SUMMARY  Metabolic networks represent the relationship between chemical reactions and compounds in cells. In useful metabolite production using microorganisms, it is often required to calculate reaction deletion strategies from the original network to result in growth coupling, which means the target metabolite production and cell growth are simultaneously achieved. Although simple elementary flux mode (EFM)-based methods are useful for listing such reaction deletions strategies, the number of cases to be considered is often proportional to the exponential function of the size of the network. Therefore, it is desirable to develop methods of narrowing down the number of reaction deletion strategy candidates. In this study, the author introduces the idea of L1 norm minimal modes to consider metabolic flows whose L1 norms are minimal to satisfy certain criteria on growth and production, and developed a fast metabolic design listing algorithm based on it (minL1-FMDL), which works in polynomial time. Computational experiments were conducted for (1) a relatively small network to compare the performance of minL1-FMDL with that of the simple EFM-based method and (2) a genome-scale network to verify the scalability of minL1-FMDL. In the computational experiments, it was seen that the average value of the target metabolite production rates of minL1-FMDL was higher than that of the simple EFM-based method, and the computation time of minL1-FMDL was fast enough even for genome-scale networks. The developed software, minL1-FMDL, implemented in MATLAB, is available on https://sunflower.kuicr.kyoto-u.ac.jp/~tamura/software, and can be used for genome-scale metabolic network design for metabolite production.

key words: metabolic network, constraint-based model, elementary flux mode, linear programming, polynomial time algorithm

1. Introduction

Life is maintained by the right amount of necessary chemical reactions occurring at the right time in cells. A series of chemical reactions is called metabolism and can be represented by a metabolic network with chemical reactions and metabolites as nodes. Many chemical reactions in metabolism are catalyzed by proteins generated by metabolism-related genes. Therefore, the gene deletions can remove the corresponding reactions to control the metabolic network. Metabolic engineering controls microorganisms metabolism by modifying the original metabolic networks through reaction deletions to produce the desired useful metabolites efficiently. Since the reaction deletions by the gene deletions are costly in terms of expense, time, and effort, it is reasonable to find appropriate reaction deletion strategies through computer simulations instead of the real biophysical deletions. For examples, such strains for Escherichia coli are successful for the production of lactate [1], ethanol from a mixture of glucose and xylose [2], isobutanol [3], 1,4-butanediol [4], malonyl-CoA [5], fatty acids [6], and itaconic acid [7].

Although there are many mathematical models for metabolic networks, the constraint-based model [8] is one of the most popular for the simulation of metabolite production using microorganisms [9]. One reason is that a metabolic network in the constraint-based model can be formulated with a linear programming (LP) problem, so there are efficient algorithms even for large networks that work in polynomial time. This model analyzes the steady-state of metabolism by focusing on the reaction speed per time unit called flux. The constraint-based model imposes three types of constraints on flux, (1) each reaction must satisfy stoichiometry, in other words, the ratio in chemical reaction formula, (2) the sum of the incoming flux into each compound must equal to the sum of the outgoing flux, (3) Each flux must satisfy given upper and lower bounds. The objective function of the LP is to maximize cell growth that is represented by a special virtual reaction that was adjusted to meet the results of real cell behavior.

The problem of designing metabolic networks to produce a useful metabolite can be generalized as the problem of properly removing reactions from the larger metabolic network, including all candidate reactions as shown in Fig. 1 (A). In the constraint-based models, it is assumed that cell growth rate (GR) is maximized. And the target metabolite production rate (PR) is evaluated under the condition that GR is maximized. When PR>0 is achieved under the condition that GR is maximized, we say that growth coupling is achieved. Many methods have been proposed to calculate reaction deletion strategies that achieve growth coupling [10]–[19].

The constraint-based models are designed in detail based on the results of biological experiments, but they do not always match the behavior of actual cells. Therefore, in the simulation for the production of useful metabolites by a metabolic network, it is reasonable to calculate multiple reaction deletion strategies and selectively use them according to the actual situation.

For this purpose, one of the promising methods is to determine core or almost core parts to achieve growth coupling and delete the other parts, as shown in Fig. 1 (B). The elementary flux mode (EFM) [20]-based methods determine such core parts that consist of the minimal number of reactions and satisfy constraints for GR and PR. In other words,
The results of computer experiments showed that minL1-FMDL could compute several tens of reaction deletion strategies even for genome-scale networks, and the average value of the target production rates was higher than that of the simple EFM-based method. In the rest of this paper, the main problem is defined mathematically and illustrated with examples, and the proposed method minL1-FMDL is explained in Sect. 2. Section 3 describes the data and environment used in computer experiments and the detailed results of the computational experiments using FMDL and the simple EFM-based methods. Section 4 evaluates minL1-FMDL based on the results of the computer experiments and discusses the meaning of this study, and Sect. 5 concludes.

