Title: Computing Smallest Intervention Strategies for Multiple Metabolic Networks in a Boolean Model

Running heads: Minimum Knockout for Multiple Boolean Metabolic Network

Author (1): Wei Lu (joint first author)

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 6110011, Japan.

Tel: +81-774-38-3018, Fax: +81-774-38-3022, e-mail: rogi@kuicr.kyoto-u.ac.jp

Author (2): Takeyuki Tamura (joint first author) *

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 6110011, Japan.

Tel: +81-774-38-3020, Fax: +81-774-38-3022, e-mail: tamura@kuicr.kyoto-u.ac.jp

Author (3): Jiangning Song

Department of Biochemistry and Molecular Biology, Monash University, McMahons Road, Frankston VIC 3199, Melbourne, Australia,
and
National Engineering Laboratory for Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China.

Tel: +61-3-990-29304, email: Jiangning.Song@med.monash.edu.au

Author (4): Tatsuya Akutsu *

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 6110011, Japan.

Tel: +81-774-38-3015, Fax: +81-774-38-3022, e-mail: takutsu@kuicr.kyoto-u.ac.jp

Key words: algorithm, integer linear programming, metabolic network, Boolean model, elementary mode, NP-complete

* Corresponding Authors

Abstract

This paper considers the problem whereby, given two metabolic networks N_1 and N_2, a set of source compounds, and a set of target compounds, we must find the minimum set of reactions whose removal (knockout) ensures that the target compounds are not producible in N_1, but are producible in N_2. Similar studies exist for the problem of finding the minimum knockout with the smallest side-effect for a single network. However, if technologies of external perturbations are advanced in near future, it may be important to develop methods of computing the minimum knockout for multiple networks (MKMN). Flux balance analysis (FBA) is efficient if a well-polished model is available. However, that is not always the case. Therefore, in this paper, we study MKMN in Boolean models and an elementary mode (EM)-based model. Integer linear programming (ILP)-based methods are developed for these models, since MKMN is NP-complete for both the Boolean model and the EM-based model. Computer experiments are conducted with metabolic networks of *clostridium perfringens SM101* and *bifidobacterium longum DJO10A* respectively known as bad bacteria and good bacteria for human intestine. The results show that larger networks are more likely to have MKMN solutions. However, solving for these larger networks takes a very long time, and often the computation cannot be completed. This is reasonable, because small networks do not have many alternative pathways, making it difficult to satisfy the MKMN condition whereas in large networks the number of candidate solutions explodes. Our developed software is available at "http://sunflower.kuicr.kyoto-u.ac.jp/~rogi/minFvsKO/minFvsKO.html".

1 Introduction

Metabolic networks represent relations between biochemical reactions and metabolites in living cells. The removal (or knockout) of metabolism-related genes is often simulated in metabolic networks, as the perturbation of genes frequently corresponds to the inhibition of certain reactions in metabolic networks.

Many types of mathematical models have been developed for this purpose. For small size networks, models using the ordinary differential equations (ODEs) are often used. Although ODEs have a detailed explanatory power, their applicability is limited by the difficulty of obtaining the necessary kinetic parameters, and their limited scalability. On the other hand, less detailed approaches like Boolean models and constraint-based models have been used in larger networks (Gonçalves et al., 2013).

Flux balance analysis (FBA) (Raman and Chandra, 2009; Varma and Palsson, 1994) is a constraint-based mathematical model of metabolic networks in which the stoichiometry and the biomass objective functions are used to predict cell growth rates (Kauffman et al., 2003). Although the standard FBA simply maximizes the biomass objective function, the minimization of metabolic adjustment method (MOMA) (Segre et al., 2002) seeks to minimize the difference between the wild and the knocked-out flows. Flux variability analysis (FVA) assesses the range of this difference (Shlomi et al., 2009), and the optimal knockout strategies have been developed based on such flow models. Optknock (Burgard et al., 2003) determines which reactions should be knocked-out to maximize the biomass objective function. Using bilevel programming, the biomass objective function is first maximized for each knockout(s), and then a reaction set is chosen from the resulting candidates. In contrast, although RobustKnock (Tepper and Shlomi, 2010) is also based on bilevel programming, it maximizes the minimized biomass objective functions by selecting different reactions. Optorf (Kim and Reed, 2010) integrates transcriptional regulatory networks and metabolic networks. The above-mentioned flow models need to define the biomass objective function, which often involves a linear combination of more than 100 metabolites (Raman and Chandra, 2009). Furthermore, it is difficult to define the biomass objective function for higher organisms such as humans, and the essential assumption concerning growth-optimal behavior is not always fulfilled (Schuster et al., 2008).

Another flow model approach for metabolic networks is based on elementary mode (EM) concept, which does not require a biomass objective function. EM is a minimal set of reactions in a steady state, in which all irreversible reactions are used in the appropriate direction (Schuster and Hilgetag, 1994; Schuster et al., 2000). Stelling et al. (2002) estimated the effect of knockouts by the number of elementary modes (EMs) that include the knocked-out reaction. Based on the idea of topological flux balance (TFB) (Smart et al., 2008), the topological impact degree (TID) calculates the number of reactions that have an EM in common with the knocked-out reactions (Jiang et al., 2009; Tamura et al., 2011). The flux balance impact degree (FBID) is the number of reactions that are not included in any EMs that do not include the knocked-out reactions (Zhao et al., 2013). A minimal cut set (MCS) is the minimal set of reactions whose inactivation leads to a failure of the specified reactions (Klamt and Gilles, 2004; Klamt, 2006). Acuña et al. (2009) proved that computing the MCS is NP-hard. MCSs in a metabolic network are

EMs in a dual network (Ballerstein et al., 2012). EMs and MCSs can be calculated by FluxAnalyzer (Klamt and Gilles, 2004) and CellNetAnalyzer (CNA) (Klamt et al., 2007), MATLAB-based tools that can be applied to many network-based problems. Because knockouts may induce side-effects that disables the desired functionality, the constrained MCS allows additional constraints that preserve a set of desired modes (Hädicke and Klamt, 2011). Based on an EM-based analysis and the idea of minimal metabolic functionality, optimal knockout strategies were analyzed to maximize industrial production, and the effect was confirmed by biological experiments (Trinh et al., 2006; Unrean et al., 2010).

Another mathematical model of metabolic networks is the Boolean model. Although not suitable for analyzing the mass flow of metabolic networks, the logical analysis of this model is relatively solid as it needs less information than the flow-based models. The synthetic accessibility is defined as the number of reactions required to transform a set of source metabolites into a set of target metabolites (Wunderlich and Mirny, 2006). For a given set of source nodes (called seeds), the scope is defined as the set of producible compounds (Handorf et al., 2005). The damage is the number of reactions that are affected by the knocked-out reactions (Lemke et al., 2004). To take side-effects into account, Sridhar et al. (2008) developed a branch-and-bound-based algorithm, OPMET, which minimizes damage to non-target nodes. Tamura et al. (2010) developed an integer linear programming (ILP)-based method for the Boolean reaction cut (BRC) problem and analyzed the computational complexity of BRC (Tamura and Akutsu, 2010), in which the number of inhibited reactions is minimized to make target compounds non-producible (Tamura et al., 2010). In the Boolean model of metabolic networks, reactions and compounds can be represented by "AND" and "OR" nodes, respectively, and the network can then be considered as a bipartite graph. Lu et al. considered the problem of adding a minimum reaction set so that the target compound becomes producible in a Boolean metabolic network (Lu et al., 2014). Although the above research focuses on metabolic networks, the basic framework of the Boolean model can be extended to other types of biological networks. For example, Flöttmann et al. formalized signal transduction networks using reaction contingency-based bipartite Boolean modeling (Flöttmann et al., 2013), and Samaga et al. studied minimal intervention sets (MISs) for Boolean signaling networks (Samaga et al., 2010).

Thus, in many cases, the effects of reaction inhibition are first estimated in each metabolic network model, and then certain optimization problem is defined for each estimation model. Furthermore, in addition to maximizing the desired functions, some problems minimize undesirable functions as side-effects in the same network.

As a reasonable extension of the above research, this paper considers the situation in which different types of cells exist in the same place. For example, many types of bacteria exist in our intestine. Some of them are good bacteria, others are bad bacteria, and they have different metabolic networks. As another example, normal cells and cancer cells may exist in the same organ. In such cases, it may be useful to compute the optimal knockout strategies that would cause the bad cells to lose some essential functionality, but allow the good cells to survive. Therefore, we consider the following minimal knockout for multiple networks (**MKMN**) problem: given two different metabolic networks with source and target compounds, obtain the minimum number of reactions whose inhibition induces the

target compounds to become non-producible in bad cells, but producible in good cells.

Although we focus in this paper on the case where two networks are given, it is straightforward to extend the proposed methods to the case where more than two networks are given.

We analyze this problem using the Boolean model (**MKMN-B**) and the EM-based model (**MKMN-EM**). As both MKMN-B and MKMN-EM are NP-complete (as shown in the "Theoretical results" section), we develop methods based on ILP (Schrijver, 1998; Li et al., 2007) for MKMN-B and MKMN-EM. ILP is often used to formalize NP-complete problems, and there is an efficient free ILP solver called CPLEX (IBM, 2010). In MKMN-B, to properly account for the effect of cycles, we utilize the notion of the maximal valid assignment (**MaxVA**) (Tamura et al., 2010) and the minimal valid assignment (**MinVA**) (Lu et al., 2014). To obtain faster ILP-based methods by reducing the number of variables, we develop **IP-FVS1** and **IP-FVS2**, which utilize the idea based on feedback vertex sets (**FVS**) (Tamura et al., 2010), MaxVA is strictly applied in IP-FVS1, but is not applied to the nodes detected by FVS in IP-FVS2. We also develop faster algorithms **IP-FVS1-approx** and **IP-FVS2-approx** by limiting the number of time steps, although the optimality of the resulting solutions is not ensured. For MKMN-EM, we develop **IP-EM**, in which CNA is used to obtain the EMs, and then formalize MKMN-EM using ILP, since MKMN-EM is NP-complete, even when EMs are given.

We apply our developed methods to the metabolic networks data of *clostridium perfringens SM101* (CPR) and *bifidobacterium longum DJO10A* (BLJ), as downloaded from the KEGG database (Kanehisa and Goto, 2000). The CPR network is denoted by N_1, and the BLJ network is denoted by N_2. Dataset 1 consists of only the central metabolism, and N_1 and N_2 consist of 73 and 82 nodes, respectively. Dataset 2 consists of the carbon metabolism, fatty acid metabolism, and biosynthesis of amino acids, and N_1 and N_2 consist of 251 and 328 nodes, respectively. Dataset 3 consists of whole metabolic networks, and N_1 and N_2 consist of 1231 and 1881 nodes, respectively. We apply IP-FVS1, IP-FVS2, and IP-EM to Datasets 1,2, and 3, where the target compounds are pyruvate, acetyl-CoA, acetate, oxaloacetate, and phosphoenolpyruvate. For most cases in Dataset3, IP-FVS1, IP-FVS2, and IP-EM could not complete the computation a provisional time limit of 2 hours. Hence, we applied IP-FVS1-approx and IP-FVS2-approx to this dataset, and limited the number of time steps to 10. For Dataset 1, IP-FVS1, IP-FVS2, and IP-EM took at most 3 s for every target compound, but in most cases there are no solutions. For Dataset 2, IP-FVS1, IP-FVS2, and IP-EM took at most 20 s for every target compound, and solutions were obtained in about half of the cases. Finally, for Dataset 3, IP-FVS1-approx(10) and IP-FVS2-approx(10) finished their calculations within 15 s for every target compound. We examine the relations between the obtained solutions and predecessors of the target compound in N_1 and N_2.

Abbreviations

Frequently used abbreviations in this paper are listed as follows:

-EM: Elementary Mode
-FBA: Flux Balance Analysis

-MKMN: Minimum Knockout for Multiple Reactions (Problem)
-MKMN-B: Minimum Knockout for Multiple Reactions in the Boolean model (Problem)
-MKMN-EM: Minimum Knockout for Multiple Reactions in the Elementary Mode model (Problem)
-ILP: Integer Linear Programming
-FVS: Feedback Vertex Set
-MaxVA: Maximal Valid Assignment
-MinVA: Minimal Valid Assignment
-IP-FVS1-LP1, IP-FVS2-LP1, IP-FVS1-LP2, IP-FVS2-LP2: Integer Linear Programming-based methods for solving MKMN-B (Method)
-IP-FVS1-approx(t), IP-FVS2-approx(t): approximation algorithms for solving MKMN-B with a time limit t (Method)
-IP-EM: Integer Linear Programming-based methods for solving MKMN-EM (Method)
-CNA: CellNetAnalyzer
-BLJ: *bifidobacterium longum DJO10A*, known as good bacteria for human intestine
-CPR: *clostridium perfringens SM101*, known as bad bacteria for human intestine
-KEGG: Kyoto Encyclopedia of Genes and Genomes

2 Materials and Methods

2.1 Main problems

In this subsection, the main problem **Minimal Knockout for Multiple Networks (MKMN)** is explained using examples. Mathematical definitions are given in the next subsection. The MKMN for the Boolean model and the EM-based model are called **MKMN-B** and **MKMN-EM** respectively.

(Figure 1)

Suppose we have two metabolic networks N_1 and N_2, as shown in Fig. 1 (A). Rectangles and circles represent reactions and compounds, respectively. For example, reaction r_1 has the substrates (reactants) $\{c_1, c_2\}$, and products $\{c_4, c_5, c_6\}$ for both N_1 and N_2. However, since the topologies of N_1 and N_2 are slightly different, N_1 does not include reaction r_4. For N_1, $\{c_1, c_3, c_7\}$ are called source nodes, and are assumed to be supplied by the external environment, whereas $\{c_1, c_2, c_3, c_7\}$ are source nodes for N_2. The purpose of MKMN is to find the minimum number of reactions whose inhibition (deletion) induces the target compound to be non-producible in N_1 but producible in N_2. In Fig. 1 (A), c_9 is the target compound. This should be non-producible in N_1, but producible in N_2 after the minimum number of reaction deletions.

MKMN-B is defined as a problem to solve MKMN in a Boolean model. In MKMN-B, if all substrates exist, the reaction occurs. For example, in Fig. 1 (A), r_1 occurs if both c_1 and c_2 exist. Note that the existence of c_2 is not always the same in N_1 and N_2. In this example, the solution of MKMN-B is to inhibit $\{r_3\}$, as c_2 becomes non-producible, r_1 cannot take place, c_4 becomes non-producible, r_2 cannot take place, and then c_9 becomes non-producible in N_1, but producible in N_2 via r_2 or r_4. Note that directed cycles often play important roles. On the other hand, the inhibition of $\{r_1\}$ is not a solution as this makes c_9 non-producible in both N_1 and N_2. Similarly, inhibiting $\{r_2\}$ is not a solution either, since c_9 would remain producible in both N_1 and N_2.

