Vol.2012-ICS-165 No.7
2012/1/12

# 極小活性パスウェイの列挙を用いた大腸菌における遺伝子欠損の影響予測

宋剛秀[†1]　井上克巳[†2]　馬場知哉[†1]
高田豊行[†3]　城石俊彦[†3]

本稿では KEGG や Ecocyc などの生物学的知識データベースを利用し，遺伝子欠損が細胞の生育に与える影響を予測する方法を提案する．提案手法では遺伝子が欠損された代謝パスウェイをデータベースから構築し，与えられた初期代謝物から目標代謝物を生成する極小なパスウェイを列挙することによって遺伝子欠損影響の予測を行う．予測結果は大腸菌の単一遺伝子欠損株の生育データを含む慶應コレクションと比較を行い，結果として解糖系における 3 つの必須遺伝子を予測することに成功した．

# Predicting Gene Knockout Effects on *E. coli* by Minimal Active Pathway Enumeration

Takehide Soh,[†1] Katsumi Inoue,[†2] Tomoya Baba,[†1]
Toyoyuki Takada[†3] and Toshihiko Shiroishi[†3]

In this paper, we propose a method to predict gene knockout effects for the cell growth by utilizing biological databases such as KEGG and Ecocyc. We construct metabolic pathways with knockout genes from the two databases and predict gene knockout effects by enumerating all minimal active pathways, which are minimal subsets of a given network using source metabolites to produce target metabolites. To evaluate our method, we apply it to the glycolysis pathway on *Escherichia coli*. In the results, our method predicts three out of four essential genes, which are confirmed by the Keio collection containing comprehensive cell growth data obtained from biological experiments.

## 1. Introduction

It is a biologically important subject to reveal the function of genes, which affect the phenotype of organisms. For model organisms such as *Escherichia coli* (*E. coli*), it has been approached by various methods. For instance, construction of gene knockout organisms is such a method [1,3]. However, it is generally taking high costs and limited by target genes and organisms.

In this paper, we propose a computation method to predict gene knockout effects by identifying *active pathways*, which are sub-pathways that produce target metabolites from source metabolites. In particular, we focus on *minimal active pathways*, which are proposed by Soh and Inoue [7] and have the property of not containing any other active pathways. In other words, all elements of each minimal active pathway are qualitatively essential to produce target metabolites. To predict gene knockout effects by the enumeration of minimal active pathways, we first introduce *extended pathways* that include relations between enzymatic reactions and genes. Then, we formalize a problem to find minimal active pathways on the extended pathway with gene knockouts. After computing the solutions of the problem, our prediction is completed by collecting minimal active pathways that are still active under given gene knockouts.

For evaluation, we have comparisons between our prediction and the cell growth of every single gene knockout *E. coli* strain, which are obtained from the Keio collection [1]. In the experiments, we apply our method to the glycolysis pathway on *E. coli* and compute which gene knockouts affect its cell growth.

## 2. Extended Pathways

To represent metabolic pathways, we use bipartite directed graph representation as follows. Let $M$ be a set of metabolites and $R$ a set of reactions. For $M$ and $R$, $M \cap R = \emptyset$

†1 新領域融合研究センター
Transdisciplinary Research Integration Center
†2 国立情報学研究所
National Institute of Informatics
†3 国立遺伝学研究所
National Institute of Genetics

holds. Let $A_M \subseteq (R \times M) \cup (M \times R)$ a set of arcs. Let $m, r$ be a metabolite and a reaction such that $m \in M$ and $r \in R$, respectively. A *metabolic pathway* is represented in a directed bipartite graph $\mathcal{G}_{\mathcal{M}} = (M \cup R, A_M)$ where $M$ and $R$ are two sets of nodes, $A_M$ is a set of arcs. Besides the metabolic pathway, we consider relations between enzymatic reactions and genes. Let $G$ be a set of genes and $A_G$ be a set of arcs such that $A_G \subseteq (G \times R)$. That is, $A_G$ represents relations between enzymatic reactions and corresponding genes. Let $N$ be a set of nodes such that $N = M \cup R \cup G$ and $A$ be a set of arcs such that $A = A_M \cup A_G$. Then, the *extended pathway* is represented in a directed graph $\mathcal{G} = (N, A)$. We will explain the interpretation of the extended pathway in detail in the next section.

