

SAT 問題への変換を用いたフィードバックを含むパスウェイの解析

Analyzing Pathways through a Translation into SAT Problems

宋 剛秀*¹ 井上 克巳*^{2,1}

Takehide Soh

Katsumi Inoue

*¹総合研究大学院大学

The graduate university for advanced studies

*²国立情報学研究所

National Institute of Informatics

Recent advance of satisfiability (SAT) technologies has been tremendous and several problems such as planning, scheduling and packing problems are successfully solved with SAT solvers. In this paper, we show a SAT-based approach for analyzing biochemical pathways. We review a previously proposed approach where a weighted Max-SAT solver was used to compute possible reaction states in biochemical pathways. We show how this problem can be solved using a SAT solver. In our approach, we translate a given biochemical pathway into a SAT problem. One of the advantages of this method is that it generates only feasible solutions by utilizing minimal models. In the experiments, we attempt to compute reaction states in the glycolytic pathway and compare them with the previous result from the literature.

1. はじめに

生命現象を司る細胞の活動は数万種類を越える分子相互間の化学反応によって維持されている。例えば人間の個性を特徴付ける上で重要な遺伝情報の伝達や、生命の活動に必要なエネルギー代謝も化学反応に基づいて行われる。近年このような化学反応に基づいた生命現象をシステムとして理解する試みとして、システム生物学が注目されている。システム生物学において、細胞などミクロなレベルにおける生体の機能は分子の相互作用のネットワーク表現(パスウェイ)によって表現される。それゆえ生命をより深く理解するためにパスウェイを解析することはシステム生物学において重要な課題である。従来の解析手法として連立方程式、ベイジアンネットワーク、整数計画法など多くの手法が提案されている [Schuster *et al.*, 2000, Tamura *et al.*, 2009]。

しかし、これらの手法を用いても実験により得られる入力データの不足等の理由により現実のパスウェイを解析することは依然として困難であり、解析手法にはまだ改善の余地があると考えられる。

本研究ではパスウェイの解析、特に入力される化合物と出力される化合物が与えられた時にパスウェイ内で活性化する反応の集合(経路)を発見することに焦点をあてる。関連する従来研究の1つとして Tiwari らによる Weighted Max-SAT 問題への変換を用いた手法が挙げられる [Tiwari *et al.*, 2007]。彼らの手法において、パスウェイ内の化学反応の活性と非活性は命題変数の真偽値で表され、求めたい経路は重み付きのモデルとして表される。しかし Weighted Max-SAT 問題を使用した変換では必ずしも全ての制約を満たす必要がなく、生物学的に不適切なモデルを出力することがあった。また多くのパスウェイでは2つの反応が1つの種を反応物として消費するような反応を含んでいるが、彼らの変換方法ではこのような反応を制限しており、適切なモデルが重み付けにおける上位に出力されないことがあった。

これらの問題に対して本論文ではパスウェイを充足可能性判定 (SAT) 問題に変換し SAT ソルバーを用いて解析する手

法を提案する。具体的には、まず変換された SAT 問題のモデルを最新の SAT ソルバーである Minisat を用いて求める。次にこのモデルの中で条件を満たす極小モデルのみを最終的なモデルとして出力する。本手法の良い点は3つある。1つめは SAT 問題のモデルが全ての制約を必ず満たすことである。これにより従来手法の問題であった生物学的に不適切なモデルを出力する可能性を削減することができる。2つめは、本手法の変換方法を用いることで1つの種を2つの反応が消費するような経路を含むようなパスウェイに対しても適切なモデルが出力できることである。3つめは、SAT ソルバーは近年活発に研究されており、本手法はその進歩の恩恵を享受することができることである。提案手法は2つの人工的に作成したパスウェイと解糖系のパスウェイを用いて従来手法と評価を行った。

2. パスウェイの論理モデル

本節では文献 [Tiwari *et al.*, 2007] に基づいて議論に必要な用語の説明を行う。種 (species) とは生化学における任意の構成要素を指す。例えば分子、イオン、タンパク質、酵素、配位子、受容体などである。反応 (reaction) とは反応物 (reactant) の集合 R 、修飾因子 (modifier) の集合 M 、生成物 (product) の集合 P で構成される。