The problem setting of MKMN-B for N_1 is the same as that for the Boolean Reaction Cut in Tamura et al. (2010). To properly account for the effect of cycles, Tamura et al. (2010) defined the *maximal valid assignment* (**MaxVA**). For example, although the optimal solution of MKMN-B for N_1 and N_2 as in Fig. 1 (A) is $\{r_2\}$, if neither c_2 nor c_6 initially exists in N_1, then neither r_1 nor r_3 can occur, and c_9 becomes non-producible. Therefore, "inhibiting no reactions" looks like the optimal solution of MKMN-B if c_2 is not supplied to N_1 from the external environment. To avoid such ambiguity, MaxVA is defined as the 0-1 assignment that satisfies the Boolean constraint for each node, and the number of 1's is maximal among such 0-1 assignments. In Tamura et al. (2010), it was proved that MaxVA can be calculated by initially assigning a value of 1 to all nodes; the effect of reaction inhibition gradually reaches other nodes, and finally converges to some 0-1 assignment, which corresponds to the MaxVA. In MKMN-B, the assumption of MaxVA is applied to N_1.

(Figure 2)

However, the assumption of MaxVA is not appropriate for N_2 in MKMN-B. For example, suppose that N_1 and N_2 in Fig. 2 (A) are given. If the assumption of MaxVA is applied to both N_1 and N_2, $\{r_1\}$ is obtained as the optimal solution of MKMN-B, since $(c_1, c_2, c_3, c_4, r_1, r_4, r_5, r_6) = (1, 1, 1, 1, 0, 1, 1, 1)$ is the MaxVA of N_2 if $\{r_1\}$ is inhibited. However, $\{r_2, r_3\}$ seems to be a more appropriate solution of MKMN-B, as the inhibition of $\{r_1\}$ disconnects the source node c_1 and the target compound c_3 in N_2. To account for the effect of cycles, and make the target compound producible, the MinVA notion has been shown to be efficient (Lu et al., 2014). Similar to MaxVA, MinVA is a valid assignment, in which the number of 1's is minimal. In MKMN-B, MaxVA and MinVA are applied to N_1 and N_2, respectively.

(Table 1, Figure 3)

MKMN-EM is defined as a problem to solve MKMN in the elementary mode (EM)-based model. An EM describes a feasible and balanced (steady-state) flux distribution through the network, which is minimal with respect to the utilized reactions (enzymes) (Klamt and Gilles, 2004). Note that $\{r_3\}$ is not the solution of MKMN-EM in Fig. 1 (A), since the target compound c_9 becomes non-producible in both N_1 and N_2. In MKMN-EM, we consider only the topology of each EM, and do not consider the magnitude of each flow. Therefore, an EM can be represented by a 0-1 assignment of reaction and compound nodes. For example, Table 1 shows the 0-1 assignments of the EMs for the example network shown in Fig. 3, where $\{c_1, c_3, c_7, c_9\}$ is a set of external compounds. Suppose that the target compound c_7 should become non-producible, where $\{c_1, c_3, c_9\}$ is a set of source nodes. As EM1, EM2, EM3, and EM5 can produce c_7, the set of reactions to be inhibited must include at least one from each of these four EMs. For example, inhibiting $\{r_3, r_5\}$ makes c_7 non-producible since r_3 is included in EM2 and EM5, and r_5 is included in EM1 and EM3.

2.2 Mathematical definitions of main problems

In this subsection, the main problems of this paper are mathematically defined. A metabolic network is defined as a directed bipartite network $N = (V_c, V_r, E)$, where V_c is a set of *compound nodes*, and V_r is a set of *reaction nodes*. E is a set of edges connecting a compound node and a reaction node. Therefore, neighbors of compound nodes are always reaction nodes, and neighbors of reaction nodes are always compound nodes. In MKMN, two metabolic networks $N_1 = (V_{c_1}, V_{r_1}, E_1)$ and $N_2 = (V_{c_2}, V_{r_2}, E_2)$ are given, where $|V_{c_1}| = m_1$, $|V_{r_1}| = n_1$, $|V_{c_2}| = m_2$, $|V_{r_2}| = n_2$, $m = m_1 + m_2$, and $n = n_1 + n_2$ hold. For example, in Fig. 1 (A), $V_{r_1} = \{r_1, r_2, r_3\}$, $V_{c_1} = \{c_1, c_2, \ldots, c_9\}$ $V_{r_2} = \{r_1, r_2, r_3, r_4\}$, $V_{c_2} = \{c_1, c_2, \ldots, c_{10}\}$, $m_1 = 9$, $n_1 = 3$, $m_2 = 10$, and $n_2 = 4$ hold.

V_{ex} is a set of *external nodes*, that are affected by the external environment not described in N. $V_{s_1} \subset V_{c_1}$ and $V_{s_2} \subset V_{c_2}$ are sets of *source nodes* in N_1 and N_2, respectively. Similarly, $V_{t_1} \subset V_{c_1}$ and $V_{t_2} \subset V_{c_2}$ are sets of *target nodes* in N_1 and N_2, respectively. It holds that $V_{s_1}, V_{s_2}, V_{t_1}, V_{t_2} \subset V_{ex}$. In Fig. 1 (A), $V_{s_1} = \{c_1, c_3, c_7\}$, $V_{s_2} = \{c_1, c_2, c_3, c_7\}$, and $V_{t_1} = V_{t_2} = \{c_9\}$ hold. Note that some $v \in V_{ex}$ may not be included by either V_s or V_t in the EM-based analysis. In this paper, we assume that source nodes do not have

incoming edges, but that target nodes are allowed to have out-going edges to take the multiple target compounds into account.

In both the Boolean model and the EM-based model, values of either "0" or "1" are assigned to each node as the magnitude of each flow is not taken into account in MKMN-EM. Suppose that "1" is assigned to a node. In the Boolean model, this means that either the reaction occurs or the compound exists. In the EM-based model, it means that the corresponding node is included in some EM. If "0" is assigned to a node, then either the reaction does not occur, or the compound does not exist in the Boolean model. In the EM-based model, it means that the corresponding node is not included in the EM. Let $V_a \subseteq V_{r_1} \cup V_{r_2}$ be perturbed reactions. If $v \in V_r$ is included in V_a, $v = 0$ always holds whatever values are assigned to other nodes.

Let A_B be such an assignment. In the Boolean model, A_B is called a *valid assignment* (**VA**) if it satisfies each of the following:

- (i) For each $v \in V_s$, $v = 1$ holds,
- (ii) For each $v \in V_c - V_s$, $v = 1$ if and only if there is some $u \in V_r$ such that $(u, v) \in E$ and $u = 1$ hold,
- (iii) For each $v \in V_r$, $v = 1$ holds if and only if $v \notin V_a$ holds and $u = 1$ holds for all $(u, v) \in E$.

Therefore, in the Boolean model, each reaction node corresponds to an "AND" node, and each compound node corresponds to an "OR" node. A is the **MaxVA** for V_a, if A is a valid assignment for the given V_a and the number of 1's is the maximal. Similarly, A is the **MinVA** for V_a, if it is a valid assignment for the given V_a and the number of 1's is minimal.

MKMN-B is mathematically defined as follows:

Problem: MKMN-B (Minimum Knockout for Multiple Networks in the Boolean model)

- **Input:** Metabolic networks $N_1 = (V_{c_1}, V_{r_1}, E_1)$, $N_2 = (V_{c_2}, V_{r_2}, E_2)$, source nodes $V_{s_1} \subset V_{c_1}$ $V_{s_2} \subset V_{c_2}$, and target nodes $V_{t_1} \subset V_{c_1}$, $V_{t_2} \subset V_{c_2}$.
- **Output:** The minimum cardinality set $V_a \subseteq V_{r_1} \cup V_{r_2}$ such that MaxVA for N_1 ensures every $v \in V_{t_1}$ is 0, and MinVA for N_2 ensures every $v \in V_{t_2}$ is 1.

Note again that the MaxVA and the MinVA are applied to N_1 and N_2, respectively. (See also the examples of Fig.s 1 and 2.)

On the other hand, an EM describes a feasible and balanced (steady-state) flux distribution through the network, which is minimal with respect to utilized reactions (enzymes) (Klamt and Gilles, 2004). An EM is said to be a *relevant* EM if $v_1 = v_2 = 1$ holds for some $v_1 \in V_s$ and $v_2 \in V_t$ in its 0-1 assignment. Then, MKMN-EM is mathematically defined as follows:

Problem: MKMN-EM (Minimum Knockout for Multiple Networks in the Elementary Mode model)

- **Input:** Metabolic networks $N_1 = (V_{c_1}, V_{r_1}, E_1)$, $N_2 = (V_{c_2}, V_{r_2}, E_2)$, external nodes $V_{ex1} \subset V_{c_1} \cup V_{r_1}$, $V_{ex2} \subset V_{c_2} \cup V_{r_2}$, target nodes $V_{t_1} \subset V_{c_1}$, $V_{t_2} \subset V_{c_2}$, and a stoichiometry matrix S.
- **Output:** The minimum cardinality set $V_a \subseteq V_{r_1} \cup V_{r_2}$, which satisfies each of the following:
  - For all $v \in V_{t_1}$, $v = 0$ holds for any relevant EM on $N_1 = (V_{c_1}, V_{r_1} \setminus V_a, E_1)$.
  - For all $v \in V_{t_2}$, $v = 1$ holds for some relevant EM on $N_2 = (V_{c_2}, V_{r_2} \setminus V_a, E_2)$.

2.3 Integer linear programming-based method for MKMN-B

Since MKMN-B is NP-hard as discussed in "Theoretical results", we develop an integer linear programming (ILP)-based method for its solution. In ILP, every Boolean constraint must be represented by linear equations or inequalities. In this paper, we use two linear representations of Boolean constraints:

LP1 (Tamura et al., 2010): Since the Boolean "AND" relation $y = x_1 \wedge x_2 \wedge \cdots \wedge x_k$ can be converted into $(y \vee \overline{x_1} \vee \overline{x_2} \vee \cdots \vee \overline{x_k}) \wedge (\overline{y} \vee x_1) \wedge (\overline{y} \vee x_2) \wedge \cdots \wedge (\overline{y} \vee x_k) = 1$, it can be represented by the following linear inequalities:

$$\begin{aligned} y + (1 - x_1) + (1 - x_2) + \cdots + (1 - x_k) &\geq 1, \\ (1 - y) + x_1 &\geq 1, \\ (1 - y) + x_2 &\geq 1, \\ \cdots & \\ (1 - y) + x_k &\geq 1, \end{aligned}$$

where all variables are binary.

Similarly, as the Boolean "OR" relation $y = x_1 \vee x_2 \vee \cdots \vee x_k$ can be converted into $(\overline{y} \vee x_1 \vee x_2 \vee \cdots \vee x_k) \wedge (y \vee \overline{x_1}) \wedge (y \vee \overline{x_2}) \wedge \cdots \wedge (y \vee \overline{x_k}) = 1$, it can be represented by the following linear inequalities:

$$\begin{aligned} (1 - y) + x_1 + x_2 + \cdots + x_k &\geq 1, \\ y + (1 - x_1) &\geq 1, \\ y + (1 - x_2) &\geq 1, \\ \cdots & \\ y + (1 - x_k) &\geq 1, \end{aligned}$$

where all variables are binary.

LP2 (Akutsu et al., 2012) : Another type of linear function representation of Boolean functions is as follows: The Boolean "AND" can be represented by the following linear inequalities:

$$\begin{aligned} ky &\leq x_1 + x_2 + \ldots + x_k, \\ y &\geq x_1 + \ldots + x_k - (k - 1), \end{aligned}$$

where all variables are binary.

Similarly, the Boolean "OR" can be represented by the following linear inequalities:

$$\begin{aligned} ky &\geq x_1 + x_2 + \ldots + x_k, \\ y &\leq x_1 + \ldots + x_k, \end{aligned}$$

where all variables are binary.

As described in Tamura et al. (2010), we should introduce notion of time for calculating the MaxVA. A naive method for calculating the MaxVA is as follows. Each node is initially assigned 1, specified reactions are inhibited, effects of the inhibition affect neighbor nodes in the next time step, and the finally converged 0-1 assignment is the MaxVA. Similarly, as described in Lu et al. (2014), the MinVA can be calculated by initially assigning 0 to each node other than the source nodes. However, in this naive method, the number of time steps should be the same as the total number of nodes to ensure that the 0-1 assignment converges. Thus, the number of variables needed for the ILP formalization is $O((m+n)^2)$.

By utilizing an FVS, the number of time steps is reduced to f, and then the number of variables in the ILP formalization becomes $O(f(m+n))$. (The FVS is a set of nodes whose removal makes the graph acyclic, and f is the size of the FVS (Tamura et al., 2010).) For example, r_1 of N_1 in Fig. 1 (A) can be decomposed into r_1 and s_1, as shown in N_1' of Fig. 1 (B). Different from the naive method, the time step advances by one only when the value of the "r" node is copied to the "s" node. Suppose that r_3 is inhibited in Fig. 1 (B). Because we consider the MaxVA for N_1 (and N_1'), all source nodes and "s" nodes are assigned 1 at $t = 0$. If r_3 is not inhibited, all nodes become 1 at $t = 0$ since the time step advances only when the value of r_1 is copied to s_1. However, since r_3 is inhibited, $(r_3, c_2, r_1) = (0, 0, 0)$ holds at $t = 0$, whereas $(s_1, c_6) = (1, 1)$ holds at $t = 0$. Then, $r_1 = 0$ at $t = 0$ is copied to s_1 at $t = 1$. Thus, all nodes other than c_1, c_3, c_7 become 0 when r_3 is inhibited. Note that the necessary time steps for calculating this is 2 ($t = 0$ and $t = 1$). This is the size of FVS +1. In the naive method, the time step advances whenever 0 affects its adjacent nodes. Therefore, the number of necessary time steps is as large as the number of nodes to ensure that the 0-1 assignment converges.

Based on this idea, we formalize an ILP-based method to solve MKMK-B. In the following, we explain our proposed methods using examples. It is straightforward to extend the examples to a general case.