2. Methods

2.1 Definition

Let $M = (M, R, S, L, U)$ be a metabolic network represented in a constraint-based model, where $M = \{m_1, \ldots, m_a\}$ and $R = \{r_1, \ldots, r_b\}$ are sets of metabolites and reactions, 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 the $r_j$ produces $k$ of $m_i$ per unit time. If $k$ is a negative number, then $m_i$ is consumed. Let $V = \{v_1, \ldots, v_b\}$ is a set of reaction speeds per unit (flux) of $R$. Let $L = \{l_1, \ldots, l_b\}$ and $U = \{u_1, \ldots, u_b\}$ be sets of the lower and upper bounds for $V$, respectively. $v_{growth}$ is called growth rate (GR). GR is maximized by the following LP.

**maximize**

$$v_{growth}$$

**such that**

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

$$l_j \leq v_j \leq u_j \text{ for all } j$$

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

If $i$th column of $S$ has only one non-zero element, in other word, $r_i$ connects to only one metabolite, then $r_i$ is called an exchange reaction, and considered as connected to external environment. Reactions that are not exchange reactions are called internal reactions. The flux of the exchange reaction producing the target metabolite under the condition that cell growth is maximized is called production rate (PR). In the production of useful metabolites by the constraint-based model, the goal is finding appropriate reaction deletion strategies to make PR exceed the certain criteria. If the target metabolite is not connected to an exchange reaction, an auxiliary exchange reaction is added, and the growth coupling is evaluated by GR and the outgoing flux from the additional exchange reaction, which is also called PR.

Since LP may have multiple solutions, there is often multiple possible PR when GR is maximized. Worst-case scenarios of PR need to be evaluated in order to guarantee PR for the reaction deletion strategy. Such an analysis is
called flux variability analysis (FVA) [22]. Therefore, the main problem is formalized as follows.

### Main problem
Reaction deletion strategy listing problem for constraint-based models for target metabolite production

**Given**

\( M, r_{\text{target}}, PR_{\text{threshold}}, \text{count} \)

**Find**

\( D_{\text{set}} \text{ where } \left| D_{\text{set}} \right| \geq \text{count}, D \in D_{\text{set}}, D \subset R, \text{ and } v_{\text{target}} \geq PR_{\text{threshold}} \)

such that minimizes

\( v_{\text{growth}} \)

such that

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

\[ v_j = 0 \text{ if } r_j \in D \]

\[ l_j \leq v_j \leq u_j, \text{ otherwise} \]

\[ i = 1, \ldots, a, j = 1, \ldots, b \]

#### 2.2 Example

Figure 2 shows an example for a constraint-based model. \( R = \{r_1, \ldots, r_9\} \) is a set of reactions, where exchange and internal reactions are represented by black and white rectangles, respectively. \( M = \{m_1, \ldots, m_7\} \) is a set of metabolites represented by circles. Suppose that chemical equations are given as follows.

\( r_1: \text{ the exchange reaction of } m_1 \text{ to uptake nutrients such as glucose} \)

\( r_2: m_1 \rightarrow m_2 \)

\( r_3: m_1 \rightarrow m_3 \)

\( r_4: m_1 \rightarrow m_4 \)

\( r_5: m_1 \rightarrow m_5 \)

\( r_6: m_2 + m_3 \rightarrow 2m_6 \)

\( r_7: m_4 + m_5 \rightarrow m_6 + m_7 \)

\( r_8: \text{ the exchange reaction of } m_9 \text{ representing the cell growth} \)

\( r_9: \text{ the exchange reaction of } m_7 \text{ that is the target metabolite} \)

Let \( r_8 \) be called the growth reaction that represents the cell growth. Since the growth flux is maximized in FBA, \( v_8 \) is maximized in the LP representing this metabolic network.

![Fig. 2](image)

Suppose that each flux is bounded by as follows,

\[ l_1 = 0, u_1 = 10 \rightarrow 0 \leq v_1 \leq 10, \]

\[ l_2 = 0, u_2 = 10 \rightarrow 0 \leq v_2 \leq 10, \]

\[ l_3 = 0, u_3 = 10 \rightarrow 0 \leq v_3 \leq 10, \]

\[ l_4 = 0, u_4 = 10 \rightarrow 0 \leq v_4 \leq 10, \]

\[ l_5 = 0, u_5 = 5 \rightarrow 0 \leq v_5 \leq 5, \]

\[ l_6 = 0, u_6 = 5 \rightarrow 0 \leq v_6 \leq 5, \]

\[ l_7 = 0, u_7 = 10 \rightarrow 0 \leq v_7 \leq 10, \]

\[ l_8 = 0, u_8 = 10 \rightarrow 0 \leq v_8 \leq 10, \]

\[ l_9 = 0, u_9 = 10 \rightarrow 0 \leq v_9 \leq 10, \]

where \( L = \{l_1, \ldots, l_9\} \) and \( U = \{u_1, \ldots, u_9\} \). Note that every reaction is irreversible in this example since any \( l_i \) is not negative.