反応 r は種の集合 S の互いに素な部分集合 R, M, P の3組で表され、入力として R, M をとり、 P を出力する。ある1つの反応 r が与えられた時、その構成要素となる種 s の集合をそれぞれ $R(r), M(r), P(r)$ で表す。ある1つの種 s が与えられた時、その種を反応物、修飾因子、生成物として使用する反応の集合をそれぞれ $R^{-1}(s), M^{-1}(s), P^{-1}(s)$ で表す。なお反応は活性 (on) と非活性 (off) の2つに近似され、種は発現と欠乏の2つに近似されるものとする。

パスウェイは上記に述べた反応の集合 $N = \{r_1, r_2, \dots, r_n\}$ で表され、その定常状態は on となる反応の集合で表される。入力種 S_{input} はパスウェイに対して入力として与えられる種であり、目標種 S_{target} はパスウェイが出力すべき種である。これら入力種と目標種が付加されたパスウェイをパスウェイ問題という。パスウェイ問題を解くとはパスウェイ内で入力種から目標種に至る反応の集合(経路)を発見することである。

連絡先: 宋剛秀 総合研究大学院大学情報学専攻

〒 101-8430 東京都千代田区一ツ橋 2-1-2

Email: soh@nii.ac.jp

3. SAT 問題と Weighted Max-SAT 問題

リテラルはアトム、もしくはその否定で表される。節はリテラルの選言であり、集合表記も用いる。連言標準形 (conjunctive normal form: CNF) 式とは節の連言であり集合 ψ でも表される。充足可能性判定 (SAT) 問題とは、CNF 式として与えられた命題論理式を充足するようなアトムへの真偽値割り当て (モデル) が存在するかどうかを判定する問題である*1。そのような真偽値割り当てが存在する場合に、その命題論理式は充足可能 (SAT) といい、存在しない場合は充足不能 (UNSAT) であるという。SAT ソルバーは与えられた命題論理式の充足可能性を判定するプログラムであり、機能として充足可能な命題論理式については判定結果とともにそのモデルを返すことができる。

一方、各節 $c \in C$ を非負の整数値をとる重み w との対 (c, w) で表し、充足される節の重み和を最大にするような真偽値割り当てを求める問題を Weighted Max-SAT 問題と呼ぶ。

4. パスウェイの Weighted Max-SAT 問題への変換

本節では Tiwari ら [Tiwari et al., 2007] による Weighted Max-SAT 問題への変換を用いた手法を説明する。この手法においてパスウェイ $N = \{r_1, r_2, \dots, r_n\}$ は n 個の命題変数 $\{b_1, b_2, \dots, b_n\}$ 上の状態遷移システムとして定義される。ここで命題変数 b_i の真偽値 *true*, *false* は反応 r_i の on, off を表す。パスウェイにおける反応間の制約は以下の式により与えられる。

$$b_i \Rightarrow \bigwedge_{s \in R(r_i) \cup M(r_i)} \bigvee_{r_j \in P^{-1}(s)} b_j \quad (1)$$

$$b_i \Rightarrow \bigwedge_{s \in R(r_i)} \bigvee_{r_j \in R^{-1}(s), j \neq i} \neg b_j \quad (2)$$

$$b_i \Rightarrow \bigwedge_{s \in M(r_i)} \bigvee_{r_j \in R^{-1}(s), j \neq i} \neg b_j \quad (3)$$

$$\neg b_i \Rightarrow \neg \bigwedge_{s \in R(r_i) \cup M(r_i)} presentAi(s, i) \quad (4)$$

ここで $presentAi(s, i)$ は以下を表す。

$$\bigvee_{r_j \in P^{-1}(s)} b_j \wedge \bigwedge_{r_j \in R^{-1}(s), j \neq i} \neg b_j$$