When N_1 and N_2 in Fig. 1 (A) are given. **IP-FVS1-LP1** is formalized as follows:

Maximize

$$\text{TE1}(0) + \text{TE2}(0) + \text{TE3}(0) + \text{TE4\_N2}(0) \tag{1}$$

Subject to

$$\text{TC9\_N1}(f_1) = 0, \tag{2}$$

$$\text{TC9\_N2}(f_2) = 1 \tag{3}$$

/* Definitions of N_1

for all $t_1 = 0,...,f_1$

/* Reactions

$$\text{TR1\_N1}(t_1) + \text{FC1\_N1}(t_1) + \text{FC2\_N1}(t_1) + \text{FE1}(t_1) \geq 1,$$
$$\text{FR1\_N1}(t_1) + \text{TC1\_N1}(t_1) \geq 1,\ \ \text{FR1\_N1}(t_1) + \text{TC2\_N1}(t_1) \geq 1,$$
$$\text{FR1\_N1}(t_1) + \text{TE1}(t_1) \geq 1 \tag{4}$$

$$\text{TR2\_N1}(t_1) + \text{FC3\_N1}(t_1) + \text{FC4\_N1}(t_1) + \text{FE2}(t_1) \geq 1,$$
$$\text{FR2\_N1}(t_1) + \text{TC3\_N1}(t_1) \geq 1,\ \ \text{FR2\_N1}(t_1) + \text{TC4\_N1}(t_1) \geq 1,$$
$$\text{FR2\_N1}(t_1) + \text{TE2}(t_1) \geq 1 \tag{5}$$

$$\text{TR3\_N1}(t_1) + \text{FC6\_N1}(t_1) + \text{FC7\_N1}(t_1) + \text{FE3}(t_1) \geq 1,$$
$$\text{FR3\_N1}(t_1) + \text{TC6\_N1}(t_1) \geq 1,\ \ \text{FR3\_N1}(t_1) + \text{TC7\_N1}(t_1) \geq 1,$$
$$\text{FR3\_N1}(t_1) + \text{TE3}(t_1) \geq 1 \tag{6}$$

/* Compounds

$$\text{TC2\_N1}(t_1) = \text{TR3\_N1}(t_1) \tag{7}$$

$$\text{TC4\_N1}(t_1) = \text{TSR1\_N1}(t_1) \tag{8}$$

$$\text{TC5\_N1}(t_1) = \text{TSR1\_N1}(t_1) \tag{9}$$

$$\text{TC6\_N1}(t_1) = \text{TSR1\_N1}(t_1) \tag{10}$$

$$\text{TC8\_N1}(t_1) = \text{TR2\_N1}(t_1) \tag{11}$$

$$\text{FC9\_N1}(t_1) + \text{TR2\_N1}(t_1) + \text{TR3\_N1}(t_1) \geq 1,$$
$$\text{TC9\_N1}(t_1) + \text{FR2\_N1}(t_1) \geq 1,$$
$$\text{TC9\_N1}(t_1) + \text{FR3\_N1}(t_1) \geq 1 \tag{12}$$

/* SR1

$$\text{TSR1\_N1}(t_1 + 1) = \text{TR1\_N1}(t_1) \tag{13}$$

/* Es

$$\text{TE1}(t_1 + 1) = \text{TE1}(t_1) \tag{14}$$
$$\text{TE2}(t_1 + 1) = \text{TE2}(t_1) \tag{15}$$
$$\text{TE3}(t_1 + 1) = \text{TE3}(t_1) \tag{16}$$

/* Source compounds

$$\text{TC1\_N1}(t_1) = 1 \tag{17}$$
$$\text{TC3\_N1}(t_1) = 1 \tag{18}$$
$$\text{TC7\_N1}(t_1) = 1 \tag{19}$$

/* Definitions of N_2

for all $t_2 = 0,...,f_2$

/* Reactions

$$\begin{aligned}&\mathrm{TR1\_N2}(t_2) + \mathrm{FC1\_N2}(t_2) + \mathrm{FC2\_N2}(t_2) + \mathrm{FE1}(t_2) \geq 1,\\ &\mathrm{FR1\_N2}(t_2) + \mathrm{TC1\_N2}(t_2) \geq 1,\ \ \mathrm{FR1\_N2}(t_2) + \mathrm{TC2\_N2}(t_2) \geq 1,\\ &\mathrm{FR1\_N2}(t_2) + \mathrm{TE1}(t_2) \geq 1\end{aligned} \tag{20}$$

$$\begin{aligned}&\mathrm{TR2\_N2}(t_2) + \mathrm{FC3\_N2}(t_2) + \mathrm{FC4\_N2}(t_2) + \mathrm{FE2}(t_2) \geq 1,\\ &\mathrm{FR2\_N2}(t_2) + \mathrm{TC3\_N2}(t_2) \geq 1,\ \ \mathrm{FR2\_N2}(t_2) + \mathrm{TC4\_N2}(t_2) \geq 1,\\ &\mathrm{FR2\_N2}(t_2) + \mathrm{TE2}(t_2) \geq 1\end{aligned} \tag{21}$$

$$\begin{aligned}&\mathrm{TR3\_N2}(t_2) + \mathrm{FC6\_N2}(t_2) + \mathrm{FC7\_N2}(t_2) + \mathrm{FE3}(t_2) \geq 1,\\ &\mathrm{FR3\_N2}(t_2) + \mathrm{TC6\_N2}(t_2) \geq 1,\ \ \mathrm{FR3\_N2}(t_2) + \mathrm{TC7\_N2}(t_2) \geq 1,\\ &\mathrm{FR3\_N2}(t_2) + \mathrm{TE3}(t_2) \geq 1\end{aligned} \tag{22}$$

$$\begin{aligned}&\mathrm{TR4\_N2}(t_2) + \mathrm{FC5\_N2}(t_2) + \mathrm{FE4\_N2}(t_2) \geq 1,\\ &\mathrm{FR4\_N2}(t_2) + \mathrm{TC5\_N2}(t_2) \geq 1,\ \ \mathrm{FR4\_N2}(t_2) + \mathrm{TE4\_N2}(t_2) \geq 1\end{aligned} \tag{23}$$

/* Compounds

$$\mathrm{TC4\_N2}(t_2) = \mathrm{TR1\_N2}(t_2) \tag{24}$$

$$\mathrm{TC5\_N2}(t_2) = \mathrm{TR1\_N2}(t_2) \tag{25}$$

$$\mathrm{TC6\_N2}(t_2) = \mathrm{TR1\_N2}(t_2) \tag{26}$$

$$\mathrm{TC8\_N2}(t_2) = \mathrm{TR2\_N2}(t_2) \tag{27}$$

$$\begin{aligned}&\mathrm{FC9\_N2}(t_2) + \mathrm{TR2\_N2}(t_2) + \mathrm{TR3\_N2}(t_2) + \mathrm{TR4\_N2}(t_2) \geq 1,\\ &\mathrm{TC9\_N2}(t_2) + \mathrm{FR2\_N2}(t_2) \geq 1,\ \ \mathrm{TC9\_N2}(t_2) + \mathrm{FR3\_N2}(t_2) \geq 1,\\ &\mathrm{TC9\_N2}(t_2) + \mathrm{FR4\_N2}(t_2) \geq 1\end{aligned} \tag{28}$$

$$\mathrm{TC10\_N2}(t_2) = \mathrm{TR3\_N2}(t_2) \tag{29}$$

/* Es

$$\mathrm{TE1}(t_2 + 1) = \mathrm{TE1}(t_2) \tag{30}$$
$$\mathrm{TE2}(t_2 + 1) = \mathrm{TE2}(t_2) \tag{31}$$
$$\mathrm{TE3}(t_2 + 1) = \mathrm{TE3}(t_2) \tag{32}$$
$$\mathrm{TE4\_N2}(t_2 + 1) = \mathrm{TE4\_N2}(t_2) \tag{33}$$

/* Source compounds

$$\mathrm{TC1\_N2}(t_2) = 1 \tag{34}$$
$$\mathrm{TC2\_N2}(t_2) = 1 \tag{35}$$
$$\mathrm{TC3\_N2}(t_2) = 1 \tag{36}$$
$$\mathrm{TC7\_N2}(t_2) = 1 \tag{37}$$

/* SR1

$$\mathrm{TSR1\_N1}(0) = 1 \tag{38}$$

$$/^* f_1 = \mid F_1 \mid + 1, \text{ where } F_1 \text{ is the FVS of } N_1$$
$$f_1 = 2 \tag{39}$$
$$/^* f_2 = \mid F_2 \mid + 1, \text{ where } F_2 \text{ is the FVS of } N_2$$
$$f_2 = 1 \tag{40}$$

$$\text{TX} + \text{FX} = 1 \text{ for any X} \tag{41}$$

where every variable takes a value of either 0 or 1. "T" and "F" stand for true(1) and false(0), respectively. "R", "C" and "E" stand for reaction, compound, and enzyme, respectively. For each reaction node, TR$i(t) = 1$ (resp., FR$i(t) = 1$) represents $r_i = 1$ (resp., $r_i = 0$). Therefore, TR$i(t)$ + FR$i(t) = 1$ holds for any reaction node r_i at time step t. "$t = 0$" means the initial time step. For example, if TR2(1) = 0, which means r_2 = 0 (the reaction node r_2 does not occur) at $t = 1$, FR2(1) = 1 automatically holds at the same time. The use of FR$i(t)$ simplifies this illustration. In the implementation, FR$i(t)$ is replaced by 1-TR$i(t)$ to reduce the number of variables. Similarly, each compound c_i is represented by TC$i(t)$ and FC$i(t)$, where TC$i(t)$ + FC$i(t) = 1$. For example, TC3(1) = 1 means that $c_3 = 1$ (the compound c_3 exists) at $t = 1$, and FC3(1) = 1 means that c_3 = 0 (the compound c_3 does not exist) at $t = 1$.
The maximal number of time steps f_1 and f_2 can be calculated by $f_1 = |F_1| + 1$ and $f_2 = |F_2| + 1$, where $|F_1|$ and $|F_2|$ are the sizes of the FVS for N_1 and N_2, respectively. Note that "+1" is necessary because at least 1 time step is necessary even if there is no FVS. Furthermore, since the number of time steps of N_1 and N_2 may be different, the ILP formalization is defined for each network. In the above example for **IP-FVS1-LP1**, (2), (4)-(19), (38), and (39) give the ILP formalization of $N1$, whereas (3), (20)-(37), and (40) give the ILP formalization of $N2$. To distinguish each node, "_N1" (resp., "_N2") is appended to each variable name to denote which network the node belongs to. For example, TR1_N1(1) = 1 means that $r_1 = 1$ (the reaction node r_1 occurs) in N_1 at $t = 1$, and FR1_N2(3) = 1 means that $r_1 = 0$ (the reaction r_1 does not occur) in N_2 at $t = 3$. "_N1" and "_N2" are not appended to the variables of nodes that are common to both N_1 and N_2.
In **IP-FVS1-LP1**, the Boolean "AND" relation of a reaction node is converted into LP1 type linear inequalities as shown in (5). The relation in (5) represents the constraints between the reaction node r_2 and each incoming compound node (c_3 and c_4) in N_1. Two additional variables TE2(t_1) and FE2(t_1) are also included in (5), since we add a virtual predecessor node e_i to each reaction node r_i in both networks. TE$i(t)$ and FE$i(t)$ denote whether r_i is inhibited. Since r_i is represented by an "AND" node, $e_i = 0$ can ensure that r_i remains inactive, even if all other predecessors of r_i are 1. Furthermore, the same TE$i(t)$ and FE$i(t)$ are used for each common reaction node r_i between N_1 and N_2. In the above example for **IP-FVS1-LP1**, the common reaction nodes r_2 of the two networks are connected by e_2 (TE2 and FE2 in (5) and (21)).

The Boolean "OR" relation of a compound node is converted into LP1-type linear inequalities, as shown in (12), when the indegree of the compound node is more than 1. If the indegree of the compound is 1, we need only copy the value from the variable of the predecessor reaction node as shown in (7).

For N_1, the values of each source node and each node newly created by s_i at $t = 0$ are

set to 1 to realize the **MaxVA** as shown in (17)-(19) and (38). The value of the target node is set to 0 at $t = f_1$ to ensure that the target node eventually cannot be produced in N_1.

For N_2, only the source nodes and target node values at $t = f_1$ are set to 1, as shown in (3) and (34)-(37), to realize **MinVA**.

The MKMN result can be calculated by maximizing TE$i(t)$, as shown in (1), since e_i represents whether the reaction node is inhibited. As mentioned above, FEi(t) is replaced by 1-TE$i(t)$. Relation (1) implies that the minimal number of reaction knockouts is calculated by maximizing the number of reaction nodes that are not knocked out. In addition, (2) indicates that the target compound c_9 should eventually not be produced in N_1, whereas (3) suggests that the target compound c_9 should eventually be produced in N_2.

Relations (4)-(19) are the node constraints for N_1. $\{r_1\}$ is chosen as the FVS F_1. The maximal time step f_1 of N_1 is 2, where $f_1 = |F_1| + 1 = 1 + 1 = 2$. Relations (4)-(6) represent the constraints on reactions r_1-r_3, respectively, and (7)-(12) represent the constraints on compounds c_2, c_4, c_5, c_6, c_8, and c_9, respectively. In (8)-(10), the predecessor reaction node r_1 of compounds c_4-c_6 is replaced by s_1, where TSR1_N1$(t_1) = 1$ represents $s_1 = 1$ (s_1 is active) at t_1 in N_1, and TSR1_N1$(t_1) = 0$ represents $s_1 = 0$ (s_1 is inactive) at t_1 in N_1. Relation (13) increases the time step by 1 when the value of r_1 is copied to s_1. Relations (14)-(16) ensure that the state of each reaction (inhibited or uninhibited) does not change during the time transition, and (17)-(19) define c_1, c_3, and c_7 as the source nodes of N_1.

Relations (20)-(37) are the node constraints for N_2. Since the FVS $F_2 = \{\}$, the maximal time step f_2 of N_2 is 1, where $f_2 = |F_2| + 1 = 0 + 1 = 1$. Relations (20)-(23) represent the constraints on reactions r_1-r_4, respectively, and (24)-(29) give the constraints on compounds c_4, c_5, c_6, c_8, c_9, and c_{10}, respectively. Relations (30)-(33) ensure that the state of each reaction (inhibited or uninhibited) does not change during the time transition. Reaction nodes r_1, r_2 and r_3 are common to both N_1 and N_2, whereas r_4 exists only in N_2. Relations (34)-(37) define c_1, c_2, c_3 and c_7 as the source nodes of N_2.

Relations (38) states that the initial value of s_1 is set to 1 ($s_1 = 1$ at $t_1 = 0$) for the **MaxVA** in N_1, and (39)-(40) give the maximal time steps of N_1 and N_2, respectively. Relation (41) denotes that "T..." represents "true (1)" and "F..." represents "false (0)."