Then, the FBA for the metabolic network of Fig. 2 is formalized as follows.

**maximize**

\[ v_8 \text{ /*maximize cell growth*/} \]

such that

\[ v_1 - v_2 - v_3 - v_4 - v_5 = 0 \text{ /*the constraint for } m_1/* \]

\[ v_2 - v_6 = 0 \text{ /*the constraint for } m_2/* \]

\[ v_3 - v_6 = 0 \text{ /*the constraint for } m_3/* \]

\[ v_4 - v_7 = 0 \text{ /*the constraint for } m_4/* \]

\[ v_5 - v_7 = 0 \text{ /*the constraint for } m_5/* \]

\[ 2v_6 + v_7 - v_8 = 0 \text{ /*the constraint for } m_6/* \]

\[ v_7 - v_9 = 0 \text{ /*the constraint for } m_7/* \]

\[ 0 \leq v_1 \leq 10 \text{ /*the bounds for } r_1/* \]

\[ 0 \leq v_2 \leq 10 \text{ /*the bounds for } r_2/* \]

\[ 0 \leq v_3 \leq 10 \text{ /*the bounds for } r_3/* \]

\[ 0 \leq v_4 \leq 10 \text{ /*the bounds for } r_4/* \]

\[ 0 \leq v_5 \leq 5 \text{ /*the bounds for } r_5/* \]

\[ 0 \leq v_6 \leq 5 \text{ /*the bounds for } r_6/* \]

\[ 0 \leq v_7 \leq 10 \text{ /*the bounds for } r_7/* \]

\[ 0 \leq v_8 \leq 10 \text{ /*the bounds for } r_8/* \]

\[ 0 \leq v_9 \leq 10 \text{ /*the bounds for } r_9/* \]

Suppose that \( PR_{\text{threshold}} \) and \( \text{count} \) of the main problem are 2 and 1, respectively. Since the solution of this LP is uniquely determined as \( (v_1, v_2, \ldots, v_9) = (10, 5, 5, 0, 0, 5, 0, 10, 0) \), GR and PR become 10 and 0, respectively. In this case, PR does not exceed \( PR_{\text{threshold}} \). However, if \( r_6 \) is deleted, \( (v_1, v_2, \ldots, v_9) = (10, 0, 0, 5, 5, 0, 5, 5, 5) \) is uniquely obtained since \( l_6 \) and \( u_6 \) become 0. The same result is obtained when \( r_2 \) or \( r_3 \) is deleted. Therefore, \( D_{\text{set}} = \{D_1, \ldots, D_7\} \) where \( D_1 = \{r_2\}, D_2 = \{r_3\}, D_3 = \{r_6\}, D_4 = \{r_2, r_3\}, D_5 = \{r_3, r_6\}, D_6 = \{r_2, r_6\}, \) and \( D_7 = \{r_2, r_3, r_6\} \) is a solution of the main problem since GR and PR are 5 and 5, respectively, even in the worst case.

To extract the core part more effectively, if \( D' \in D_{\text{set}} \), \( D'' \in D_{\text{set}} \), and \( D' \subseteq D'' \) hold, then minL1-FMDL excludes \( D' \) from \( D_{\text{set}} \). For the example of Fig. 2, minL1-FMDL outputs only \( \{r_2, r_3, r_6\} \).
2.3 Algorithm

The minL1-FMDL algorithm is based on the solution space division-based method developed in [18]. Suppose that the $x$ and $y$ axes of the solution space represent GR and PR, respectively. Let $TMGR$ and $TMPR$ represent the theoretical maximum values of GR and PR. Then, $x = [0, TMGR]$ and $y = [0, TMPR]$ can represent a rectangle representing the solution space of the constraint-based model. By dividing each axis into $P$ equivalents, we can divide the entire solution space into $P^2$ small sub-solution spaces. It is known that effective reaction deletion strategies can be obtained with a high success rate by solving LP for each small rectangle and deleting reactions whose fluxes are zero in the solution of the LP [18].

On the other hand, this method yields a variety of effective reaction deletion strategies for each small sub-solution space. Existing solution space division-based methods such as GridProd select the reaction deletion strategies with the highest PR, and discard the other candidates. However, minL1-FMDL utilizes the other candidates as well and refines all candidates so that any reaction deletion strategy is not a subset of another strategy.

