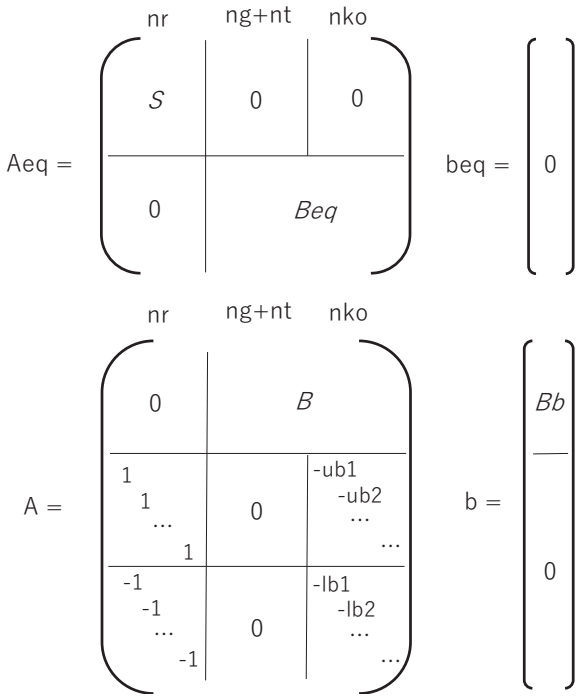


(A)

ID	Gene KO		v_1	v_2	v_3	v_4	v_5	v_6	v_7
1	none	best	10	5	5	5	0	10	5
2		worst	10	10	0	0	0	10	0
3	g1	best	5	0	0	0	5	0	5
4		worst	0	0	0	0	0	0	0
5	g2	best	5	0	0	0	5	0	5
6		worst	0	0	0	0	0	0	0
7	g3	best	10	0	5	5	5	5	10
8		worst	5	0	5	5	0	5	5
9	g4	best	10	5	5	5	0	10	5
10		worst	10	10	0	0	0	10	0
11	g5	both	10	10	0	0	0	10	0
12	g1, g2	best	5	0	0	0	5	0	5
13		worst	0	0	0	0	0	0	0
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(B)

Figure 2. (A) A toy example of the constraint-based model. Circles and rectangles represent metabolites and reactions, respectively. Black and white rectangles are external and internal reactions. r_1, r_6, and r_7 are the substrate uptake, cell growth, and target metabolite production reactions. $[\alpha, \beta]$ represents the lower and upper bounds of the reaction speeds. (B) The optimistic and pessimistic flux distributions from the viewpoints of PR for each gene deletion strategy when GR is maximized. Deleting g_3 achieves growth-coupled production since PR≥PRLB and GR≥GRLB are satisfied even for the pessimistic case of PR.



(A)

$$Beq = \begin{pmatrix} x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} \\ 0 & 0 & 1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

$$(Bb)^T = (0, 0, 0, 2, 0, 1, 0, 1)$$

(B)

Figure 3. (A) How to construct the components Aeq, beq, A, and b for the MILP formalization that gDel_minRN searches a gene deletion strategy candidate. (B) Beq, B, Bb for the example of the network of Fig. 2(A).

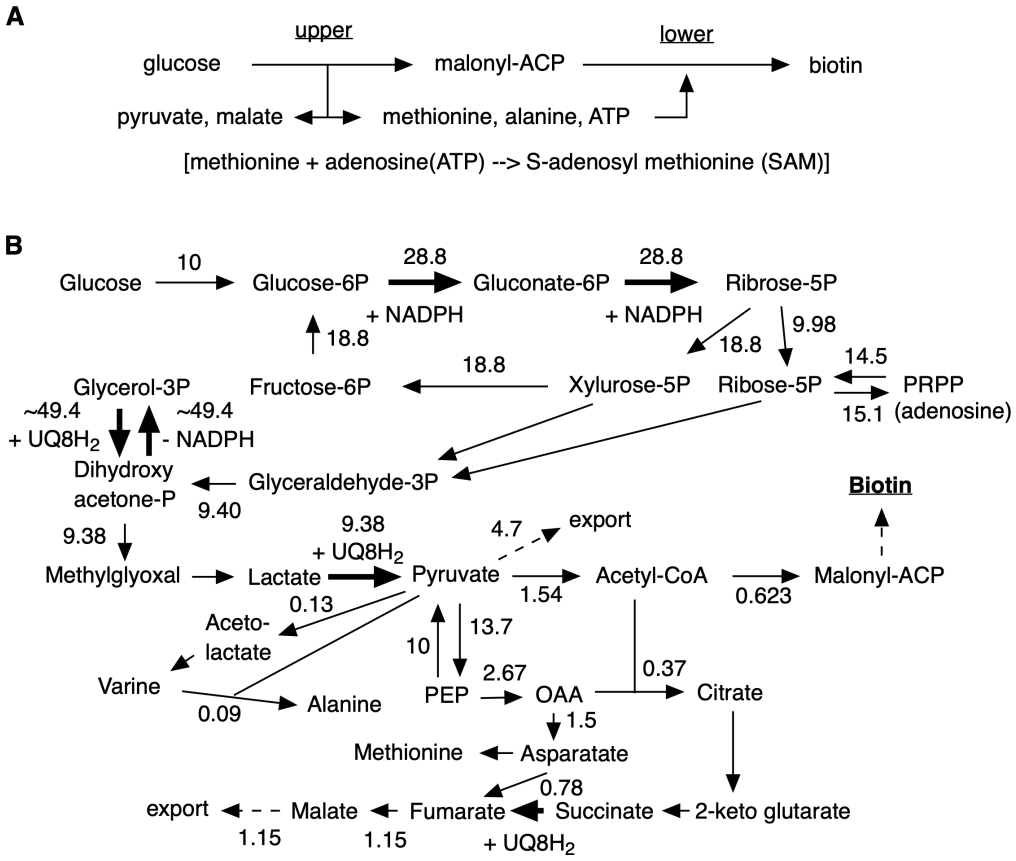


Figure 4. The constructed pathway for biotin production. (A) Overview of the biotin synthesis pathway from iML1515 classified into two pathways as upper and lower pathway. (B) Precise flow of upper pathway, from glucose to malonyl-ACP. The number indicated with each arrows shows the flux value of each reaction. The abbreviations are as follows; NADPH, Nicotinamide adenine dinucleotide phosphate reduced form; UQ8, ubiquinone-8; ACP, acyl carrier protein; PEP, phosphoenolpyruvate; OAA, oxaloacetate; PRPP, Phosphoribosyl diphosphate.