Since MinVA is calculated for N_2, **IP-FVS1-LP2** for MKMN-B shown in Fig. 2 is as follows:

Maximize

$$\text{TE1}(0) + \text{TE2\_N1}(0) + \text{TE3\_N1}(0) + \text{TE4\_N2}(0) + \text{TE5\_N2}(0) + \text{TE6\_N2}(0) \quad (42)$$

Subject to

$$\text{TC3\_N1}(f_1) = 0, \quad (43)$$

$$\text{TC3\_N2}(f_2) = 1 \quad (44)$$

/* Definitions of N_1

for all $t_1 = 0,...,f_1$

/* Reactions

$2\ \mathrm{TR1\_N1}(t_1) - \mathrm{TC1\_N1}(t_1) - \mathrm{TE1}(t_1) \leq 0,$
$\mathrm{TR1\_N1}(t_1) - \mathrm{TC1\_N1}(t_1) - \mathrm{TE1}(t_1) \geq -1$ (45)

$2\ \mathrm{TR2\_N1}(t_1) - \mathrm{TC2\_N1}(t_1) - \mathrm{TE2\_N1}(t_1) \leq 0,$
$\mathrm{TR2\_N1}(t_1) - \mathrm{TC2\_N1}(t_1) - \mathrm{TE2\_N1}(t_1) \geq -1$ (46)

$2\ \mathrm{TR3\_N1}(t_1) - \mathrm{TC2\_N1}(t_1) - \mathrm{TE3\_N1}(t_1) \leq 0,$
$\mathrm{TR3\_N1}(t_1) - \mathrm{TC2\_N1}(t_1) - \mathrm{TE3\_N1}(t_1) \geq -1$ (47)

/* Compounds

$\mathrm{TC2\_N1}(t_1) = \mathrm{TR1\_N1}(t_1)$ (48)

$2\ \mathrm{TC3\_N1}(t_1) - \mathrm{TR2\_N1}(t_1) - \mathrm{TR3\_N1}(t_1) \geq 0,$
$\mathrm{TC3\_N1}(t_1) - \mathrm{TR2\_N1}(t_1) - \mathrm{TR3\_N1}(t_1) \leq 0$ (49)

/* Es

$\mathrm{TE1}(t_1 + 1) = \mathrm{TE1}(t_1)$ (50)
$\mathrm{TE2\_N1}(t_1 + 1) = \mathrm{TE2\_N1}(t_1)$ (51)
$\mathrm{TE3\_N1}(t_1 + 1) = \mathrm{TE3\_N1}(t_1)$ (52)

/* Source compounds

$\mathrm{TC1\_N1}(t_1) = 1$ (53)

/* Definitions of N_2

for all $t_2 = 0,...,f_2$

/* Reactions

$2\ \mathrm{TR1\_N2}(t_2) - \mathrm{TC1\_N2}(t_2) - \mathrm{TE1}(t_2) \leq 0,$
$\mathrm{TR1\_N2}(t_2) - \mathrm{TC1\_N2}(t_2) - \mathrm{TE1}(t_2) \geq -1$ (54)

$2\ \mathrm{TR4\_N2}(t_2) - \mathrm{TC2\_N2}(t_2) - \mathrm{TE4\_N2}(t_2) \leq 0,$
$\mathrm{TR4\_N2}(t_2) - \mathrm{TC2\_N2}(t_2) - \mathrm{TE4\_N2}(t_2) \geq -1$ (55)

$2\ \mathrm{TR5\_N2}(t_2) - \mathrm{TC4\_N2}(t_2) - \mathrm{TE5\_N2}(t_2) \leq 0,$
$\mathrm{TR5\_N2}(t_2) - \mathrm{TC4\_N2}(t_2) - \mathrm{TE5\_N2}(t_2) \geq -1$ (56)

$2\ \mathrm{TR6\_N2}(t_2) - \mathrm{TC4\_N2}(t_2) - \mathrm{TE6\_N2}(t_2) \leq 0,$
$\mathrm{TR6\_N2}(t_2) - \mathrm{TC4\_N2}(t_2) - \mathrm{TE6\_N2}(t_2) \leq -1$ (57)

/* Compounds

$2\ \mathrm{TC2\_N2}(t_2) - \mathrm{TR1\_N2}(t_2) - \mathrm{TR5\_N2}(t_2) \geq 0,$
$\mathrm{TC2\_N2}(t_2) - \mathrm{TR1\_N2}(t_2) - \mathrm{TR5\_N2}(t_2) \leq 0$ (58)

$\mathrm{TC3\_N2}(t_2) = \mathrm{TR6\_N2}(t_2)$ (59)

$$\mathrm{TC4\_N2}(t_2) = \mathrm{TSR4\_N2}(t_2) \tag{60}$$

/* SR4
$$\mathrm{TSR4\_N2}(t_2 + 1) = \mathrm{TR4\_N2}(t_2) \tag{61}$$

/* Es
$$\mathrm{TE1}(t_2 + 1) = \mathrm{TE1}(t_2) \tag{62}$$
$$\mathrm{TE4\_N2}(t_2 + 1) = \mathrm{TE4\_N2}(t_2) \tag{63}$$
$$\mathrm{TE5\_N2}(t_2 + 1) = \mathrm{TE5\_N2}(t_2) \tag{64}$$
$$\mathrm{TE6\_N2}(t_2 + 1) = \mathrm{TE6\_N2}(t_2) \tag{65}$$

/* Source compounds
$$\mathrm{TC1\_N2}(t_2) = 1 \tag{66}$$

/* SR4
$$\mathrm{TSR4\_N2}(0) = 0 \tag{67}$$

/* $f_1 = |F_1| + 1$, where F_1 is the FVS size of N_1
$$f_1 = 1 \tag{68}$$
/* $f_2 = |F_2| + 1$, where F_2 is the FVS size of N_2
$$f_2 = 2 \tag{69}$$

where each variable takes a value of either 0 or 1. The Boolean relations of the reaction nodes and compound nodes are converted into LP2 type linear inequalities here. As (42) maximizes the number of reactions that are not inhibited, it corresponds to minimizing the number of inhibited reactions. Relation (43) ensures that the target compound c_3 finally becomes non-producible in N_1, whereas (44) guarantees that c_3 remains producible in N_2. Relations (45)-(47) and (54)-(57) represent the Boolean relations of reactions in N_1 and N_2, respectively. Similarly, (48)-(49) and (58)-(60) represent the Boolean relations of compounds in N_1 and N_2, respectively. Relations (50)-(52) and (62)-(65) ensure that the state of each reaction (inhibited or uninhibited) does not change in that time transition. Relations (53) and (66) mean that c_1 is the source compound in both N_1 and N_2, whereas (68)-(69) give the numbers of time steps for N_1 and N_2, which are calculated from the sizes of the FVS. Here, $\{r_4\}$ is chosen as the FVS. In (60), the predecessor reaction node r_4 of the compound c_4 is replaced by s_4. Relation (67) states that the initial value of s_4 is set to 0 ($s_4 = 0$ at $t = 0$) as MinVA is applied to N_2. All variables represented by "F..." are replaced with "T..." variables in the implementation.

Although the above ILP-formalizations do not include cycles in both N_1 and N_2, in the general case, MaxVA and MinVA are applied to N_1 and N_2, respectively.

2.4 Exception of the maximal valid assignment

(Figure 4)

MKMN-B can be solved by IP-FVS1. However, the problem setting of MKMN-B, in

which MaxVA and MinVA are applied to N_1 and N_2 respectively, may be inappropriate for some cases. For example, suppose that N_1 and N_2 are as in Fig. 4 (A). Originally, p_2 and q_2 are from one reversible reaction (r_2), which is decomposed into two irreversible reactions. Both p_2 and q_2 can be inhibited by the inhibition for r_2. If MaxVA is assumed in N_1, neither p_2 nor q_2 becomes 0 unless r_2 is inhibited. Thus, there is a case where an original reversible reaction works as if it were a source node in N_1.

To handle such reversible reactions more appropriately, we define a variant of MaxVA by modifying the FVS-based method as follows: if the cycle detected by the FVS-based method does not include an original irreversible reaction, 0 is assigned to the newly created node at $t = 0$.

For example, suppose that N_1 and N_2 are as in Fig. 4 (A). Since the cycle consisting of $\{r_1, c_6, r_3, c_2\}$ includes an original irreversible reaction, after decomposing r_1 of N_1 into r_1 and s_1 (see Fig. 4 (B)), $s_1(0)$ is assigned a value of 1. Furthermore, cycles consisting of $\{c_3, p_2, c_8, q_2\}$, $\{c_3, p_2, c_9, q_2\}$, $\{c_4, p_2, c_8, q_2\}$, and $\{c_4, p_2, c_9, q_2\}$ are also decomposed. Suppose that p_2 is decomposed as shown in N_1 and N_2 in Fig. 4 (B). In this case, as the cycle does not include an original irreversible reaction, $s_2(0)$ is assigned a value of 0 for both N_1 and N_2 (see also Table 2). Note that if $s_2(0) = 1$, the values of c_9 in N_1 are fixed to 1; and r_2 must be inhibited to make $c_9 = 0$, as it operates as if a reversible reaction was a kind of source node.

(Table 2)

2.5 Fast approximation algorithm for MKMN-B with large networks

Although IP-FVS1 and IP-FVS2 successfully reduce the number of variables in the ILP formalization from $O((m+n)^2)$ to $O(f(m+n))$, $O(f(m+n))$ may be still too large if the network is very big. One reasonable strategy for handling this problem is to limit the number of time steps to some small constant. IP-FVS1 and IP-FVS2 need f time steps to ensure that the MaxVA and MinVA are always calculated, and hence the optimal solution of MKMN-B is always obtained. However, as the number of time steps necessary to obtain the optimal solution of MKMN-B depends on the topology of the network obtained after removing the FVS, we generally require fewer than f time steps. Although the proposed method does not ensure that the solution is optimal (as f time steps are necessary for the worst case), it is possible to confirm that MaxVA and MinVA in the obtained solution satisfy the condition of MKMN-B; that is, the target compound is not producible in N_1, but is producible in N_2. Note that this validation process can be conducted in polynomial time, since MKMN-B is NP-complete (see "Theoretical results" section). Let **IP-FVS1-approx(t)** and **IP-FVS2-approx(t)** be the IP-FVS1 and IP-FVS2 algorithms in which the number of time steps is limited to t. If t is not large, the numbers of variables used in IP-FVS1-approx(t) and IP-FVS2-approx(t) are $O(m+n)$.

2.6 Elementary mode-based method for MKMN-EM

To solve MKMN-EM, we utilize CNA to calculate the EMs. Since MKMN-EM is NP-complete even if the EMs are given (see "Theoretical results" section), we develop an ILP-based method **IP-EM** to solve MKMN-EM for given EMs. For example, suppose that N_1 and N_2 have four and three relevant EMs, respectively as shown in Table 3.

(Table 3)

Then, **IP-EM** can be formulated as follows:

maximize $r_1 + r_2 + r_3 + r_4 + r_5 + r_6 + r_7$

such that

$$r_1 \wedge r_2 \wedge r_4 = 0, \tag{70}$$
$$r_1 \wedge r_3 \wedge r_6 = 0, \tag{71}$$
$$r_2 \wedge r_4 \wedge r_5 = 0, \tag{72}$$
$$r_5 \wedge r_7 = 0, \tag{73}$$
$$(r_1 \wedge r_2 \wedge r_7) \vee (r_2 \wedge r_3 \wedge r_4) \vee (r_4 \wedge r_5) = 1. \tag{74}$$

where all variables are binary, and Boolean relations should be converted into linear inequalities. Constraints (70)-(73) are converted into

$$r_1 + r_2 + r_4 \leq 2, \tag{75}$$
$$r_1 + r_3 + r_6 \leq 2, \tag{76}$$
$$r_2 + r_4 + r_5 \leq 2, \tag{77}$$
$$r_5 + r_7 \leq 1. \tag{78}$$

For example, since (70) requires that at least one of $\{r_1, r_2, r_4\}$ must be 0, it can be represented by (75). Furthermore, the constraint (74) is converted into

$$\begin{aligned} A &= r_1 \wedge r_2 \wedge r_7, \\ B &= r_2 \wedge r_3 \wedge r_4, \\ C &= r_4 \wedge r_5, \\ A \vee B \vee C &= 1, \end{aligned}$$

and then converted into

$$\begin{aligned} A &\leq \frac{1}{3}(r_1 + r_2 + r_7), \\ A &\geq (r_1 + r_2 + r_7) - 2, \\ B &\leq \frac{1}{3}(r_2 + r_3 + r_4), \\ B &\geq (r_2 + r_3 + r_4) - 2, \\ C &\leq \frac{1}{2}(r_4 + r_5), \\ C &\geq (r_4 + r_5) - 1, \\ A + B + C &\geq 1, \end{aligned}$$

using LP2 where all variables are binary. In this example, inhibiting $\{r_1, r_5\}$ is an optimal solution, since either r_1 or r_5 is included in each of EM1, EM2, EM3, and EM4, and neither r_1 nor r_5 is included in EM6.

3 Computational Experiments

We conducted computational experiments on an Intel(R) Xeon(R) CPU E5-2690 0, 2.90GHz with cache size 20.48 MB, operating system on SUSE Linux Enterprise Server 11 SP3 (x86 64). CPLEX version 12.4.0.0 was used as the ILP solver to obtain the global optimality (IBM, 2010).

Metabolic networks of *clostridium perfringens SM101* (CPR) and *bifidobacterium longum DJO10A* (BLJ) were represented by N_1 and N_2, respectively, in the MKMN. KGML (KEGG Markup Language) of BLJ and CPR were downloaded from the KEGG PATHWAY database (Kanehisa and Goto, 2000). Datasets 1, 2, and 3 were extracted and used as small, medium, and large networks, respectively. Details of Datasets 1, 2, and 3 are given in Tables 4, 5, and 6, respectively. We considered the target compounds to be C00022(pyruvate), C00024(acetyl-CoA), C00033(acetate), C00036(oxaloacetate), and C00074(phosphoenolpyruvate). In addition, "5 compounds" indicates the problem where none of the above five compounds are producible in N_1, but all are producible in N_2.

(Tables 4,5,6)

3.1 Datasets

In Dataset 1, N_1 and N_2 correspond to the central metabolism of BLJ and CPR respectively. N_1 includes 36 compounds and 37 reactions (26 reversible reactions), and consists of glycolysis, gluconeogenesis, and pentose phosphate pathway. N_2 includes 43 compounds and 39 reactions (31 reversible), and consists of glycolysis, gluconeogenesis, citrate (TCA) cycle, and pentose phosphate pathway. Thus, there are a total of 155 nodes for Dataset 1; increasing to 212 nodes after decomposing each reversible reactions into two irreversible reactions.

In Dataset 2, N_1 and N_2 correspond to the carbon metabolism, fatty acid metabolism, and biosynthesis of amino acids of BLJ and CPR respectively. N_1 includes 132 compounds and 119 reactions (33 reversible), whereas N_2 includes 157 compounds and 171 reactions (38 reversible). Thus, there are a total of 579 nodes for Dataset 2, rising to 650 nodes after decomposing each reversible reaction into two irreversible reactions.

In Dataset 3, N_1 and N_2 correspond to the whole pathways of BLJ and CPR, respectively. N_1 includes 622 compounds and 609 reactions (209 reversible), whereas N_2 includes 1189 compounds and 1151 reactions (414 reversible). Thus, there are a total of 2340 nodes for Dataset 3, increasing to 2714 nodes after decomposing each reversible reaction into two irreversible reactions.

3.2 Results for Dataset 1

(Tables 7,8)

Tables 7 and 8 show the computation time and size of the solutions obtained for each target compound by each MKMN method for Dataset 1. As Dataset 1 has small networks, the computation time is less than 2 s for every case, and there is generally no MKMN solution.

(Table 9)

However, when the target compound is C00022 (pyruvate), {R01541, R04779} and {R04779, R05605} are solutions for IP-FVS2 and IP-EM, respectively as shown in Table 9, where RXXXXX ($X \in [0-9]$) are the IDs in KEGG. R04779 is a reaction from beta-D-Fructose 6-phosphate to beta-D-Fructose 1,6-bisphosphate located seven steps before pyruvate in the pathway of glycolysis. R01541 is a reaction from 2-Dehydro-3-deoxy-D-gluconate to 2-Dehydro-3-deoxy-6-phospho-D-gluconate locating two steps before pyruvate in the pentose phosphate pathway, and R05605 is a reaction between 2-Dehydro-3-deoxy-6-phospho-D-gluconate and pyruvate in the pentose phosphate pathway.