Figure 3 (A) is an example of the constraint-based model for illustrating the behavior of minL1-FMDL. $M = \{m_1, \ldots, m_6\}$ and $R = \{r_1, \ldots, r_7\}$ are sets of metabolites and reactions, respectively. Let $r_6$ and $r_7$ be the reactions for cell growth and the target metabolite production. $[a, b]$ represents $l$ and $u$; in other words, the lower and upper bounds of the corresponding fluxes. In the original state, when $v_6$ is maximized, the best and worst PR is 8 and 0, respectively, since $(r_1, \ldots, r_6) = (10, 2, 5, 3, 10, 8)$ and $(r_1, \ldots, r_6) = (10, 10, 0, 0, 0, 0, 10, 0)$ are obtained as shown in the first and second rows in Fig. 3 (B). The fluxes for the best and worst PR for each reaction deletion strategy are shown in Fig. 3 (B). If $PR_{\text{threshold}}$ and $count$ are 2 and 2, then $D_{\text{set}} = D_1, D_2$ where $D_1 = \{r_2, r_3\}$ and $D_2 = \{r_2, r_4, r_5\}$ is the solution. Note that $D' = \{r_2\}$ is not included in $D_{\text{set}}$ by minL1-FMDL since $D' \subseteq D_1$ and $D' \subseteq D_2$ hold. If $PR_{\text{threshold}}$ and $count$ are 4 and 2, then there is no solution since $D$ whose worst-case PR is more than or equal to 4 is only $\{r_2, r_3, r_5\}$.

minL1-FMDL first computes $TMGR$ and $TMPR$. These values are 10 and 8, respectively, in this example. Since the theoretical minimum values are 0, the solution space can be represented by $GR=[0,10]$ and $PR=[0,8]$. According to the designated value $P$, minL1-FMDL divides the whole solution space into $P^2$ pieces of the smaller solution spaces.

If 0 is allowed for GR and PR constraints, 0 is assigned to all fluxes in many cases. Therefore, in minL1-FMDL, certain lower bounds are applied to the original solution space. Let such lower bounds be 1 for both GR and PR. Then, the modified whole solution space is $GR=[1,10]$ and $PR=[1,8]$. Suppose that $P = 3$ is given. Then the ranges $1 \leq GR \leq 10$ and $1 \leq PR \leq 8$ are divided into $[1 \leq GR \leq 4, 4 \leq GR \leq 7, 7 \leq GR \leq 10]$ and $[1 \leq PR \leq 3.33, 3.33 \leq PR \leq 5.57, 5.57 \leq PR \leq 8]$, respectively.

By considering all these combinations, we can obtain nine small solution spaces since $P^2 = 9$ as shown in Table 1. In each small solution space, the two corresponding constraints are added, and LP is conducted to minimize the absolute sum of fluxes. Then, the reactions that are assigned 0 in the solution of the LP compose the reaction deletion strategies. For example, as shown in Table 1, when the constraints are $1 \leq GR \leq 4$ and $1 \leq PR \leq 3.33$, the LP outputs $(v_1, \ldots, v_7) = (1, 0, 1, 0, 0, 1, 1)$. Then, $\{r_2, r_4, r_5\}$ are selected for the reaction deletion strategies. When, $\{r_2, r_4, r_5\}$ are deleted, the worst case PR is 5 since $(v_1, \ldots, v_7) = (5, 0, 5, 0, 0, 5, 5)$ is obtained. The obtained flux set, its corresponding reaction deletion strategy, the consequent fluxes for each of the nine cases are summarized in Table 1. Then, $D_{\text{set}} = \{r_2, r_2, r_4, r_5\}$ is obtained. Since $\{r_2\}$ is a subset of $\{r_2, r_4, r_5\}$, it is discarded. Finally, $\{r_2, r_2, r_5\}$ is obtained, and the worst-case PR is 5.

The pseudo code of minL1-FMDL is shown below.

![minL1-FMDL](image)

Fig. 3: Illustration of minL1-FMDL. (A) an example of the constraint-based model, where $[a, b]$ means the lower and upper bounds of each flux. (B) Fluxes corresponding to the best and worst case of PR for each reaction deletion strategy under the condition that GR is maximized.

**Procedure** minL1-FMDL($M, r_{\text{target}}, PR_{\text{threshold}}, P$)

\[
TMGR = \max \ v_{\text{growth}}
\]

s.t. $\sum_j S_{ij} \cdot v_j = 0$ for all $1 \leq i \leq a$

$LB_i \leq v_j \leq UB_j$ for all $1 \leq j \leq b$

\[
TMPR = \max \ v_{\text{target}}
\]

s.t. $\sum_j S_{ij} \cdot v_j = 0$ for all $1 \leq i \leq a$

$LB_j \leq v_j \leq UB_j$ for all $1 \leq j \leq b$

$D_{\text{set}} = \emptyset$, $x = 1$

for $i = 1$ to $P$ do
Note that $\epsilon$ is the lower bounds for GR and PR constraints, and $v_{\text{growth}}^\text{min}$ is the minimum required GR in the target production.