各式の意味は以下のように与えられる。

式 (1) ある反応 r_i が on であれば、その反応物と修飾因子を供給する反応を on にする

式 (2) ある反応 r_i が on であれば、 r_i が反応物として使用する種 s を反応物として共有する反応 r_j を抑制する

*1 本論文ではモデルを真に割り当てられたアトムの集合で表すことにする。例えばモデル $a = true, b = true, c = false$ は $M_1 = \{a, b\}$ で表現され、モデル $a = true, b = true, c = true$ は $M_2 = \{a, b, c\}$ で表現される。

式 (3) ある反応 r_i が on であれば、 r_i が修飾因子として使用する種 s を反応物として共有する反応 r_j を抑制する。

式 (4) ある反応 r_i に必要な全ての反応物、修飾因子が発現しており、かつそれらを使用する反応が off であれば、その反応 r_i は on になる

また上記の式に加えて 3 つの制約が必要である。ダミー反応制約とは、どの反応からも供給されることがない種 S_{init} を初期種として追加する際に必要な制約である。ダミー入力種の追加は $R = \emptyset, M = \emptyset, P = S_{init}$ で構成されるダミー反応を追加することにより実現する。ダミー反応は本来存在する反応でないため、できるだけ活性化しないように制約を追加する必要がある。ダミー反応を r とした時に、活性を制限する制約は以下で表される。

$$\neg b \quad (5)$$

入力種・目標種制約とは、パスウェイ問題において発現されるべき種を指定する際に必要な制約であり、以下で表される。

$$\bigvee_{r_i \in P^{-1}(s_{input}) \cup P^{-1}(s_{target})} b_i \quad (6)$$

モード切り替え制約とは、反応を可能な限り活性化させるのか、可能な限り非活性化させるのかを指定するための制約である。可能な限り非活性化させる場合は、ダミー反応を除いた反応に対して以下を追加する。

$$\neg b_i \quad (7)$$

以上が与えられたパスウェイに対する制約である。先述したように Tiwari らによる手法では Weighted Max-SAT 問題を用いるため、各制約はそれぞれ重みを持つ。節への重み付けは以下ようになる。式 (1)、式 (2)、式 (4)、式 (6) で表される節には最も大きい重み W を与える。ここで W は、ダミー反応を含むネットワーク内の反応の総数と等しいものとする。式 (3) で表される節には $\lfloor \frac{W}{2} \rfloor$ を与える。式 (5) で表される節には $\lfloor \frac{W}{k+1} \rfloor$ を与える。ここで k はダミー反応以外に供給反応が存在しない種の総数を表す。式 (7) には重み 1 を与える。

5. パスウェイの SAT 問題への変換

5.1 反応間における制約

現実のパスウェイにはフィードバックや 1 つの種を 2 つが共有するような反応が含まれている。このようなパスウェイにおいて Tiwari らの変換方法では式 (2) と式 (3) において反応物や修飾因子を共有する反応を抑制しているため生物学的立場から見て不適切な経路が出力されることがあった。それに対して提案手法では 1 つの種を 2 つが共有するような反応経路をモデルとして出力するために、制約を以下のように与える。

$$b_i \Rightarrow \bigwedge_{s \in R(r_i) \cup M(r_i)} \bigvee_{r_j \in P^{-1}(s)} b_j \quad (8)$$

$$\neg b_i \Rightarrow \neg \bigwedge_{s \in R(r_i) \cup M(r_i)} presentAi(s, i) \quad (9)$$

ここで $presentAi(s, i)$ は以下を表す。

$$\bigvee_{r_j \in P^{-1}(s)} b_j \wedge \bigwedge_{r_j \in R^{-1}(s), j \neq i} \neg b_j$$

各式の意味は以下のように与えられる。

モデル生成手続き (CNF Ψ , 入力反応 $\{b_{input}\}$)

```

(1) begin
(2)   while (solve( $\Psi$ ))           /*モデル M 生成*/
(3)      $\Psi_{verify} = \Psi \cup \{f_1\} \cup F_2$ ;
(4)     if (!solve( $\Psi_{verify}$ ))      /*極小性検査*/
(5)       if ( $\{b_{input}\} \subset M$ )
(6)         /*極小モデル生成*/
(7)       end if
(8)     end if
(9)      $\Psi = \Psi \cup \{f_1\}$ 
(10)  end while
(11) end
    
```