When the target compound is C00074 (phosphoenolpyruvate), {R04779, R01541} and {R04779} are solutions for IP-FVS2 and IP-EM respectively. When all five compounds are targeted, only MKMN-EM has a solution {R04779, R05605}.

Tables S1-S3 in supplemental data show the relation among the target compound, predecessors of the target compound in N_1 and N_2, and solutions given by each MKMN method for each target compound. For example, Table S1 shows that R00200 and R00703 are predecessors of the target compound C00022 for both N_1 and N_2. R05605 is a predecessor of C00022 in N_1, but not in N_2. R04779 (or R01541) is not a predecessor of C00022 in either N_1 or N_2. In the tables of supporting information, "ko" indicates reactions included in the solution by each MKMN method. We can see that solutions for MKMN are not trivial, as they do not consist of only predecessors of the target compound.

3.3 Results for Dataset 2

(Tables 10,11)

Tables 10 and 11 show the computation time and size of the solutions obtained for each target compound by each MKMN method for Dataset 2. As the networks are not very large, the computation time is generally less than 20 s. MKMN-B-FVS1 only has solutions when the target compound is C00022 or C00024, and MKMN-B-FVS2 only has solutions for C00022, C00024, or C00033; in contrast MKMN-EM has solutions for all cases.

(Table 12, Figure 5)

When the target compound is C00022(pyruvate), the solutions obtained for IP-FVS1, IP-FVS2, and IP-EM are {R00200, R00214, R00216, R00586, R00945, R01196, R05605}, {R00200, R00214, R00216, R01196, R05605}, and {R00214, R00216, R01512, R05605}, respectively listed in Table 12. As shown in Fig. 5, among the above reactions, {R00200, R00220} are predecessors of C00022 in both N_1 and N_2, whereas {R00214, R00216, R001196, R05605} are predecessors of C00022 in N_1, but not in N_2 (see Table S4 in supplemental data). However, {R00586, R00945, R01512} are not predecessors of C00022 in either N_1 or N_2. R00586 is a reaction between L-Serine and O-Acetyl-L-serine in the pathway of cysteine and methionine metabolism. R00945 is a reaction between L-Serine and glycine in the pathway of glycine, serine, and threonine metabolism. R01512 is a reaction between 3-Phospho-D-glyceroyl phosphate and 3-Phospho-D-glycerate in the pathway of Glycolysis. The relations among the reactions included in the obtained solutions and predecessors of N_1 and N_2 are described in Table S4.

When the target compound is C00024(Acetyl-CoA), the solutions for IP-FVS1, IP-FVS2, and IP-EM are {R00230, R00238, R01196}, {R01196}, and {R01196}, respectively (see Table 12). R00230 is a predecessor of C00024 in both N_1 and N_2, whereas R00238 and R01196 are predecessor of C00024 in N_1, but not in N_2 (see Table S5 in supplemental data).

When the target compound is C00033(acetate), IP-FVS1 does not obtain a solution, but IP-FVS2 and IP-EM both gives the solution {R00196}. As described in Table S6 in supplemental data, R00196 is not the predecessor of either N_1 or N_2. Next, when the target compound is C00036(oxaloacetate), only MKMN-EM has a solution, {R00345}. R00345 is a predecessor of C00036 in both N_1 and N_2 as shown in Table S7 of supplemental data. When the target compound is C00074(phosphoenolpyruvate), only MKMN-EM has a solution {R00199, R00206, R01015, R01829, R04533}. As described in Table S8 of supplemental data, R00199 and R00206 are predecessors of C00074 in N_1, but not in N_2. However, none of R01015, R01829, or R04533 are predecessors of C00074 in either N_1 or N_2.

Finally, when all five compounds are targeted, MKMN-B does not have the solution, but MKMN-EM has the solution {R00214, R00216, R01512, R05605}. As described in Table S9 of supplemental data, R00214, R00216 and R05605 are predecessors of at least

one of the five target compounds in N_1, but not in N_2. R01512 is not the predecessor of any of the five target compounds in either N_1 or N_2, and is a reaction between 3-Phospho-D-glyceroyl phosphate and 3-Phospho-D-glycerate locating two steps before C00074 (phosphoenolpyruvate) in the glycolysis pathway.

3.4 Results for Dataset 3

(Tables 13,14)

Tables 13 and 14 show the computation time and size of the solutions obtained for each target compound by each MKMN method for Dataset 3. For IP-FVS1, although the calculation was completed for every target compound, no solutions were found when the target compound was either C00033(acetate), C00074(phosphoenolpyruvate) or "5 compounds." For IP-FVS2, the calculation only completed when the target compound was C00074(phosphoenolpyruvate) or "5 compounds." CNA could not calculate the EMs for the EM-based method, as the networks are too large.

As listed in Table 13, IP-FVS1 can compute solutions with either LP1 or LP2 within 70 min for each target compound. However, IP-FVS2 requires a computation time of over 2 h, and frequently cannot obtain a solution.

(Tables 15,16,17)

As the computation time is large and solutions are not always obtained, we applied **IP-FVS1-approx(10)** and **IP-FVS2-approx(10)** to Dataset 3. Tables 15 and 16 show the computation time and size of the solutions given by IP-FVS1-approx(10) and IP-FVS2-approx(10) for each target compound for MKMN-B for Dataset 3. As can be seen in Table 15, the computation time of IP-FVS1-approx(10) and IP-FVS2-approx(10) is at most 15 s, and computation is always completed although there is no solution for some cases. The relations among solutions of IP-FVS1, IP-FVS1-approx(10), IP-FVS2, and IP-FVS2-approx(10) are summarized in Table 17.

When the target compound is C00022 (pyruvate), the solution obtained by IP-FVS1 is {R00200, R00212, R00214, R00220, R00470, R00703, R00704, R00782, R00896, R02320, R05605, R05636}; the solution given by IP-FVS1-approx(10) is almost the same, but R00896 is replaced by R03105. Although IP-FVS2 cannot obtain a solution, IP-FVS2-approx(10) obtains the solution {R00200, R00470, R01220}. The relations among solution obtained by each method and the predecessors of the target compound in N_1 and N_2 are described in Table S10 of supplemental data.

When the target compound is C00024(acetyl-CoA), IP-FVS1 and IP-FVS1-approx(10) obtain the same solution {R00212, R00228, R00230, R00238, R01177, R04386, R05351}. Although IP-FVS2 cannot obtain a solution, IP-FVS2-approx(10) gives {R00212, R03026, R05351}. The relations among the selected reactions and predecessors of the target compound are described in Table S11 of supplemental data.

Next, when the target compound is C00033(acetate), FVS-type1 has no solution and IP-FVS2 cannot complete the calculation within 2 h, whereas IP-FVS2-approx(10) ob-

tains the solution {R00212, R00238, R05351} (See also Table S12 in supplemental data). When the target compound is C00036(oxaloacetate), the solutions given by IP-FVS1 and IP-FVS1-approx(10) are {R00345, R00355, R00483} and {R00345, R00355, R00357} respectively. Although IP-FVS2 cannot complete the calculation, IP-FVS2-approx(10) obtains the solution {R00345, R00485}. The relation is summarized in Table S13 of supplemental data.

Finally, when the target compound is C00074(phosphoenolpyruvate) or "5 compounds," neither FVS-type1 nor FVS-type2 has solutions.

4 Theoretical Results

Solving ILP is NP-complete. However, it does not mean that a problem that can be formalized as ILP is NP-complete. Therefore, in this section, we prove that MKMN-B and MKMN-EM are NP-complete, and show the appropriateness of formalizing MKMN-B and MKMN-EM as ILP.

Theorem 1 MKMN-B is NP-complete even if the maximum degree is limited to 2, and the substrates and products of a reaction in N_1 are the same as those in N_2.

Proof: The proof of the NP-hardness is similar to that of Theorem 1 in Tamura and Akutsu (2010). The NP-completeness of the decision version of MKMN-B is shown by the polynomial time reduction from the decision version of the Hitting Set Problem (HSP). The former aims to determine whether there exists $|V_a| \leq z$, whose MaxVA satisfies that $v_1 = 0$ for all $v_1 \in V_{t_1}$ and $v_2 = 1$ for all $v_2 \in V_{t_2}$ hold. Given a set of elements $X = \{x_1, \ldots, x_n\}$, a set of subsets $S = \{S_1, \ldots, S_m\}$, and an integer z, the HSP determines whether Y ($Y \subset X$) exists such that $S_i \cap Y \neq \phi$ for any $i = 1, \ldots, m$ and $|Y| \leq z$ (Ausiello, 1999).

When an instance of HSP with X, S and z is given, N_1 and N_2 of MKMN-B are constructed as follows.

$$
\begin{aligned}
V_{c_1} &= \{c_{x_1}, \ldots, c_{x_n}, c_0, c_t\}, V_{s_1} = \{c_0\}, V_{t_1} = \{c_t\}, \\
V_{r_1} &= \{r_{x_1}, \ldots, r_{x_n}\} \cup \{r_{s_1} \ldots, r_{s_m}\} \\
E_1 &= \{(c_0, r_{x_1}), \ldots, (c_0, r_{x_n})\} \cup \{(r_{x_1}, c_{x_1}), \ldots, (r_{x_n}, c_{x_n})\} \\
&\quad \cup \{(r_{s_1}, c_t), \ldots, (r_{s_m}, c_t)\} \cup \{(c_{x_i}, r_{s_j}) | x_i \in S_j\}, \\
V_{c_2} &= \{c_1, c_t\}, V_{s_2} = \{c_1\}, V_{t_2} = \{c_t\}, \\
V_{r_2} &= \{r_1\}, \\
E_2 &= \{(c_1, r_1), (r_1, c_t)\}.
\end{aligned}
$$

For example, if an instance of the HSP is given as $X = \{1, 2, 3, 4\}$ and $S = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{3, 4\}\}$, then N_1 and N_2 of MKMN-B are constructed as shown in Fig. 6 (A). This conversion can be conducted in polynomial time.

In the following, we show that HSP has a solution with $|Y| = z$ if and only if MKMN-B has a solution with $|V_a| = z$. Suppose that HSP has a solution $Y = \{x_i | x_i \in Y\}$. Then, $V_a = \{r_{x_i} | x_i \in Y\}$ is a solution of MKMN-B since $|Y| = |V_a|$ holds. Next, suppose that MKMN-B has a solution V_a. If $r_{s_i} \in V_a$ holds, then V_a', where V_a' is obtained by replacing r_{s_i} with r_{x_j} satisfying $(c_{x_j}, r_{s_i}) \in E_1$ is also a solution. Therefore, we can assume, without generality that V_a does not include r_{s_i}. Then, $Y = \{x_j | r_{x_j} \in V_a\}$ is a solution of the HSP. Since the decision version of MKMN-B is clearly in NP, it belongs to NP-complete.

Each node with degree greater than 2 can be converted into nodes with degree at most 2 by the methods shown in Fig. 6 (B) and (C). Since reactions in N_1 and N_2 created by this reduction do not intersect, MKMN-B is NP-hard, even when reactions common to both N_1 and N_2 have the same substrates and the same products.

□

Theorem 2 MKMN-EM is NP-hard even if the maximum degree is limited to 2, and the substrates and products of a reaction in N_1 are the same as those in N_2.

Proof: In N_1 of the proof of Theorem 1 with $c_0 \in V_{ex1}$, there are m EMs $\{A_{EM_1}, \ldots, A_{EM_m}\}$ with $c_t = 1$ where $r_{si} = 1$ holds only for A_{EM_i}. If the HSP has a solution $Y = \{x_i | x_i \in Y\}$, then $N_1 = \{V_{c_1}, V_{r_1} \setminus V_a, E_1\}$ for $V_a = \{r_{x_i} | x_i \in Y\}$ does not have any EM with $c_t = 1$. On the other hand, suppose that $N_1 = \{V_{c_1}, V_{r_1} \setminus V_a, E_1\}$ does not have any EM with $c_t = 1$. If V_a includes some r_{s_i}, then r_{s_i} can be replaced with some r_{x_j} satisfying $(c_{x_j}, r_{s_i}) \in E_1$ since $N_1 = \{V_{c_1}, V_{r_1} \setminus \{V_a \setminus \{r_{s_i}\}\} \setminus \{r_{x_j}\}, E_1\}$ does not have any EM with $c_t = 1$ either. Therefore we can assume without generality that V_a does not include any r_{s_i}. Then, $Y = \{x_i | r_{x_i} \in V_a\}$ is the solution of HSP. The other parts of the proof are as in those in Theorem 1 other than $c_1 \in V_{ex2}$ is assumed.

□

Theorem 3 MKMN-EM is NP-complete even if EMs of N_1 and N_2 are given.

Proof: If the sets of reactions of N_1 and N_2 have no intersections, then the problem can clearly be reduced to the HSP for EMs of N_1. Note that the completeness holds for the number of EMs, but not for the number of nodes since the number of EMs may be exponential to the number of nodes.

□

5 Discussion

In this paper, we have studied the MKMN problem to determine the minimum set of reactions whose inhibition induces that the target compounds are not producible in N_1, but are producible in N_2. MKMN-B and MKMN-EM are the Boolean version and elementary mode (EM)-based version of MKMN respectively.

For MKMN-B, we developed an integer linear programming (ILP)-based methods IP-FVS1 and IP-FVS2 utilizing the idea of feedback vertex sets (FVS) to reduce the number of variables present in the ILP formalizations. In IP-FVS1, MaxVA and MinVA are strictly applied to N_1 and N_2 respectively to appropriately take the effect of cycles into account. However, since each node in FVS works as if it were a source node in the MaxVA, the solution obtained by IP-FVS1 is not always realistic. To avoid this problem, IP-FVS2 does not apply MaxVA to cycles consisting only of original reversible reactions in N_1. To solve MKMN-B for large networks, we developed IP-FVS1-approx and IP-FVS2-approx. These algorithms are fast since the number of time steps is limited to a small constant, but the optimality of their solutions is not ensured. We also developed IP-EM, an ILP-based method for MKMN-EM in which every EM including a source node and a target compound is inhibited in N_1, but at least one such EM remains in N_2.

We implemented IP-FVS1, IP-FVS2, IP-FVS1-approx, IP-FVS2-approx, and IP-EM in computational experiments, using metabolic networks of *bifidobacterium longum DJO10A* (BLJ) and *clostridium perfringens SM101* (CPR) for N_1 and N_2, respectively. The Datasets 1, 2 and 3 include 155, 650 and 2340 nodes in total respectively. From Tables 8, 11, 14 and 16, it is seen that MKMN tends to have no solution for smaller networks, but has solutions for larger networks. However, Tables 13 and 14 show that solving MKMN for large networks is computationally very expensive. In particular, IP-FVS2 and the IP-EM could not finish the computation within 2 hours for most cases. Tables 15 and 16 show that IP-FVS1-approx and IP-FVS2-approx can solve MKMN-B very efficiently even for large networks. Although optimality is not ensured by IP-FVS1-approx and IP-FVS2-approx, the solutions they produced are optimal for most cases in the Dataset 3 experiment as shown in Table 17. Furthermore, since MKMN-B belongs to NP, it is not difficult to confirm that the obtained solutions by IP-FVS1-approx and IP-FVS2-approx satisfy that the target compound is not producible in N_1 but producible in N_2 even for non-restricted time steps. It remains as a future work to develop a method of finding the smallest number of the time steps to ensure the optimality of IP-FVS1-approx and IP-FVS2-approx. It is to be noted that some bilevel programming-based method seems to be necessary to define MKMN in an FBA model since FBA needs another objective function in addition to minimizing the number of reactions. Applying Petri-net-based methods (Jin et al., 2011) is also interesting since they may extract the good points of both Boolean-based methods and FBA-based methods.