## 3. Minimal Active Pathways with Gene Knockouts

While the minimal active pathway is defined only on metabolic pathways in the literature [7], we explain the definition of the minimal active pathway on the extended pathway in this section. We here define $M_S \subset M$ as a set of source metabolites and $M_T \subset M$ as a set of target metabolites such that $M_S \cap M_T = \emptyset$. An *extended pathway instance* is represented in a four tuple $\pi = (N, A, M_S, M_T)$, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. Let $K$ be a set of knockout genes that are disabled in a given pathway. A *knockout instance* is represented in a five tuple $\pi_K = (N, A, M_S, M_T, K)$. If $K = \emptyset$ then $\pi_K$ corresponds to $\pi$.

A metabolite $m \in M$ is called a *reactant* of a reaction $r \in R$ when there is an arc $(m, r) \in A$. On the other hand, a metabolite $m \in M$ is called a *product* of a reaction $r \in R$ when there is an arc $(r, m) \in A$. Besides, a gene $g \in G$ is called a *corresponding gene* of a reaction $r \in R$ when there is an arc $(g, r) \in A$. A reaction is called a *reversible reaction* if it can occur in both directions between reactants and products. In this paper, we distinguish a reversible reaction as two reactions. For instance, if there is a reversible reaction $r_1$ that has $m_1$ and $m_2$ as reactants and $m_3$ and $m_4$ as products. In this case, we split the reaction $r_1$ into two reactions $r_{1a}$ and $r_{1b}$ such that one of them has $m_1$ and $m_2$ as products and $m_3$ and $m_4$ as reactants.

Let $s : R \to 2^M$ be a mapping from a set of reactions to a power set of metabolites such that $s(r) = \{m \in M \mid (m, r) \in A\}$ represents the set of metabolites that are needed to turn the reaction $r$ activatable. Let $p : R \to 2^M$ be a mapping from a set of reactions to a power set of metabolites such that $p(r) = \{m \in M \mid (r, m) \in A\}$ represents the set of metabolites that are produced by the reaction $r$. Let $c : R \to 2^G$ be a mapping from a set of reactions to a power set of genes such that $c(r) = \{g \in G \mid (g, r) \in A\}$ represents the set of genes that are corresponding genes of the reaction $r$. Let $s' : M \to 2^R$ be a mapping from a set of metabolites to a power set of reactions such that $s'(m) = \{r \in R \mid (m, r) \in A\}$. Let $p' : M \to 2^R$ be a mapping from a set of metabolites to a power set of reactions such that $p'(m) = \{r \in R \mid (r, m) \in A\}$. Let $c' : G \to 2^R$ be a mapping from a set of genes to a power set of reactions such that $c'(g) = \{r \in R \mid (g, r) \in A\}$.

Let $t$ be an integer variable representing time. In this paper, the time is used to represents order relation between reactions to produce target metabolites from source metabolites. In the following, we explain important notions related to production of metabolites, activation of reactions, and expression of genes. Since we focus on gene knockouts, we suppose that almost all genes exist in the cell of a given organism. We also suppose that if genes exist then they are expressed and available to construct enzymes needed for enzymatic reactions. The reason of this condition is that we want to simulate how the lack of corresponding genes affects metabolic pathway rather than how the existence of genes affects other elements. Although our pathway modeling is simple, it allows us to analyze a whole cell scale pathway. Let $\pi_K = (N, A, M_S, M_T, K)$ be a knockout instance, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. Let $\mathcal{G} = (N, A)$ be an extended pathway. Let $M' \subset M$ be a subset of metabolites. A metabolite $m \in M$ is obviously *producible* at time $t = 0$ from $M'$ on $\mathcal{G}$ if $m \in M'$ holds. A reaction $r \in R$ is *activatable* at time $t > 0$ from $M'$ on $\mathcal{G}$ if following two conditions are satisfied: (i) for every $m \in s(r)$, $m$ is producible at time $t - 1$ from $M'$, (ii) at least one corresponding gene $g \in c(r)$ is not included in $K$. A metabolite $m \in M$ is *producible* at time $t > 0$ from $M'$ on $\mathcal{G}$ if there is at least one activatable reaction $r$ at time $t$ such that $m \in p(r)$. If $r$ is activatable at time $t$ then $r$ is activatable at a time $t + 1$. If $m$ is producible at time $t$ then $m$ is producible at time $t + 1$.