図 1: SAT ソルバを用いたモデル生成手続き

式 (8) ある反応 r_i が on であれば, その反応物と修飾因子を供給する反応を on にする

式 (9) ある反応 r_i に必要な全ての反応物, 修飾因子が発現しており, かつそれらを使用する反応が off であれば, その反応 r_i は on になる

また目標種 $target$ については, それを生成する反応を必ず活性させるために対応する命題変数を単位節 $\{b_{target}\}$ として加える. ここで注意するのは, 単位節として加えるのは目標種を生成する反応 $\{b_{target}\}$ のみであって入力種 $input$ に対応する反応 b_{input} は加えない. b_{input} については後述するモデル生成手続きの中で使用する. また SAT 問題への変換を用いることで全てのモデルはこの 3 つの制約を必ず充足する. しかし, ただ単に制約を減らすだけでは冗長なモデルが多数出力されることになり, 適切なモデルを発見することが困難となる. この問題を解決するために提案手法では極小モデルに着目する. その生成手続きを次節で説明する.

5.2 モデル生成手続き

ある SAT 問題に対してモデル $M_1 = \{a, b\}$ とモデル $M_2 = \{a, b, c\}$ が得られたとする. 2 つのモデルには包含関係 $M_1 \subset M_2$ が成り立ち, この時に M_1 は M_2 より小さいという. 与えられた SAT 問題に対して M より小さいモデルが存在しない場合に M は極小モデルであるという. 本論文ではこの極小モデルを利用して適切な, すなわち目標種の生成に寄与する反応のみを on にするような経路を求める. 極小モデルを現実の問題へ適用した研究として越村ら [Koshimura et al., 2008] によるジョブショップスケジューリング問題 (JSSP) への適用がある. 彼らは極小モデルを求めることで, 最適化問題である JSSP の最適値を求める方法を報告している. 本節では彼らの手法を応用して前節で生成した SAT 問題から目的とするモデルのみを求める方法を説明する.

CNF 式 Ψ を式 (8), 式 (9) そして目標種に対応する単位節 $\{b_{target}\}$ からなる節で構成されるものとする. 図 1 に提案手法のモデル生成手続きを示す.

まず (2) でサブルーチン solve が呼ばれている. これは SAT ソルバに対して Ψ を与えて充足可能性判定を行わせることを表している. UNSAT であればループを抜け, SAT であれば M にモデルが格納される. (3) では得られたモデルの極小性を判定するために Ψ_{verify} を作成している. アトム集合であるモデル $M = \{m_1, m_2, \dots, m_e\}$ とその補集合 $\bar{M} = \{n_1, n_2, \dots, n_f\}$ が得られた時に f_1 は $\neg m_1 \vee \neg m_2 \vee \dots \vee \neg m_e$ となる節を表しており, F_2 は $\neg n_1 \wedge \neg n_2 \dots \wedge \neg n_f$ となる単位節の集合を表している. ここで得られた Ψ_{verify} のが充足不能であれば M は Ψ に対して極小のモデルである. これは越村ら [Koshimura et al., 2008]