We also analyzed the obtained solutions by checking the relation among the predecessors of N_1 and N_2, and the selected reactions. From Tables S5, S10 and S11, it may be thought that the solutions given by IP-FVS1 and IP-FVS1-approx are trivial when the predecessors of N_2 are not a subset of those of N_1 since choosing all predecessors of N_1 clearly satisfies the condition of MKMN-B. However, Tables S3 and S7 show that this method cannot always obtain the correct solution. On the other hand, when the prede-

cessors of N_2 are a subset of those of N_1, the obtained solution is complex as shown in Table S4. It seems that the solutions given by IP-FVS2, IP-FVS2-approx, and IP-EM are somewhat involved, and it is difficult to infer solutions from these tables. Since a smaller set of inhibited reactions is preferable, solutions of IP-FVS2 and IP-EM are considered to be better than those from IP-FVS1. Moreover, for large networks, since IP-FVS2 and IP-EM are generally unable to complete the calculation, IP-FVS2-approx is considered to be the best method, although the optimality of its solution is not ensured in the worst-case scenario.

As theoretical results, we proved that MKMN-B is NP-complete for the number of nodes, MKMN-EM is NP-hard for the number of nodes, and MKMN-EM is NP-complete for the number of EMs. The software developed in this study and reported here is available at "http://sunflower.kuicr.kyoto-u.ac.jp/~rogi/minFvsKO/minFvsKO.html".

Acknowledgments

TT was partially supported by JSPS, Japan (Grant-in-Aid for Young Scientists (B) 25730005 and Grant-in-Aid for Scientific Research (A) 25250028). JS is an Australian National Health and Medical Research Council (NHMRC) Peter Doherty Fellow and a recipient of the Hundred Talents Program of the Chinese Academy of Sciences (CAS). TA was partly supported by the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists, China, and Grant-in-Aid #22240009 from JSPS, Japan.

Disclosure statement

No competing financialinterests exist.

References

IBM ILOG CPLEX Optimizer. 2010. urlhttp://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

Acuña, V., Chierichetti, F., Lacroix, V., et al. 2009. Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems*, 95(1):51–60.

Akutsu, T., Zhao, Y., Hayashida, M., et al. 2012. Integer programming-based approach to attractor detection and control of boolean networks. *IEICE Transactions on Information and Systems*, 95(12):2960–2970.

Ausiello, G., 1999. *Complexity and approximation: Combinatorial optimization problems and their approximability properties*. Springer.

Ballerstein, K., von Kamp, A., Klamt, S., et al. 2012. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28(3):381–387.

Burgard, A.P., Pharkya, P., and Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657.

Flöttmann, M., Krause, F., Klipp, E., et al., 2013. Reaction-contingency based bipartite boolean modelling. *BMC Systems Biology*, 7(1):58.

Gonçalves, E., Bucher, J., Ryll, A., et al. 2013. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Molecular BioSystems*, 9(7):1576–1583.

Hädicke, O., and Klamt, S., 2011. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering*, 13(2):204–213.

Handorf, T., Ebenhöh, O., and Heinrich, R., 2005. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *Journal of Molecular Evolution*, 61(4):498–512.

Jiang, D., Zhou, S., and Chen Y.P.,. 2009. Compensatory ability to null mutation in metabolic networks. *Biotechnology and Bioengineering*, 103(2):361–369.

Jin, G., Zhao, H., Zhou, X., et al. 2011. An enhanced petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data. *Bioinformatics*, 27(13):i310–i316.

Kanehisa, M., and Goto, S., 2000. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30.

Kauffman, K.J., Prakash, P., and Edwards J.S., 2003. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5):491–496.

Kim, J., and Reed J.L., 2010. Optorf: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Systems Biology*, 4(1):53.

Klamt, S., 2006. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2):233–247.

Klamt, S., and Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234.

Klamt, S., Saez-Rodriguez, J., and Gilles, E.D., 2007. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Systems Biology*, 1(1):2.

Lemke, N., Herédia, F., Barcellos, C.K., et al. 2004. Essentiality and damage in metabolic networks. *Bioinformatics*, 20(1):115–119.

Li, Z., Zhang, S., Wang, Y., et al. 2007. Alignment of molecular networks by integer quadratic programming. *Bioinformatics*, 23(13):1631–1639.

Lu, W., Tamura, T., Song, J., et al. 2014. Integer programming-based method for designing synthetic metabolic networks by minimum reaction insertion in a boolean model. *PLOS ONE*, 9(3):e92637.

Raman, K., and Chandra, N., 2009. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435–449.

Samaga, R., Von Kamp, A., and Klamt, S., 2010. Computing combinatorial intervention strategies and failure modes in signaling networks. *Journal of Computational Biology*, 17(1):39–53.

Schrijver, A., 1998. *Theory of linear and integer programming.* John Wiley & Sons.

Schuster, S., Fell, D.A., and Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332.

Schuster, S., and Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02):165–182.

Schuster, S., Pfeiffer, T., and Fell, D.A., 2008. Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, 252(3):497–504.

Segre, D., Vitkup, D., and Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99: 23: 15112–15117.

Shlomi, T., Cabili, M.N., and Ruppin, E., 2009. Predicting metabolic biomarkers of human inborn errors of metabolism. *Molecular Systems Biology*, 5: 263.

Smart, A.G., Amaral, L.A.N., and Ottino, J.M., 2008. Cascading failure and robustness in metabolic networks. *Proceedings of the National Academy of Sciences*, 105(36):13223–13228.

Sridhar, P., Song, B., Kahveci, T., et al. 2008. Mining metabolic networks for optimal drug targets. In *Pacific Symposium on Biocomputing*, volume 13, pages 291–302.

Stelling, J., Klamt, S., Bettenbrock, K., et al. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193.

Tamura, T., and Akutsu, T., 2010. Exact algorithms for finding a minimum reaction cut under a boolean model of metabolic networks. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 93(8):1497–1507.

Tamura, T., Takemoto, K., and Akutsu, T., 2010. Finding minimum reaction cuts of metabolic networks under a boolean model using integer programming and feedback vertex sets. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, 1(1):14–31.

Tamura, T., Cong, Y., Akutsu, T., et al. 2011. An efficient method of computing impact degrees for multiple reactions in metabolic networks with cycles. *IEICE Transactions on Information and Systems*, 94(12):2393–2399.

Tepper, N., and Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, 26(4): 536–543.

Trinh, C.T., Carlson, R., Wlaschin, A., et al. 2006. Design, construction and performance of the most efficient biomass producing e. coli bacterium. *Metabolic Engineering*, 8(6): 628–638.

Unrean, P., Trinh, C.T., and Srienc F., 2010. Rational design and construction of an efficient e.coli for production of diapolycopendioic acid. *Metabolic Engineering*, 12(2): 112–122.

Varma, A., and Palsson, B.O., 1994. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12:994–998.

Wunderlich, Z., and Mirny, L.A., 2006. Using the topology of metabolic networks to predict viability of mutant strains. *Biophysical Journal*, 91(6):2304–2311.

Zhao, Y., Tamura, T., Akutsu, T., et al. 2013. Flux balance impact degree: a new definition of impact degree to properly treat reversible reactions in metabolic networks. *Bioinformatics*, 29(17):2178–2185.

Table 1: The 0-1 assignments corresponding to the EMs of the example network of Fig. 3, where $\{c_1, c_3, c_7, c_9\}$ is a set of external compounds.

	p_1	q_1	p_2	q_2	r_3	r_4	r_5	r_6	r_7	r_8	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
EM1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	0
EM2	0	1	0	0	1	0	0	0	1	0	1	1	0	0	0	1	1	0	0
EM3	0	1	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1
EM4	1	0	1	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0
EM5	0	0	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0

Table 2: Initial values of nodes included by the FVS.

	IP-FVS1	IP-FVS2
$s_2(0)$ of N_1	1	0
$s_2(0)$ of N_2	0	0

In IP-FVS1, N_1 and N_2 are based on MaxVA and MinVA, respectively. However, in IP-FVS2, in addition to the above assumption, nodes included by the FVS consisting of only reversible reactions are assigned an initial value of 0.

Table 3: Example for MKMN-EM.

N_1		N_2	
EM1	r_1, r_2, r_4	EM5	r_1, r_2, r_7
EM2	r_1, r_3, r_6	EM6	r_2, r_3, r_4
EM3	r_2, r_4, r_5	EM7	r_4, r_5
EM4	r_5, r_7		

After obtaining the EMs that produce the target compound using CellNetAnalyzer, MKMN-EM can be formalized by ILP.

Table 4: Details of the networks of Dataset 1.

	N_1	N_2	total (N_1+N_2)
#compound	36	43	79
#irreversible reaction	11	8	19
#reversible reaction	26	31	57
#node	73	82	155

N_1 and N_2 are the central metabolism of *bifidobacterium longum DJO10A* (BLJ) and *clostridium perfringens SM101* (CPR) respectively. N_1 consists of {cpr00010.xml cpr00030.xml}, and N_2 consists of {blj00010.xml,blj00020.xml,blj00030.xml} in the KEGG database.

Table 5: Details of the networks of Dataset 2.

	N_1	N_2	total (N_1+N_2)
#compound	132	157	289
#irreversible reaction	86	133	219
#reversible reaction	33	38	71
#node	251	328	579

N_1 and N_2 consist of carbon metabolism, fatty acid metabolism, and biosynthesis of amino acids of BLJ and CPR, respectively. N_1 is {cpr01200.xml, cpr01212.xml cpr01230.xml}, and N_2 is {blj01200.xml,blj01212.xml,blj01230.xml} in the KEGG database.

Table 6: Details of the networks of Dataset 3.

	N_1	N_2	total (N_1+N_2)
#compound	622	567	1189
#irreversible reaction	400	337	737
#reversible reaction	209	205	414
#node	1231	1881	2340

N_1 and N_2 consist of whole metabolic networks of BLJ and CPR, respectively, downloaded from KGML in the KEGG database.

Table 7: The computation time of IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 1.

\ type target compound	IP-FVS1 (with LP1)	IP-FVS1 (with LP2)	IP-FVS2 (with LP2)	IP-EM
C00022(Pyruvate)	2.07 sec	1.45 sec	0.65 sec	0.47 sec
C00024(Acetyl-CoA)	<0.01sec	<0.01sec	<0.01sec	<0.01sec
C00033(Acetate)	0.21 sec	0.18 sec	0.19 sec	0.27 sec
C00036(Oxaloacetate)	<0.01sec	<0.01sec	<0.01sec	<0.01sec
C00074(Phosphoenolpyruvate)	0.18 sec	0.18 sec	0.57 sec	0.49 sec
All the above 5 compounds	0.18 sec	0.18 sec	0.18 sec	0.49 sec

Table 8: Size of the solutions given by IP-FVS1, IP-FVS2, and IP-EM for each target compound with Dataset 1.

target compound \ type	IP-FVS1	IP-FVS2	IP-EM
C00022(Pyruvate)	no solution	2	2
C00024(Acetyl-CoA)	no solution	no solution	no solution
C00033(Acetate)	no solution	no solution	no solution
C00036(Oxaloacetate)	no solution	no solution	no solution
C00074(Phosphoenolpyruvate)	no solution	2	1
All the above 5 compounds	no solution	no solution	2

Table 9: Solutions given by IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 1.

target compound	type	the obtained solution
C00022	IP-FVS1	no solution
	IP-FVS2	R01541(N_1), R04779(N_1)
	IP-EM	R04779(N_1), R05605(N_1)
C00074	IP-FVS1	no solution
	IP-FVS2	R04779(N_1), R01541(N_1)
	IP-EM	R04779(N_1)
5 compounds	IP-FVS1	no solution
	IP-FVS2	no solution
	IP-EM	R04779(N_1), R05605(N_1)

"(N_1)" indicates that the reaction appears in N_1, but not in N_2.

Table 10: Computation time of IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 2.

\ type target compound	IP-FVS1 (with LP1)	IP-FVS1 (with LP2)	IP-FVS2 (with LP2)	IP-EM
C00022(Pyruvate)	8.95 sec	11.19 sec	10.5 sec	14.59 sec
C00024(Acetyl-CoA)	8.57 sec	11.31 sec	11.91 sec	4.62 sec
C00033(Acetate)	0.74 sec	0.74 sec	9.82 sec	17.48 sec
C00036(Oxaloacetate)	8.17 sec	7.05 sec	5.32 sec	10.92 sec
C00074(Phosphoenolpyruvate)	0.72 sec	0.74 sec	0.72 sec	13.36 sec
All the above 5 compounds	0.74 sec	0.76 sec	0.74 sec	4.64 sec

Table 11: Size of the solutions given by IP-FVS1, IP-FVS2, and IP-EM for each target compound with Dataset 2.

target compound \ type	IP-FVS1	IP-FVS2	IP-EM
C00022(Pyruvate)	7	5	4
C00024(Acetyl-CoA)	3	1	1
C00033(Acetate)	no solution	1	1
C00036(Oxaloacetate)	no solution	no solution	1
C00074(Phosphoenolpyruvate)	no solution	no solution	5
All the above 5 compounds	no solution	no solution	4

Table 12: Solutions given by IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 2.

target compound	type	the obtained solution
C00022	IP-FVS1	R00200, R00214(N_1), R00216(N_1), R00586(N_1) R00945, R01196(N_1), R05605(N_1)
	IP-FVS2	R00200, R00214(N_1), R00216(N_1), R01196(N_1), R05605(N_1)
	IP-EM	R00214(N_1), R00216(N_1), R01512, R05605(N_1)
C00024	IP-FVS1	R00230, R00238(N_1), R01196(N_1)
	IP-FVS2	R01196(N_1)
	IP-EM	R01196(N_1)
C00033	IP-FVS1	no solution
	IP-FVS2	R01196(N_1)
	IP-EM	R01196(N_1)
C00036	IP-FVS1	no solution
	IP-FVS2	no solution
	IP-EM	R00345
C00074	IP-FVS1	no solution
	IP-FVS2	no solution
	IP-EM	R00199(N_1), R00206(N_1), R01015(N_1), R01829(N_1), R04533(N_1)
5 compounds	IP-FVS1	no solution
	IP-FVS2	no solution
	IP-EM	R00214(N_1), R00216(N_1), R01512, R05605(N_1)

"(N_1)" indicates that the reaction appears in N_1, but not in N_2.

Table 13: The computation time by IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 3.