Let $\mathcal{G}' = (N', A')$ be a sub-graph of $\mathcal{G}$, where $N' = M' \cup R' \cup G'$ and $A' = A'_M \cup A'_G$. Then, an active pathway of $\pi_K$ is defined as follows.

**Definition1** Active Pathway of Knockout Instance

A bipartite directed graph $\mathcal{G}'$ is an *active pathway* of $\pi_K$ if it satisfies the following conditions:

- $M_T \subset M'$
- $M' = M_S \cup \{m \in M \mid (m,r) \subseteq A, r \in R'\} \cup \{m \in M \mid (r,m) \subseteq A, r \in R'\}$
- $A' = \{(m,r) \in A \mid r \in R'\} \cup \{(r,m) \in A \mid r \in R'\} \cup \{(g,r) \in A \mid g \notin K, r \in R'\}$
- $G' = \{g \in G \mid (g,r) \in A', r \in R'\}$
- For every $m \in M'$, $m$ is producible from $M_S$ on $\mathcal{G}'$

From Definition 1, active pathways include a set of metabolites, reactions, and genes, which are producible and activatable from $M_S$ on $\mathcal{G}'$ such that all target metabolites $M_T$ are producible. The number of active pathways depends on the combination of $M_S$ and $M_T$ but there are generally a large number of active pathways in an extended pathway. We thus particularly focus on minimal ones rather than active pathways. We give the definition of minimal active pathways of $\pi_K$ as follows. Let $\mathcal{G}$ and $\mathcal{G}'$ be extended pathways. We say that $\mathcal{G}$ is *smaller* than $\mathcal{G}'$ and represented in $\mathcal{G} \subset \mathcal{G}'$ if $R \subset R'$. An active pathway $\mathcal{G}$ is *minimal active pathway* of $\pi_K$ *iff* there is no active pathway of $\pi_K$ which is smaller than $\mathcal{G}$. As this definition shows, we only need to see sets of reactions to compare two pathways. Thus, in the rest of this paper, we sometimes represent a minimal active pathway as a set of reactions.

Any reactions included in a minimal active pathway cannot be deleted to produce target metabolites. Although all minimal active pathways are considered to be candidates, ones including a large number of reactions are considered to be biologically inefficient in practice. We thus introduce a time limitation $z$ and pathways that can make all target metabolites producible by $t = z$. In the following, we consider the problem to find minimal active pathways with respect to $\pi_K$ and $z$.

## 4. Knockout Effects

This section provides how to predict knockout effects. In the following, we give some definitions for the prediction. Let $\pi = (N, A, M_S, M_T)$ and $\pi_K = (N, A, M_S, M_T, K)$ be an extended pathway instance and a knockout instance, respectively. Besides, we denote the number of minimal active pathways of $\pi$ as $|\pi|$ and the number of minimal active pathways of $\pi_K$ as $|\pi_K|$. Obviously, $|\pi_K| \leq |\pi|$ holds. Let $K_a$ and $K_b$ be sets of knockout genes respectively. If $|\pi_{K_a}| \leq |\pi_{K_b}|$ holds then we say that the knockout effect of $K_a$ has stronger effect than the knockout effect of $K_b$. If $|\pi_K| = 0$ then we say that the knockout effect of $K$ is *critical* to produce target metabolites. Various metabolites are known as vital metabolites, which means organisms cannot survive without them. That is, if some gene knockout is critical to produce such metabolites, then a given organism cannot grow any more or die. If $|K| = 1$ and its effect is critical to produce vital metabolites then we say that the gene $g \in K$ is *essential*.