2.4 Simple Elementary Flux Mode-Based Method

Elementary flux modes (EFMs) are non-zero flux (reaction speed) distributions that is minimal for the number of reactions. In other words, a reaction deletion strategy $D$ yields an EFM if $D^* = D \cup r_i$ for any $x$ results in $v_x = 0$ for any $y$. For example, in Fig. 3 (A), there are three EFMs that are represented in Table 2.

### Table 2

The elementary flux modes (EFMs) for the constraint-based model of Fig. 3 (A). Since EFM2 and EFM3 satisfy GR>0 and PR>0, the simple EFM-based method finds two reaction deletion strategies, $(v_2, v_4, v_5)$ and $(v_2, v_3)$, that achieve growth coupling.

| GR constraints | PR constraints | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ |
|---------------|---------------|------|------|------|------|------|------|------|
| $1 \leq \text{GR} \leq 4$ | $1 \leq \text{PR} \leq 3.33$ | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| $1 \leq \text{GR} \leq 4$ | $3.33 \leq \text{PR} \leq 5.67$ | 5 | 0 | 5 | 0 | 0 | 5 | 5 |
| $1 \leq \text{GR} \leq 4$ | $5.67 \leq \text{PR} \leq 8$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $4 \leq \text{GR} \leq 7$ | $1 \leq \text{PR} \leq 3.33$ | 4 | 3 | 1 | 0 | 0 | 4 | 1 |
| $4 \leq \text{GR} \leq 7$ | $3.33 \leq \text{PR} \leq 5.67$ | 10 | 10 | 0 | 0 | 0 | 10 | 0 |
| $4 \leq \text{GR} \leq 7$ | $5.67 \leq \text{PR} \leq 8$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $7 \leq \text{GR} \leq 10$ | $1 \leq \text{PR} \leq 3.33$ | 7 | 6 | 1 | 0 | 0 | 7 | 1 |
| $7 \leq \text{GR} \leq 10$ | $3.33 \leq \text{PR} \leq 5.67$ | 10 | 10 | 0 | 0 | 0 | 10 | 0 |
| $7 \leq \text{GR} \leq 10$ | $5.67 \leq \text{PR} \leq 8$ | 7 | 3.66 | 3.33 | 0 | 0 | 7 | 3.33 |
| $7 \leq \text{GR} \leq 10$ | $8 \leq \text{PR} \leq 10$ | 10 | 10 | 0 | 0 | 0 | 10 | 0 |

Computational experiments were conducted for two constraint-based models of $E. coli$. iML1515 [23] is one of the most recent genome-scale models that contains 2712 reactions and 1877 metabolites, while $e_{coli}\_core$ [24] contains the only essential part of metabolism of $E. coli$ (See also Table 3). If the target metabolite does not have an exchange reaction to the external environment, an auxiliary exchange reaction was temporarily added to the model to
simulate the target metabolite production as in existing studies for metabolite production [17]–[19]. Because some exchange reactions, including these additional reactions, never excrete, PR of some target metabolites is zero, even if it is maximized. Such metabolites are excluded from the target metabolites. Since TMPRs were 0 for such exchange reactions for 20 and 871 metabolites of e.coli_core and iML1515, respectively, the number of target metabolites in this study were 52 and 1006.

The values of $PR_{\text{threshold}}$, $v_{\text{min}}$, and $\epsilon$ were set to 0.00001 for all computational experiments. The reason why $PR_{\text{threshold}}$ was set to almost 0 was in order not to discard any solutions to evaluate the detailed performance of each method. All procedures in the computational experiments were implemented on a CentOS 7 machine with an Intel Xeon Processor with 2.30 GHz 18C/36T, and 128 GB memory. This workstation had CPLEX, COBRA Toolbox [9], and MATLAB.

In the first experiments, minL1-FMDL for $P=2$, 5, 10, and 15 and the simple EFM-based method were applied for the 52 target metabolites of e.coli_core. The results are summarized in Table 4. The simple EFM-based method could complete the computation, obtained 3998 reaction deletion strategies, and the elapsed time was 24.05s. The computation time of minL1-FMDL was between 1.854s and 7.52s, however, the average numbers of obtained reaction deletion strategies were at most 7.52. Therefore, if the simple EFM-based method can complete the calculation, it has the ability to yield an overwhelming number of solutions. The average GR/TMGR were 27.82% for the simple EFM-based method, and between 50% and 55% for minL1-FMDL. It was seen that the average minPR/TMPR of minL1-FMDL were higher than those of the simple EFM-based method.