\ type target compound	IP-FVS1 (with LP1)	IP-FVS1 (with LP2)	LP-type2 IP-FVS2	IP-EM
C00022(Pyruvate)	26min37sec	54min56sec	$>$ 2 hours	NA
C00024(Acetyl-CoA)	32min6sec	56min19sec	$>$ 2 hours	NA
C00033(Acetate)	16min17sec	12min14sec	$>$ 2 hours	NA
C00036(Oxaloacetate)	32min38sec	1hr8min35sec	$>$ 2 hours	NA
C00074(Phosphoenolpyruvate)	21min1sec	12min34sec	5min29sec	NA
All the above 5 compounds	21min44sec	12min15sec	7min2sec	NA

Table 14: Size of the solutions given by IP-FVS1, IP-FVS2 and IP-EM for each target compound with Dataset 3.

target compound \ type	IP-FVS1	IP-FVS2	IP-EM
C00022(Pyruvate)	12	NA	NA
C00024(Acetyl-CoA)	7	NA	NA
C00033(Acetate)	no solution	NA	NA
C00036(Oxaloacetate)	3	NA	NA
C00074(Phosphoenolpyruvate)	no solution	no solution	NA
All the above 5 compounds	no solution	no solution	NA

Table 15: Computation time of IP-FVS1-approx(10) and IP-FVS2-approx(10) for each target compound with Dataset 3.

\ type target compound	IP-FVS1-approx(10) (with LP1)	IP-FVS1-approx(10) (with LP2)	IP-FVS2-approx(10) (with LP2)
C00022(Pyruvate)	9.43sec	6.92sec	12.05sec
C00024(Acetyl-CoA)	9.18sec	6.71sec	14.95sec
C00033(Acetate)	4.46sec	1.22sec	14.01sec
C00036(Oxaloacetate)	8.78sec	7.81sec	10.66sec
C00074(Phosphoenolpyruvate)	2.81sec	0.97sec	1.32sec
All the above 5 compounds	2.72sec	0.96sec	1.37sec

Table 16: The size of the obtained solutions by IP-FVS1-approx(10), IP-FVS2-approx(10) for each target compound with Dataset 3.

target compound \ type	IP-FVS1-approx(10)	IP-FVS2-approx(10)	IP-EM
C00022(Pyruvate)	12	3	NA
C00024(Acetyl-CoA)	7	3	NA
C00033(Acetate)	no solution	3	NA
C00036(Oxaloacetate)	3	2	NA
C00074(Phosphoenolpyruvate)	no solution	no solution	NA
All the above 5 compounds	no solution	no solution	NA

Table 17: Solutions given by IP-FVS1, IP-FVS1-approx(10), IP-FVS2, IP-FVS2-approx(10) and IP-EM for each target compound with Dataset 3.

target compound	type	the obtained solution
C00022	IP-FVS1	R00200, R00212, R00214(N_1), R00220, R00470(N_1), R00703, R00704(N_1), R00782, R00896, R02320, R05605(N_1), R05636
	IP-FVS1-approx(10)	R03105 is chosen instead of R00896.
	IP-FVS2	NA
	IP-FVS2-approx(10)	R00200, R00470(N_1), R01220
	IP-EM	NA
C00024	IP-FVS1	R00212, R00228, R00230, R00238(N_1), R01177(N_1), R04386(N_1), R05351
	IP-FVS1-approx(10)	the same as the above
	IP-FVS2	NA
	IP-FVS2-approx(10)	R00212, R03026(N_1), R05351
	IP-EM	NA
C00033	IP-FVS1	no solution
	IP-FVS2	NA
	IP-FVS2-approx(10)	R00212, R00238(N_1), R05351
	IP-EM	NA
C00036	IP-FVS1	R00345, R00355, R00483(N_1)
	IP-FVS1-approx(10)	R00345, R00355, R00357
	IP-FVS2	NA
	IP-FVS2-approx(10)	R00345, R00485(N_1)
	IP-EM	NA

"(N_1)" indicates that the reaction appears in N_1, but not in N_2.

Figure legends

Figure 1: **An example of Minimum Knockout for Multiple Networks (MKMN) problem.** MKMN is a problem to find the minimum number of reactions whose inhibition makes the target compound non-producible in N_1 but producible in N_2. (A)The MKMN solution for this example is $\{r_3\}$, whose inhibition prevents production of v_{c_9} in N_1 but not in N_2. (B) r_1 of N_1 in Fig. 1 is decomposed into r_1 and s_1.

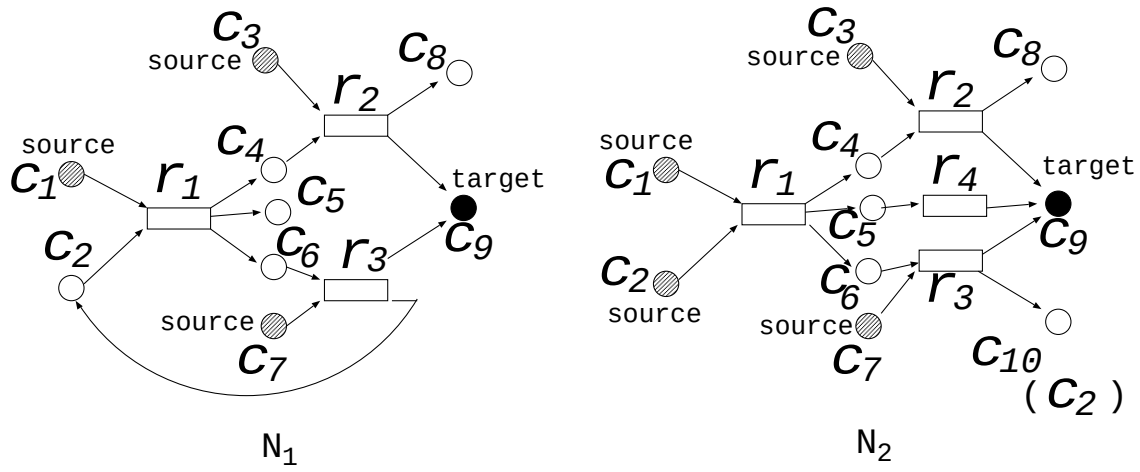
Figure 2: **The minimal valid assignment (MinVA) is applied to N_2.** The FVS-based operation for N_2 is different from that of N_1. (A)$\{r_2, r_3\}$ is the optimal solution of MKMN-B for N_1 and N_2. (B)However, if s_4 is initially assigned a value of 1 in N_2', $\{r_1\}$ is obtained as the solution.

Figure 3: **The 0-1 assignments of elementary modes (EMs) of this network are listed in Table 1.**

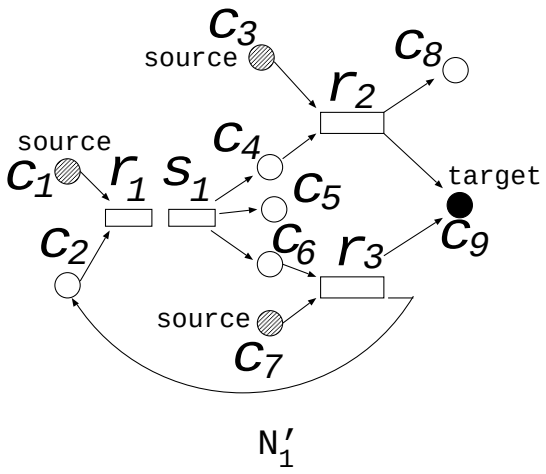
Figure 4: **A reversible reaction r_2 decomposed into p_2 and q_2.** r_1 is decomposed in N_1, and p_2 is decomposed in N_1 and N_2 so that there is no directed cycle in the resulting networks. $s_1(0) = 1$ since the detected cycle in (A) includes an irreversible reaction. However, $s_2(0) = 0$ since the detected cycle in (A) does not include an irreversible reaction.

Figure 5: **Solutions for Dataset 2 with the target compound C00022(Pyruvate).** When the target compound is C00022(Pyruvate), solutions of IP-FVS1, IP-FVS2, and IP-EM for Dataset 2 are represented by red, blue, and green lines, respectively.

Figure 6: **Example of the polynomial time reduction from the HSP.** (A)N_1 and N_2 of MKMN-B are constructed from an instance of the HSP with $X = \{1, 2, 3, 4\}$ and $S = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{3, 4\}\}$. (B)Compound nodes with outdegree greater than 2 can be converted to nodes with outdegree at most 2. (C)Reaction nodes with indegree greater than 2 can be converted to nodes with indegree at most 2.


c3
source
c8
r2
source
c1
r1
c4
c5
target
c2
c6
r3
c9
source
c7
N1
c3
source
c8
r2
source
c1
r1
c4
r4
target
c5
c9
c2
source
c6
r3
source
c7
c10
(c2)
N2

(A)


c3
source
c8
r2
source
c1
r1
s1
c4
c5
target
c2
c6
r3
c9
source
c7
N1′

(B)

Figure 1:

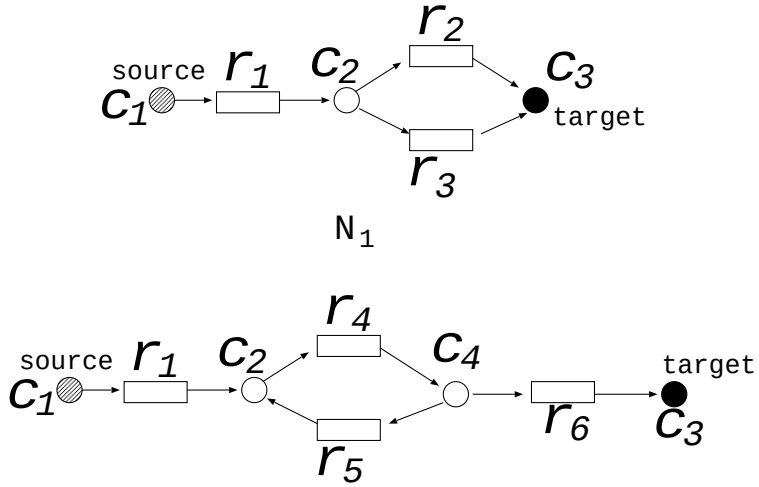

source
r_1
c_2
r_2
c_3
c_1
target
r_3
N_1
r_4
source
r_1
c_2
c_4
target
c_1
r_6
c_3
r_5
N_2
(A)
r_4
s_4
source
r_1
c_2
c_4
target
c_1
r_6
c_3
r_5
N_2'
(B)

Figure 2:

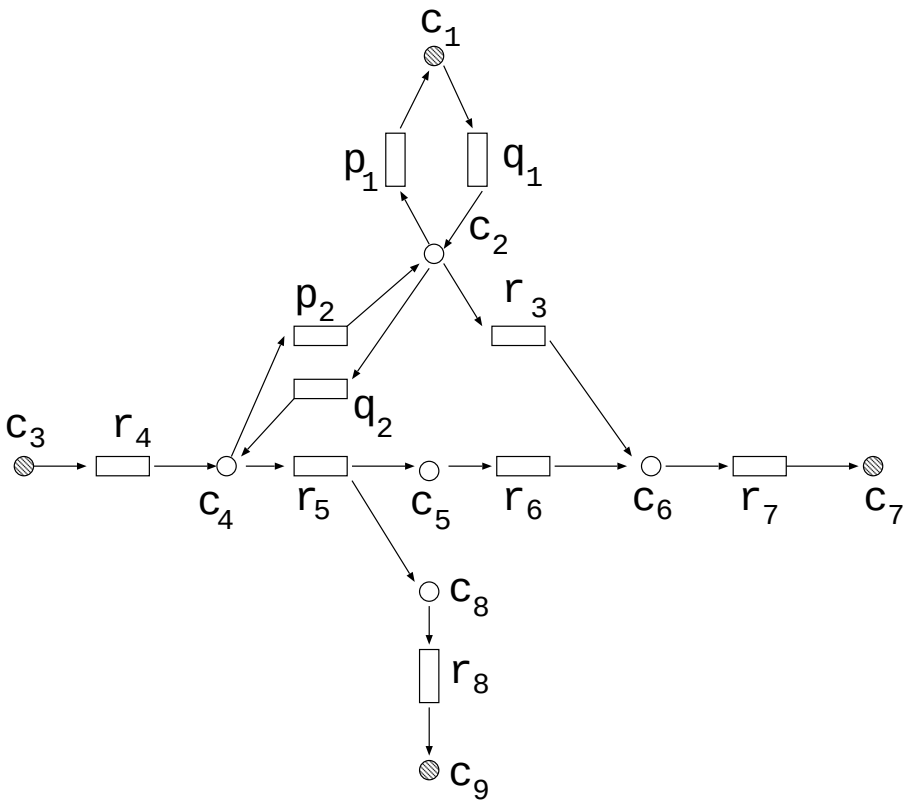


Figure 3:

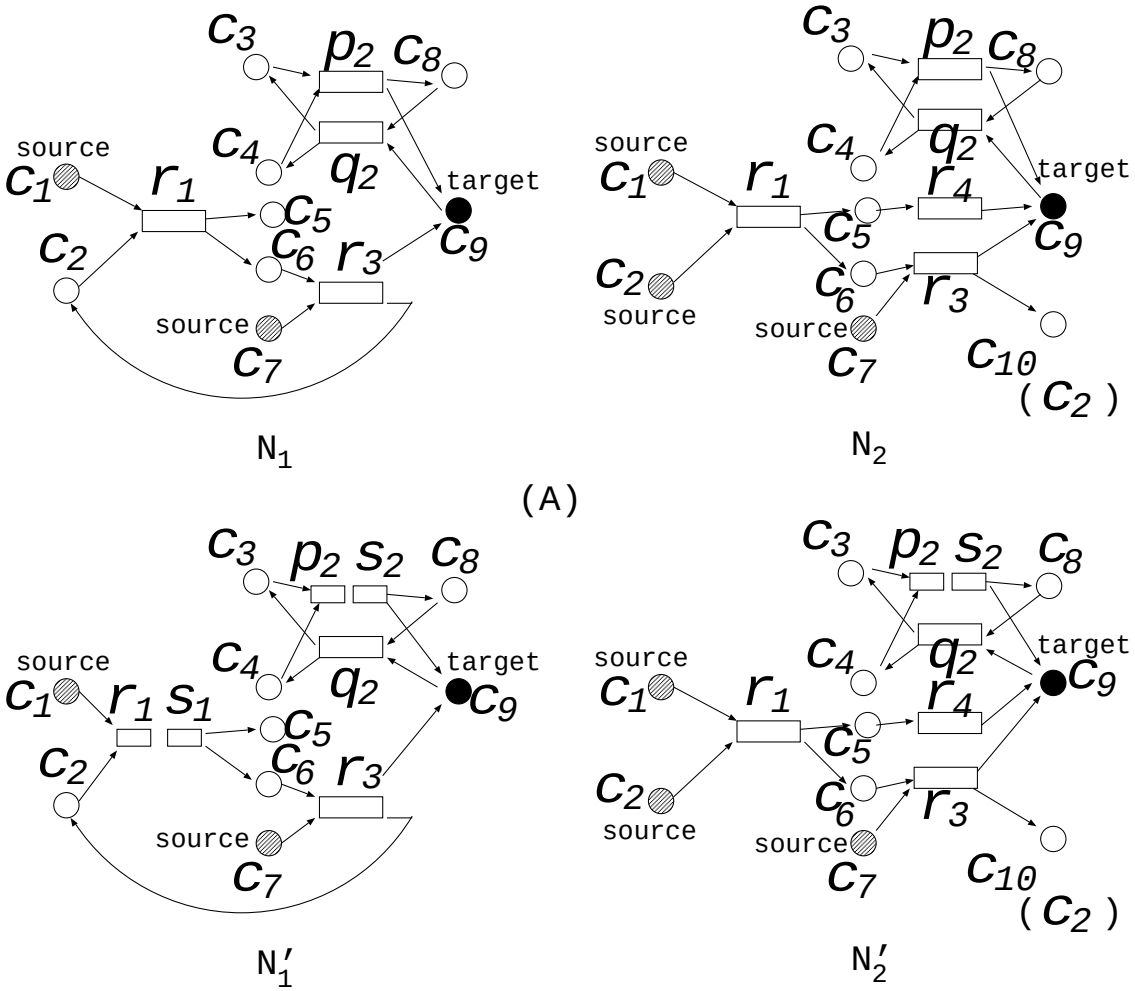

c3
p2
c8
source
c1
r1
c4
q2
target
c5
c2
c6
r3
c9
source
c7
N1
(A)
N2
c10
(c2)
s2
s1
N′1
N′2
(B)