In addition to the number of remaining minimal active pathways after knockouts, an important factor of the prediction is the gain of ATPs. This is because inefficient pathways in terms of energy consumption will not be used in organisms. In particular, it is important when we consider the pathways of glycolysis since a main function of glycolysis is to generate ATPs. Thus, the following number is considered to measure the knockout effects: the number of pathways, which gain ATPs or consume less ATPs, lost by $K$. Obviously, critical knockouts always have the strongest effect.

## 5. Computation Method

This section provides how to compute $|\pi_K|$. In this paper, we use the method of computing all minimal active pathways of $\pi$ proposed by Soh and Inoue [7]. This method computes pathways through propositional encoding and minimal model generation. An advantage is that this method is flexible for adding biological constraints. Moreover, we can utilize SAT technologies, which have been developed actively in recent years.

In the following, we briefly explain the propositional encoding to compute minimal active pathways of $\pi$. Let $i$, $j$ be integers denoting indices for metabolites and reactions. Let $t$ be an integer variable representing time. Let $\pi = (N, A, M_S, M_T)$ be an extended pathway instance, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. We introduce two kinds of propositional variables. Let $m^*_{i,t}$ be a propositional variable which is *true* if a metabolite $m_i \in M$ is producible at time $t$. Let $r^*_{j,t}$ be a propositional variable which is *true* if a reaction $r_j \in R$ is activatable at time $t$.

The encoding of the problem to find minimal active pathways with respect to $\pi_K$ and $z$ is as follows.

Vol.2012-ICS-165 No.7
2012/1/12

$$\psi_1 = \bigwedge_{0\le t<z}\bigwedge_{m_i\in M}\left(m^*_{i,t}\to m^*_{i,t+1}\right), \qquad \psi_2 = \bigwedge_{0\le t<z}\bigwedge_{r_j\in R}\left(r^*_{j,t}\to r^*_{j,t+1}\right)$$

$$\psi_3 = \bigwedge_{1\le t\le z}\bigwedge_{r_j\in R}\left(r^*_{j,t}\to \bigwedge_{m_i\in s(r_j)} m^*_{i,t-1}\right), \quad \psi_4 = \bigwedge_{1\le t\le z}\bigwedge_{r_j\in R}\left(r^*_{j,t}\to \bigwedge_{m_i\in p(r_j)} m^*_{i,t}\right)$$

$$\psi_5 = \bigwedge_{m_i\in(M\setminus M_S)}\bigwedge_{1\le t\le z}\left(m^*_{i,t}\to m^*_{i,t-1}\vee \bigvee_{r_j\in p'(m_i)} r^*_{j,t}\right)$$

$$\psi_6 = \bigwedge_{m_i\in M_S} m^*_{i,0}\wedge \bigwedge_{m_{i'}\in M\setminus M_S}\neg m^*_{i',0}, \qquad \psi_7 = \bigwedge_{m_i\in M_T} m^*_{i,z}$$

The formulas $\psi_1$ and $\psi_2$ represent that once a metabolite (or a reaction) is changed to producible (or activatable), then it remains in the producible state (or the activatable state). The formula $\psi_3$ represents that if a reaction $r_j$ is activatable at time $t$ then its reactants must be producible at time $t-1$. The formula $\psi_4$ represents that if a reaction $r_j$ is activatable at time $t$ then its products must be producible at time $t$. The formula $\psi_5$ represents that if a reaction $m_i$ is producible then either two states hold: the metabolite $m_i$ is producible at $t-1$ or at least one reaction $r_j$ is activatable. The formulas $\psi_6$ and $\psi_7$ represent source metabolites and target metabolites. We denote the conjunction of $\psi_1,\ldots,\psi_7$ as $\Psi_z$. Then, we can enumerate minimal active pathways of $\pi$ by computing minimal models of $\Psi_z$ with respect to $V^z=\{r^*_{i,z}|r_i\in R\}$.