For each obtained reaction deletion strategies, the worst-case PR/TMPR under the condition that GR is maximized were evaluated. The average minPR/TMPR of the simple EFM-based method was 21.41%, while the values were between 45% and 50% for minL1-FMDL.

For $P=2$ of minL1-FMDL, no reaction deletion strategies were obtained for four target metabolites (fail ratio=4/52=0.0769). For $P=5$, 10, and 15 of minL1-FMDL and the simple EFM-based method, no reaction deletion strategies were obtained for two target metabolites (fail ratio=2/52=0.0385).

For iML1515, the simple EFM-based method could not complete the calculation due to out of memory. The results by minL1-FMDL were summarized in Table 5. The elapsed time for $P=2$ and $P=15$ was 11.25s and 222.12s. The increase in computation time associated with the increase in $P$ was higher than the case for e.coli_core. It seemed possible to apply larger $P$ until as long as the computation time is acceptable. The average number of the obtained reaction deletion strategies was 1.024 for $P=2$ and 20.88 for $P=15$. The same tendency of large increases in values with increasing $P$ was observed for minPR/TMPR as well. However, such a tendency was not observed for GR/TMGR. For the fail ratio, the opposite tendency was clearly observed. It was 0.7594 for $P=2$, but it was decreased to 0.1123 for $P=15$.

The distribution of PR and GR were investigated for e_ecoli_core for the case where the target metabolite was pyruvate since pyruvate is a vital metabolite but not produced in the natural state in either e.coli_core or iML1515. Figure 4 represents the distribution of GR and PR yielded by the reaction deletion strategies obtained by the simple EFM-

### Table 3

| Model     | #Reactions | #Metabolites | #Targets |
|-----------|------------|--------------|----------|
| e.coli_core | 95         | 72           | 52       |
| iML1515   | 2712       | 1877         | 1006     |

### Table 4

|                          | $P=2$ | $P=5$ | $P=10$ | $P=15$ | Simple EFM |
|--------------------------|-------|-------|--------|--------|------------|
| Elapsed time             | 0.413s| 2.827s| 6.764s | 9.819s | 24.05s     |
| Avg #obtained designs    | 1.854 | 4.22  | 6.42   | 7.52   | 3998       |
| Avg minPR/TMPR           | 47.87%| 47.63%| 46.78% | 46.84% | 21.41%     |
| Avg GR/TMGR              | 51.95%| 53.41%| 54.35% | 54.17% | 27.82%     |
| Fail ratio               | 7.69% | 3.85% | 3.85%  | 3.85%  | 3.85%      |

### Table 5

|                          | $P=2$ | $P=5$ | $P=10$ | $P=15$ | Simple EFM |
|--------------------------|-------|-------|--------|--------|------------|
| Elapsed time             | 11.25s| 33.03s| 104.08s| 222.12s| NA         |
| Avg #obtained designs    | 1.024 | 4.099 | 11.334 | 20.88  | NA         |
| Avg minPR/TMPR           | 9.12% | 25.67%| 38.27% | 42.09% | NA         |
| Avg GR/TMGR              | 63.40%| 63.35%| 54.62% | 52.29% | NA         |
| Fail ratio               | 75.94%| 30.62%| 15.51% | 11.23% | NA         |
Fig. 4 The distribution of GR and PR for the reaction deletion strategies for ecoli_core obtained by the simple EFM-based method when the target metabolite was pyruvate and GR is maximized.

Fig. 5 The distribution of GR and PR for the reaction deletion strategies for ecoli_core obtained by minL1-FMDL for P=2, 5, 10, and 15 when the target metabolite was pyruvate and GR is maximized. Each elapsed time is described in parentheses.

Fig. 6 The distribution of GR and PR for the reaction deletion strategies for iML1515 obtained by minL1-FMDL for P=5, 10, 20, and 30, and the target metabolite was pyruvate. Different from the cases for ecoli_core, it was seen that the number of obtained reaction deletion strategies increased as P increased.

4. Discussion

For computing reaction deletion strategies for growth coupling, many methods, including based on MILP, local search, and genetic algorithms, are known. Among them, this study focused on the methods that determine the core or almost core parts and then deleting reactions included in the other parts, as shown in Fig. 1 (A). For this purpose, the developed method minL1-FMDL was mainly compared with the simple EFM-based method to evaluate the performance in this study.

Since the simple EFM-based method determines the core part that is minimal for the number of reactions under the condition that GR > 0 and PR > 0, it can be considered as an L0 norm minimal mode-based method. On the other hand, since minL1-FMDL determines the almost core part where the total sum of the absolute values of fluxes is minimum under the condition that PR and GR range constraints are given, it can be considered as an L1 norm minimal mode-based method.