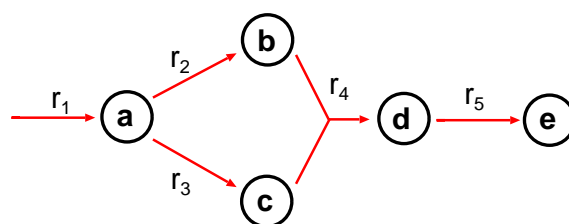


図 2: 2 つの反応が 1 つの種を共有するパスウェイ

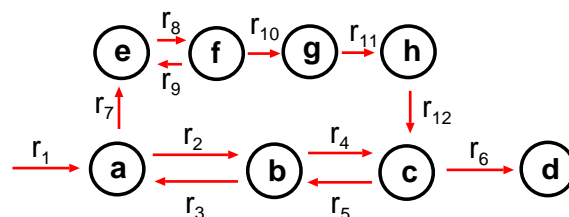


図 3: フィードバックを含むパスウェイ

によって証明されている. (5) では得られた極小モデルが入力種を生成する反応 $\{b_{input}\}$ を包含するかどうかを判定している. ここで $\{b_{input}\}$ を包含していれば, 得られた極小モデルは全て入力種を使用した上で目標種を生成する経路を表す. (9) では $\{f_1\}$ を Ψ に加えている. こうすることで次のループ以降で M より大きいモデルを探索しないようになる. 以上の手続きを (2) で SAT ソルバが UNSAT を返すまで継続することで Ψ の全ての極小モデルおよびパスウェイ経路を求めることができる.

6. 実験によるモデルの評価

6.1 Weighted Max-SAT 問題への変換手法との比較

実験は CPU1.8GHz, メモリ 2GB を搭載した計算機上で行った. まず人工的に作成した 2 つのパスウェイ (図 2, 図 3) に対して, Tiwari ら [Tiwari et al., 2007] の手法と提案手法のモデル比較を行った. Tiwari らの手法ではモデルと共に重みが出力されるが, 今回の実験では最大重みより上位 3 位までの重みを持つモデルを結果として示す. まず初めに図 2 に示す 2 つの反応 r_1, r_2 が 1 つの種 a を共有するパスウェイに対する結果を示す. このパスウェイには入力種として a , 目標種として e を設定する. Tiwari ら [Tiwari et al., 2007] の手法を用いて解析した結果は次のようになる. 重み 74 において 2 つのモデル $\{r_1, r_2, r_4, r_5\}$, $\{r_1, r_3, r_4, r_5\}$, 重み 73 において 2 つのモデル $\{r_1, r_2, r_5\}$, $\{r_1, r_3, r_5\}$, 重み 70 において 1 つのモデル $\{r_1, r_2, r_3, r_4, r_5\}$ が出力される. 最大重み 74 の 2 つのモデルであるが, これらのモデルは 2 つの種 b, c のどちらかを生成しておらず, 反応 r_4 が 2 つの種 b, c を用いて生成されるという制約を無視している. 重み 73 では反応 r_4 が on になっていないにも関わらず, r_5 が on になっており r_5 が種 d を反応物として消費するという制約を無視している. 重み 70 では全ての反応が on になっておりこれは適切なモデルである. このように彼らのモデルにおいて制約は必ず満たされる必要がないため, 適切でないモデルが適切なモデルよりも重みにおいて上位に出力される場合があることが分かる.

次に図 3 に示されるフィードバックを含むパスウェイについての結果を考察する. このパスウェイにおいては入力種として a , 目標種として d を設定する. Tiwari らの手法の結果は次のようになる. 重み 500 において $\{r_1, r_2, r_4, r_6\}$, 重み 498 において $\{r_1, r_2, r_4, r_6, r_8, r_9\}$, 重み 497 において $\{r_1, r_6, r_7, r_8, r_{10}, r_{11}, r_{12}\}$ が出力される.

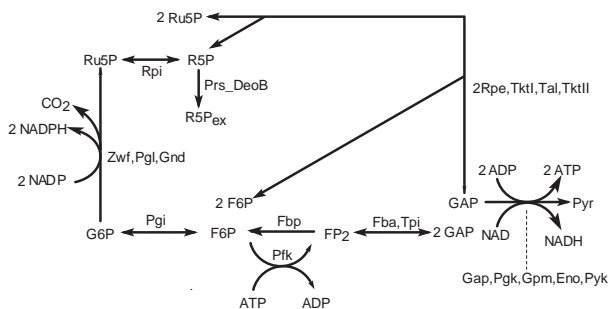


図 4: 解糖系のパスウェイ [Schuster *et al.*, 2000]

最大重み 500 のモデルであるが, a を入力として d を生成する冗長性のない経路であり良いモデルといえる. 重み 498 のモデルであるが $\{r_8, r_9\}$ は目標種 d の生成に寄与せず, 冗長な経路を含んだモデルだといえる. 重み 497 のモデルは目標種 d を生成するための別経路であり, 適切なモデルである.