The computation for $\pi$ is always necessary to have comparison between a wild type cell and its mutant. We then explain a method to compute all minimal active pathways of $\pi_K$ for a set of knockout genes $K$ from $|\pi|$. Actually, when the minimal active pathways of $\pi$ are obtained, we do not need much additional computation. The all minimal active pathways of $\pi_K$ are obtained by selecting pathways that do not contain some $r\in R_K$, where $R_K=\{r\in c'(g)\mid g\in K\}$. The procedure is given as follows: (i) Enumerate all minimal active pathways with respect to $\pi$ and $z$, (ii) Delete minimal active pathways including some $r\in R_K$, where $R_K=\{r\in c'(g)\mid g\in K\}$. Except the above procedure, there is another way to compute all minimal active pathways with respect to $\pi_K$ and $z$. The same is achieved by adding constraints, which inhibit the activation of each reaction in $R_K$, to the formula $\Psi_z$.

## 6. Experimental Results

At first, we show experimental conditions. We construct extended pathways from EcoCyc [5] and KEGG [4]. Specifically, we use EcoCyc to construct metabolic pathways, which consists of 1222 metabolites and 1920 reactions. Besides, we use KEGG to construct relations between enzymatic reactions and genes. In the following experiments, we use the entire extended pathways constructed from the two databases. Each experiment has been done using a PC (3.2GHz CPU) running on OS X 10.6. For computation, we use a SAT solver Minisat2 [2]. Koshimura *et al.* propose a procedure computing minimal models with SAT solvers [6]. We follow their procedure based on SAT solvers to generate minimal models.

To evaluate our method, we use the Keio collection [1]. In particular, we use their result on the MOPS medium whose main nutrient is glucose. Since this comparative data is obtained from every single gene knockout, in the following, we basically consider that the set of knockout genes $K$ consists of one gene. Moreover, in the Keio collection, if a cell growth is less than 0.1 or not applicable (N. A.) then we say that the cell is strongly affected by a gene knockout.

According to the MOPS medium of the Keio collection [1], we choose source metabolites as follows: D-glucose-6-phosphate, $H^+$, $H_2O$, ATP, ADP, phosphate and $NAD^+$. In addition, pyruvate is given as the target metabolite to analyze glycolysis.

We then compute all minimal active pathways from the entire metabolic pathway of *E. coli*. As is known in Biology, the reference pathway glycolysis is constructed by 8 steps. However, if some reactions are disabled then *E. coli* is expected to use other bypass pathway by additional reactions. In this experiment, we thus give $z=12$. Besides, the number of reactions included in each pathway is limited to less than or equal to 12.

At first, we compute all minimal active pathways using above conditions and obtained 75 minimal active pathways. We then connect 61 genes to reactions included in them by information on KEGG. Secondly, we compute minimal active pathways of for each gene knockout. This experiment is completed within 4 seconds. Figure 1 shows the result of 61 gene knockouts. The x-axis denotes each gene knockout and the y-axis denotes the number of minimal active pathways. In the figure, we compute minimal active pathways

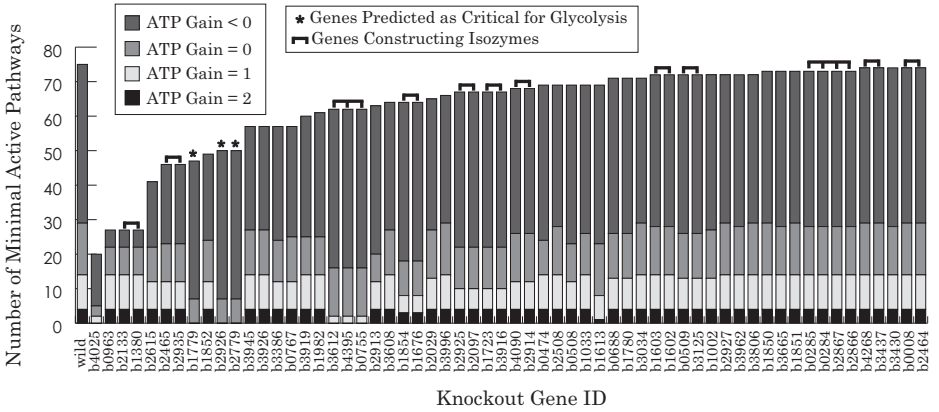


Fig. 1 The Number of Minimal Active Pathways for each Gene Knockout on Glycolysis