While the L0 norm minimal mode-based method minimizes the number of used reactions, the L1 norm minimal mode-based method minimizes the sum of absolute values of fluxes. In addition to this difference, minL1-FMDL uses the constraints for GR and PR that are divided into smaller pieces. Although GridProd[18] is also based on the L1 norm minimal mode-based method, it was designed for finding a single solution.

In the performance comparison between minL1-FMDL and the simple EFM-based method, for small networks with less than 100 reactions, the simple EFM-based method was superior because all EFMs can be enumerated quickly. However, for a genome-scale network of about 3,000 reactions, the computer used in this study could not complete the calculation due to out of memory. Since the number of EFMs increases exponentially with the number of reactions, it is unlikely that all EFMs can be enumerated for a genome-scale network, even with increased computing resources.

While the number of reaction deletion strategies obtained by minL1-FMDL is much less than that obtained by the simple EFM-based method for small networks, it increases as P increases and tends to stop increasing once P reaches a certain value. Users who give importance to
fast computation time should use smaller \( P \), and users who give importance to a large number of solutions should use larger \( P \). If the network is the same, the approximate computation time increases in proportion to the square of \( P \). Since minL1-FMDL can fastly obtain dozens of reaction deletion strategies that achieve growth coupling even for genome-scale networks, the L1 norm minimal mode-based method can be considered effective from the viewpoint of balancing computation time and the number of obtained solutions. However, it is not guaranteed that minL1-FMDL can find a solution. If the goal is to find a single solution, the experiments described in Table 5 of [18] have shown that \( P=25 \) was the most effective for a genome-scale network iAF1260 [25]. However, it should be noted that the optimal value of \( P \) varies depending on the networks.

Instead of the simple EFM-based method, it may be possible to add some constraints to the search for EFMs to narrow the scope and reduce the number of candidate EFMs. Such constrained EFMs may be obtained using MILP or successive LPs. Since MILP is NP-complete, it falls outside the purpose of this study to develop a polynomial-time algorithm. Solving successive LPs may be promising for listing the effective reaction deletion strategies for metabolite production, however, non-straightforward extensions may be necessary to consider cell growth maximization.

Another possible method is to extend MILP-based methods, such as OptKnock [10] and RobustKnock [26]. For example, the fail ratio of OptKnock for e.coli_core was 0.0577, and the elapsed time was 1.05s. However, Opt-Knock could not determine the appropriate reaction deletion strategies for any target metabolite of iML1515. GDLS [11], which is a local search-based method, also failed to determine reaction deletion strategies within one hour for any target metabolite of iML1515. Since many existing algorithms could not complete the calculation for listing reaction deletion strategies for growth coupling, it was important to develop a polynomial-time algorithm that is fast even for genome-scale networks.

In this study, the worst-case PR was evaluated, and minL1-FMDL is described to find only reaction deletion strategies where \( PR \geq PR_{\text{threshold}} \) holds. However, in the computational experiments, \( PR_{\text{threshold}} \) was set to 0.00001 to evaluate the detailed ability to list reaction deletion strategies.

The author recently developed CubeProd [27], which is an extension of GridProd by adding another solution space axis based on the absolute sum of fluxes. Since CubeProd minimizes the absolute sum of fluxes while the lower bounds are imposed for the absolute sum of fluxes, CubeProd is not based on L1 minimal modes. Therefore, the performance of minL1-FMDL was not compared with that of CubeProd in this study.

Klamt et al. was extended the notion of EFMs to elementary flux vectors (EFVs) to consider the upper and lower bounds of fluxes [28]. The number of EFVs may decrease from the number of EFMs due to the additional constraints, but the number of EFVs is still proportional to the exponential function of the number of reactions. Therefore, the performance of the EFV-based method when compared to minL1-FMDL is expected to be not fundamentally different from that of the simple EFM-based method. Namely, the performance of the EFV-based method is expected to be superior for the small network, while minL1-FMDL is expected to be superior for genome-scale networks.

The two-step fermentation-based approach is also popular in metabolic engineering. In this approach, the cells are initially cultured under growth conditions and then shifted to the production phase by preventing cell growth by some method such as nitrogen starvation [29]. Therefore, small PR with large GR and large PR with zero GR are required.

For this problem setting, Toya et al. developed SSDesign that determines gene deletion strategies that make the solution space in the desired shape described above using EFMs as a navigator [29]. SSDesign can design strains that have the appropriate solution space shape for the two-step fermentation approach. Because minL1-FMDL evaluates PR when GR is maximized, and the problem setting is different from that of SSDesign, the performance of minL1-FMDL was not compared with SSDesign in this study.