Figure 4:

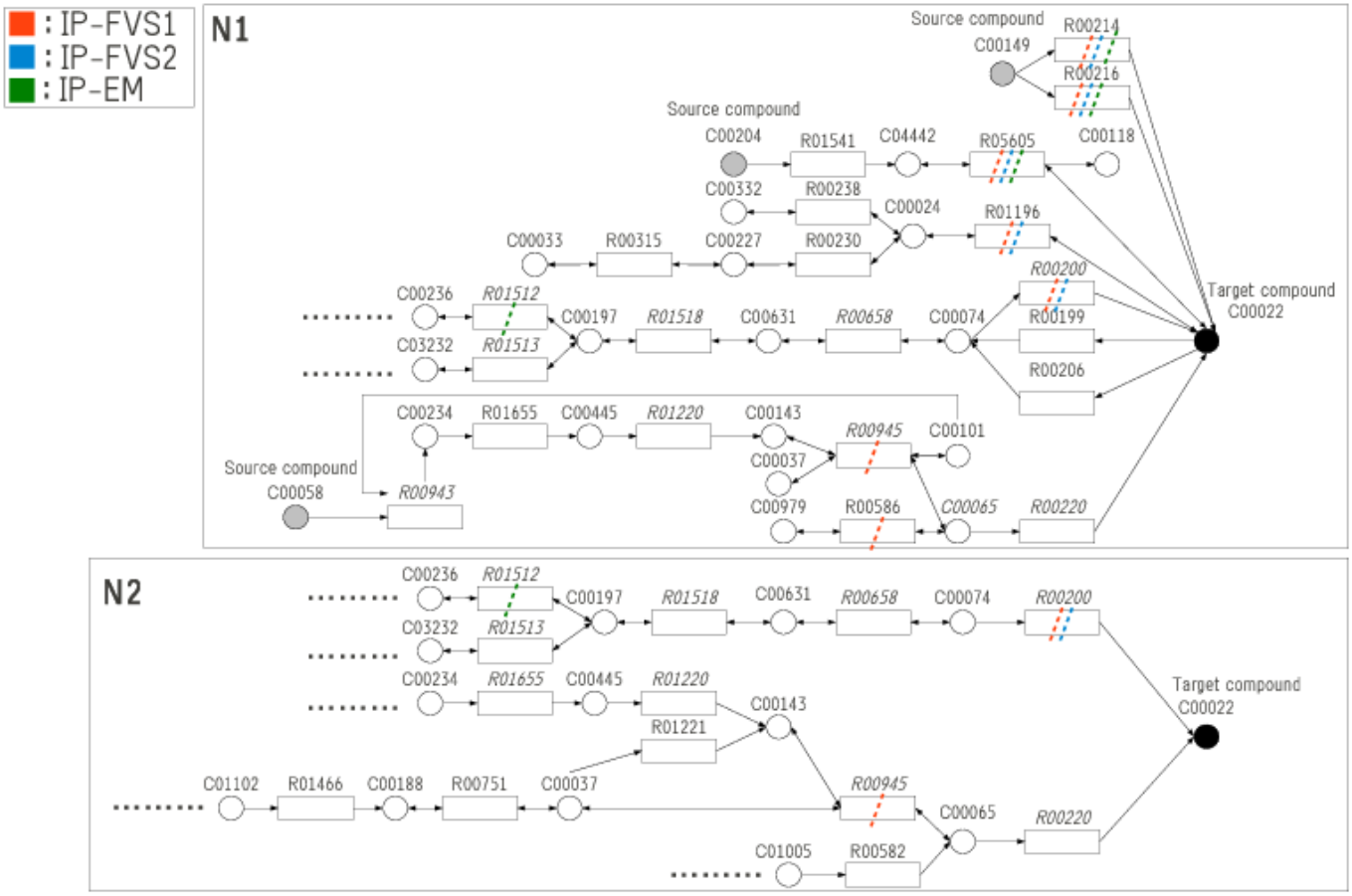

: IP-FVS1
: IP-FVS2
: IP-EM
N1
Source compound
C00149
R00214
R00216
Source compound
C00204
R01541
C04442
R05605
C00118
C00332
R00238
C00024
R01196
C00033
R00315
C00227
R00230
R00200
C00236
R01512
C00197
R01518
C00631
R00658
C00074
R00199
Target compound
C00022
C03232
R01513
R00206
C00234
R01655
C00445
R01220
C00143
R00945
C00101
C00037
Source compound
C00058
R00943
C00979
R00586
C00065
R00220
N2
C00236
R01512
C00197
R01518
C00631
R00658
C00074
R00200
C03232
R01513
C00234
R01655
C00445
R01220
C00143
Target compound
C00022
R01221
C01102
R01466
C00188
R00751
C00037
R00945
C00065
R00220
C01005
R00582

Figure 5:

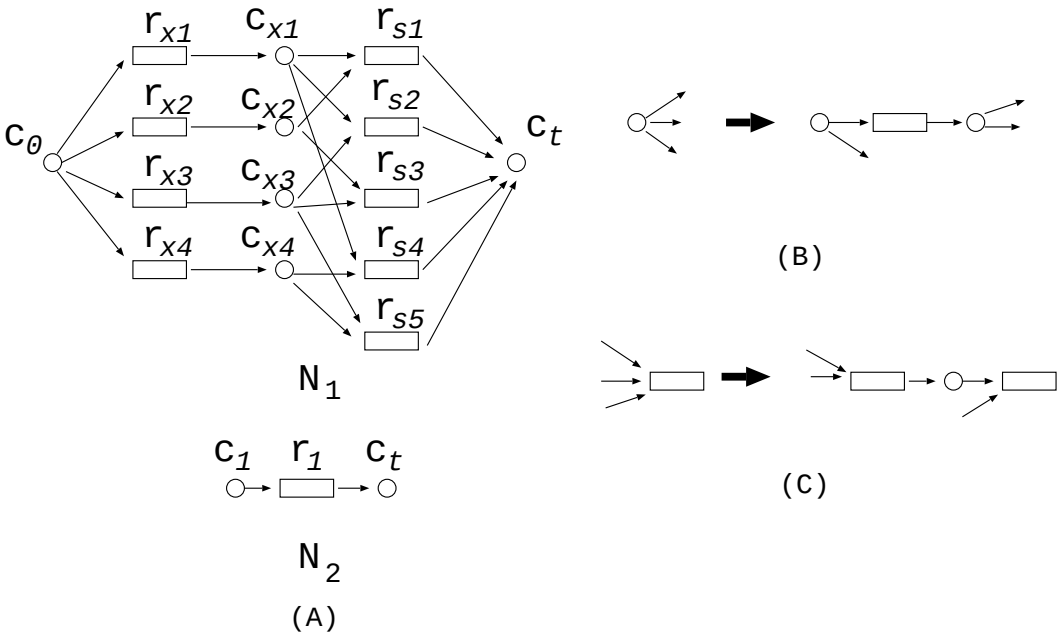

r_{x1}
c_{x1}
r_{s1}
r_{x2}
c_{x2}
r_{s2}
c_0
c_t
r_{x3}
c_{x3}
r_{s3}
r_{x4}
c_{x4}
r_{s4}
r_{s5}
N_1
c_1
r_1
c_t
N_2
(A)
(B)
(C)

Figure 6:

Supplemental data

S1: Relationship between inhibited reactions and predecessors of C00022 in N_1 and N_2 in Dataset 1.

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00200	N_1	N_2	-	-	-
R00703	N_1	N_2	-	-	-
R05605	N_1	-	-	-	ko
R04779	-	-	-	ko	ko
R01541	-	-	-	ko	-
size	3	2	no solution	2	2

S2: Relationship between inhibited reactions and predecessors of C00074 in N_1 and N_2 in Dataset 1.

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00658	N_1	N_2	-	-	-
R04779	-	-	-	ko	ko
R01541	-	-	-	ko	-
size	1	1	no solution	2	1

S3: Relationship between knocked-out nodes and predecessors of all the above 5 compounds are targets in Data1

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00200	N_1	N_2	-	-	-
R00658	N_1	N_2	-	-	-
R00703	N_1	N_2	-	-	-
R00710	N_1	N_2	-	-	-
R05605	N_1	-	-	-	ko
R00351	-	N_2	-	-	-
R04779	-	-	-	-	ko
size	5	5	no solution	no solution	2

S4: Relationship among inhibited reactions and predecessors of C00022 in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00200	N_1	N_2	ko	ko	-
R00220	N_1	N_2	-	-	-
R00214	N_1	-	ko	ko	ko
R00216	N_1	-	ko	ko	ko
R01196	N_1	-	ko	ko	-
R05605	N_1	-	ko	ko	ko
R00586	-	-	ko	-	-
R00945	-	-	ko	-	-
R01512	-	-	-	-	ko
size	6	2	7	5	4

S5: Relationship among inhibited reactions and predecessors of C00024 in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00230	N_1	N_2	ko	-	-
R00238	N_1	-	ko	-	-
R01196	N_1	-	ko	ko	ko
R00209	-	N_2	-	-	-
R00351	-	N_2	-	-	-
R00352	-	N_2	-	-	-
size	3	4	3	1	1

S6: Relationship among inhibited reactions and predecessors of C00033 in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00315	N_1	N_2	-	-	-
R01196	-	-	-	ko	ko
size	1	1	no solution	1	1

S7: Relationship among inhibited reactions and predecessors of C00036 in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00345	N_1	N_2	-	-	ko
R00355	N_1	-	-	-	-
R00351	-	N_2	-	-	-
R00352	-	N_2	-	-	-
size	2	3	no solution	no solution	1

S8: Relationship among inhibited reactions and predecessors of C00074 in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00658	N_1	N_2	-	-	-
R00199	N_1	-	-	-	ko
R00206	N_1	-	-	-	ko
R01015	-	-	-	-	ko
R01829	-	-	-	-	ko
R04533	-	-	-	-	ko
size	3	1	no solution	no solution	5

S9: Relationship among inhibited reactions and predecessors of all the above 5 compounds are targets in Dataset 2

Node	Predecessors		IP-FVS1	IP-FVS2	IP-EM
R00200	N_1	N_2	-	-	-
R00220	N_1	N_2	-	-	-
R00230	N_1	N_2	-	-	-
R00315	N_1	N_2	-	-	-
R00345	N_1	N_2	-	-	-
R00658	N_1	N_2	-	-	-
R00199	N_1	-	-	-	-
R00206	N_1	-	-	-	-
R00214	N_1	-	-	-	ko
R00216	N_1	-	-	-	ko
R00238	N_1	-	-	-	-
R00355	N_1	-	-	-	-
R01196	N_1	-	-	-	-
R05605	N_1	-	-	-	ko
R00209	-	N_2	-	-	-
R00351	-	N_2	-	-	-
R00352	-	N_2	-	-	-
R01512	-	-	-	-	ko
size	14	9	no solution	no solution	4

S10: Relationship among inhibited reactions and predecessors of C00022 in Dataset 3

Node	Predecessors		IP-FVS1	IP-FVS1-approx(10)	IP-FVS2	IP-FVS2-approx(10)
R00200	N_1	N_2	ko	ko	-	ko
R00212	N_1	N_2	ko	ko	-	-
R00220	N_1	N_2	ko	ko	-	-
R00703	N_1	N_2	ko	ko	-	-
R00782	N_1	N_2	ko	ko	-	-
R02320	N_1	N_2	ko	ko	-	-
R05636	N_1	N_2	ko	ko	-	-
R00214	N_1	-	ko	ko	-	-
R00470	N_1	-	ko	ko	-	ko
R00704	N_1	-	ko	ko	-	-
R03105	N_1	-	-	ko	-	-
R05605	N_1	-	ko	ko	-	-
R00258	-	N_2	-	-	-	-
R00896	-	-	ko	-	-	-
R01220	-	-	-	-	-	ko
size	12	8	12	12	NA	3

S11: Relationship among inhibited reactions and predecessors of C00024 in Dataset 3

Node	Predecessors		IP-FVS1	IP-FVS1-approx(10)	IP-FVS2	IP-FVS2-approx(10)
R00212	N_1	N_2	ko	ko	-	ko
R00228	N_1	N_2	ko	ko	-	-
R00230	N_1	N_2	ko	ko	-	-
R05351	N_1	N_2	ko	ko	-	ko
R00238	N_1	-	ko	ko	-	-
R01177	N_1	-	ko	ko	-	-
R04386	N_1	-	ko	ko	-	-
R00351	-	N_2	-	-	-	-
R03026	-	-	-	-	-	ko
size	7	5	7	7	NA	3

S12: Relationship among inhibited reactions and predecessors of C00033 in Dataset 3

Node	Predecessors		IP-FVS1	IP-FVS-1-approx(10)	IP-FVS2	IP-FVS2-approx(10)
R00315	N_1	N_2	-	-	-	-
R00317	N_1	N_2	-	-	-	-
R00710	N_1	N_2	-	-	-	-
R00212	-	-	-	-	-	ko
R00238	-	-	-	-	-	ko
R05351	-	-	-	-	-	ko
size	3	3	no solution	no solution	NA	3

S13: Relationship among inhibited reactions and predecessors of C00036 in Dataset 3

Node	Predecessors		IP-FVS1	IP-FVS1-approx(10)	IP-FVS2	IP-FVS2-approx(10)
R00345	N_1	N_2	ko	ko	-	ko
R00355	N_1	N_2	ko	ko	-	-
R00357	N_1	N_2	-	ko	-	-
R00351	-	N_2	-	-	-	-
R00483	-	-	ko	-	-	-
R00485	-	-	-	-	-	ko
size	3	4	3	3	NA	2