of $\pi_{K_1}, \ldots, \pi_{K_{61}}$ such that $K_1 = \{b4025\}, K_2 = \{b0963\}, \ldots, K_{61} = \{b2464\}$. However, since some of 61 genes construct isozymes, such single gene knockout $K_i$ does not affect the number of minimal active pathways $|\pi_{K_i}|$. However, for reference, we compute the effect of the gene knockouts that disables all of them. For instance, b2133 and b1380 construct isozymes. In this case, the number of minimal active pathways in the figure shows the case of the gene knockout of both b2133 and b1380. For each gene knockout, we compute the gain of ATPs in each minimal active pathway, which is calculated by counting the number of both reactions with the coefficient of ATP: ones consuming ATP and the other ones producing ATP. Minimal active pathways that produce the positive number of ATPs are more important than others because producing ATP is a main function of glycolysis.

From the figure, we can read that *E. coli* keeps almost all minimal active pathways even by more than half of single gene knockouts. It is considered to indicate the robustness of *E. coli*. However, some gene knockouts dramatically reduces the number of minimal active pathways. In particular, single gene knockouts of b1779, b2926 and b2779 destroy all minimal active pathways that gain the positive number of ATPs. In our prediction method, those gene knockouts are expected to have the strongest effect for the cell growth of *E. coli*.

In order to evaluate the above predictions, we have comparisons with the Keio collection. Table 1 shows the comparison for the first 11 gene knockouts with regard to the number of lost minimal active pathways. Column 1, Gene ID, shows identifiers of genes except *wild*, which denotes an empty set of knockout genes, e.g., $K = \{\}$. Other rows denote the result of single gene knockout. Column 2, Total, shows the total number of minimal active pathways, i.e., $|\pi_{K_i}|$. Columns 3 to 6 show the number of minimal active pathways, each of which denotes the gain of ATPs. Column 7, MOPS24hr, and Column 8, MOPS48hr, show the cell growth of *E. coli* after 24 hours and 48 hours, respectively. Note that N.A. means "not applicable" because of essential gene [1]. As the first row of Table 1 shows, we found 14 minimal active pathways that produce the positive number of ATPs on the wild type cell of *E. coli* while there are 75 in total.

Table 1 11 Single Gene Knockouts for Glycolysis

| Gene ID | # of Minimal Active Pathways | | | | | Keio Collection [1] | |
|---|---|---|---|---|---|---|---|
| | Total | ATP Gain | | | | MOPS24hr | MOPS48hr |
| | | 2 | 1 | 0 | <0 | | |
| wild | 75 | 4 | 10 | 15 | 46 | 0.219-0.392 | 0.216-0.480 |
| b4025 | 20 | 0 | 2 | 3 | 15 | 0.137 | 0.542 |
| b0963 | 27 | 4 | 10 | 8 | 5 | 0.293 | 0.371 |
| b2133[a] | 27 | 4 | 10 | 8 | 5 | 0.303 | 0.366 |
| b1380[a] | 27 | 4 | 10 | 8 | 5 | 0.357 | 0.393 |
| b2615 | 41 | 4 | 8 | 10 | 19 | N.A. | N.A. |
| b2465[b] | 46 | 4 | 8 | 11 | 23 | 0.311 | 0.315 |
| b2935[b] | 46 | 4 | 8 | 11 | 23 | 0.317 | 0.327 |
| b1779* | 47 | 0 | 0 | 7 | 40 | N.A. | N.A. |
| b1852 | 49 | 4 | 8 | 12 | 25 | 0.231 | 0.223 |
| b2926* | 50 | 0 | 0 | 7 | 43 | N.A. | N.A. |
| b2779* | 50 | 0 | 0 | 7 | 43 | N.A. | N.A. |

Distinguished single gene knockouts are b1779, b2926 and b2779; they destroy all minimal active pathways that gain ATPs. It predicts that they have the strongest effect for the cell growth *E. coli* because they disable an important function, ATP production. For this prediction, the Keio collection shows "N.A" for each gene knockout. Thus, in glycolysis, our method successfully predicts the gene knockout effects.