5. Conclusion

In this study, the author developed minL1-FMDL, which calculates multiple reaction deletion strategies to achieve growth coupling in polynomial time. minL1-FMDL divides the solution space where the \( x \) and \( y \) axes are cell growth and target production constraints and determines the L1 minimal mode in each sub-solution space to determine the core part for growth coupling. Reactions that are outside the core part are included in the reaction deletion strategies. Although the simple EFM-based method, which can be considered as L0 norm minimal mode-based method, can find an overwhelming number of reaction deletion strategies for small networks, it is difficult to complete the calculation for genome-scale networks. Although it may be possible to speed up the simple EFM-based method by narrowing down the scope of candidate reaction deletion strategies, such extension is not straightforward since cell growth maximization must be considered. Since minL1-FMDL works in polynomial time for the number of reactions, it can be considered as an effective alternative method for listing reaction deletion strategies for growth coupling to balance the computation time and the number of obtained solutions.

Acknowledgements

The author was partially supported by grants from JSPS, KAKENHI #20H04242.

References

[1] S.S. Fong, A.P. Burgard, C.D. Herring, E.M. Knight, F.R. Blattner, C.D. Maranas, and B.O. Palsson, “In silico design and adaptive evolution of escherichia coli for production of lactic acid,” Biotechnology and bioengineering, vol.91, no.5, pp.643–648, 2005.
TAMURA: L1 NORM MINIMAL MODE-BASED METHODS FOR LISTING REACTION NETWORK DESIGNS FOR METABOLITE PRODUCTION

[1] C.T. Trinh, P. Unrean, and F. Szience, “Minimal escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses,” Applied and environmental microbiology, vol.74, no.12, pp.3634–3643, 2008.

[2] C.T. Trinh, J. Li, H.W. Blanch, and D.S. Clark, “Redesigning escherichia coli metabolism for anaerobic production of isobutanol,” Applied and environmental microbiology, vol.77, no.14, pp.4894–4904, 2011.

[3] H. Yim, R. Haselbeck, W. Niu, C. Pujol-Baxley, A. Burgard, J. Boldt, J. Khandurina, J.D. Trwack, R.E. Osterhout, R. Stephen, J. Estadilla, T. Seisam, H.B. Schreyer, S. Andrae, T.H. Yang, S.Y. Lee, M.J. Burk, and S. Van Dien, “Metabolic engineering of escherichia coli for direct production of 1,4-butanediol,” Nature chemical biology, vol.7, no.7, pp.445–452, 2011.

[4] P. Xu, S. Ranganathan, Z.L. Fowler, C.D. Maranas, and M.A. Koffas, “Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-
coa,” Metabolic engineering, vol.13, no.5, pp.578–587, 2011.

[5] S. Ranganathan, T.W. See, A. Chowdhury, A.R. Zomorrodi, J.M. Yoon, Y. Tu, J.V. Shanks, and C.D. Maranas, “An integrated computational and experimental study for overproducing fatty acids in escherichia coli,” Metabolic engineering, vol.14, no.6, pp.687–704, 2012.

[6] B.-J. Harder, K. Bettenbrock, and S. Klamt, “Model-based metabolic engineering enables high yield itaconic acid production by escherichia coli,” Metabolic engineering, vol.38, pp.29–37, 2016.

[7] N. Tepper and T. Shlomi, “Predicting metabolic engineering knockout strategies for metabolic engineering,” Bioinformatics, vol.13, no.4, pp.264–276, 2003.

[8] J.M. Monk, C.J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, A.M. Feist, and B.O. Palsson, “imi1515, a knowledgebase that computes escherichia coli traits,” Nature biotechnology, vol.35, no.10, pp.904–908, 2017.

[9] J.D. Orth, B.O. Palsson, and R.M.T. Fleming, “Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide,” EcoSal plus, 2010.

[10] P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.O. Palsson, “A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information,” Molecular systems biology, vol.3, no.1, p.121, 2007.

[11] N. Tepper and T. Shlomi, “Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways,” Bioinformatics, vol.26, no.4, pp.536–543, 2010.

[12] T. Tamura, “Efficient reaction deletion algorithms for redesign of constraint-based metabolic networks for metabolic production with weak coupling,” bioRxiv, p.563437, 2020.

[13] S. Klamt, G. Regensburger, M.P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, S. Müller, and C. Kaleta, “From elementary flux modes to elementary flux vectors. Metabolic pathway analysis with arbitrary linear flux constraints,” PLoS computational biology, vol.13, no.4, 2017.

[14] T. Yota, T. Shiraki, and H. Shimizu, “Sdesign: Computational metabolic pathway design based on flux variability using elementary flux modes,” Biotechnology and bioengineering, vol.112, no.4, pp.759–768, 2015.

Takeyuki Tamura received B.E., M.E. and Ph.D. 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 April, 2006. He worked as an assistant professor from December 2007 to September 2017, and works as an associate professor from October 2017. His research interests are bioinformatics and the theory of combinatorial optimization for graphs and networks.