On the other hand, our method predicts there are still minimal active pathways that produce the positive number of ATPs after the single gene knockouts of b4025, b0963 and b1852. Those remaining pathways are supposed to be used as bypass pathways. For instance, b4025 encoding glucosephosphate isomerase gene of glycolysis pathway that

transfer D-glucose-6-phosphate to D-fluctose-6-phosphate. However, pentose phosphate pathway is available as bypass pathway from D-glucose-6-phosphate, resulted in the gene knockout slow-growth at starting MOPS24hr and same level of wild type final growth at MOPS48hr. Besides them, considering the knockouts of b2133, b1380, b2465 and b2935, they do not affect to the cell growth since those construct isozymes.

The single gene knockout of b2615 is different to the above gene knockouts. Our method predicts that this knockout does not affect to the cell growth in terms of glycolysis. However, the Keio collection shows that this is essential gene for *E. coli*. One assumption is that it affects other function in the cell. To explore this, to have experiments on other target metabolites such as amino acids is an important future topic.

## 7. Conclusion and Future Work

In this paper, we propose a method to predict the knockout effect by enumerating minimal active pathways and the gain of ATPs. We formalize the extended pathways and show the definition of minimal active pathways on it. In addition, we presented a computation method for the prediction. An advantage of our method is capable to treat a whole cell scale metabolic pathway. Besides, our method allows us to trace the reason of the prediction results, e.g., we can suggest the reason of the essentiality of three genes in the glycolysis pathway. This is an important feature that other methods do not have.

As far as future work is concerned, to apply our method to other organism such as mouse is a very important topic. In addition to *E. coli*, mouse is a well known model organism for human study and its available information has been accumulated in the last decade. In particular, chromosome substitution strains are used to reveal the function of genes [8]. In addition to gene knockouts, to adapt our method to such strains is an interesting topic. Although there is a large difference between *E. coli* and mouse, the basic metabolism is same. This fact tells us that our method can also be a potential prediction method to mouse.

## Acknowledgements

This research is supported in part by the 2008-2011 JSPS Grant-in-Aid for Scientific Research (A) No. 20240016, and the research project of Systems Biology of Genetic Function and the 2nd Transdisciplinary Seeds Research in the Transdisciplinary Research Integration Center.

## References

1) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H.: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Molecular Systems Biology*, Vol.2, No.2006.0008 (2006).
2) Eén, N. and Sörensson, N.: An Extensible SAT-solver, *Proceedings of the 6th International Conference on Theory and Applications of Satisfiability Testing (SAT 2003)*, pp.502–518 (2003).
3) Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H. and Tomita, M.: Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations, *Science*, Vol.316, No.5824, pp.593–597 (2007).
4) Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M.: KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Research* (2011).
5) Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., LuisMu 単 iz-Rascado, C 辿 sarBonavides-Martinez, Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A.G., Mackie, A., Paulsen, I., Gunsalus, R.P., Karp, P.D.：EcoCyc: a comprehensive database of *Escherichia coli* biology, *Nucleic Acids Research*, Vol.39, No.suppl 1, pp.D583–D590 (2011).
6) Koshimura, M., Nabeshima, H., Fujita, H. and Hasegawa, R.: Minimal model generation with respect to an atom set, *Proceedings of the the 7th International Workshop on First-Order Theorem Proving (FTP '09)*, pp.49–59 (2009).
7) Soh, T. and Inoue, K.: Identifying Necessary Reactions in Metabolic Pathways by Minimal Model Generation, *Proceedings of ECAI 2010*, pp.277–282 (2010).
8) Takada, T., Mita, A., Maeno, A., Sakai, T., Shitara, H., Kikkawa, Y., Moriwaki, K., Yonekawa, H. and Shiroishi, T.: Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits, *Genome Research*, Vol.18, No.3, pp.500–508 (2008).