実験結果より Tiwari らの変換方法の問題点は 2 点あると考えられる. 1 つめは図 2 のパスウェイに対しては不適切なモデルが適切なモデルよりも重みについて上位に出力されることである. 2 つめは図 3 のパスウェイに対しては冗長な経路を含むモデルが適切なモデルよりも上位に出力されることである. まとめて彼らの手法の問題点は不適切なモデルと適切なモデルの重みによる順序づけが適切でないことであるといえる. このため人が判断してモデルを抽出する作業が必要であるが, 大規模なパスウェイに対しては制約の無視や冗長な経路だと判断することは非常に困難である.

上記の結果に対して提案手法の結果を次に示す. 図 2 のパスウェイに対する結果としては 1 つのモデル $\{r_1, r_2, r_3, r_4, r_5\}$ が得られた. また図 3 のパスウェイに対しては 2 つのモデル $\{r_1, r_2, r_4, r_6\}$ と $\{r_1, r_6, r_7, r_8, r_{10}, r_{11}, r_{12}\}$ が得られた. 提案手法は適切なモデルのみを出力しており, これら 2 つのパスウェイに対して提案手法は従来手法より優れた手法であるといえる.

6.2 解糖系パスウェイへの適用

前節では人工的に作成したパスウェイで実験を行ったが, 文献 [Schuster *et al.*, 2000] より引用した実際に人のエネルギー代謝を担うパスウェイの 1 つである解糖系パスウェイ (図 4) に対しても実験を行った. このパスウェイにおける化学反応式を表 1 に示す. 紙面の都合上全ての反応と略記を説明することはできないので, 詳細は文献 [Schuster *et al.*, 2000] を参照されたい. このパスウェイの概要を説明すると図 4 に G6P と記載された種はグルコース 6 リン酸 (glucose 6-phosphate), Pyr と記載された種はピルビン酸 (pyruvate) を表しており, それぞれこのパスウェイにおける入力種と目標種となる.

このパスウェイに対して提案手法が出力したモデルは次の 3 つである:

- $\{r_1, r_3, r_5, r_7, r_{14}, r_{16}, r_{17}\},$
- $\{r_3, r_5, r_7, r_8, r_9, r_{11}, r_{14}, r_{16}, r_{17}\},$
- $\{r_3, r_4, r_7, r_8, r_9, r_{11}, r_{14}, r_{16}, r_{17}\}.$

1 つめのモデルはグルコース 6 リン酸が GAP を経てピルビン酸を生成する経路を表している. 2 つめのモデルは, ピルビン酸を生成する別の経路を表している. この 1 つめのモデルは一般的に認知されている代謝経路であり, 2 つめのモデルとあわせて生物学的に起こり得ることが文献 [Schuster *et al.*, 2000] で述べられている. このことより提案手法が正しい経路を発見していることが分かる. 3 つめのモデルは FP2 と F6P 間

表 1: 解糖系パスウェイにおける反応

反応番号	反応名	反応式
r_1, r_2	Pgi	$G6P = F6P$
r_3	Pfk	$F6P + ATP \rightarrow ADP + FP2$
r_4	Fbp	$FP2 \rightarrow F6P$
r_5, r_6	Fba	$FP2 = GAP$
r_7	Gap,Pgk, Gpm,Eno,Pyk	$GAP + NAD + ADP = Pyr + NADH + 2ATP$
r_8	Zwf,Pgl,Gnd	$G6P + 2NADP \rightarrow NADPH + CO_2 + Ru5P$
r_9, r_{10}	2Rpe,Tk1l, Tal,Tk1ll	$Ru5P + R5P = F6P + GAP$
r_{11}, r_{12}	Rpi	$Ru5P = R5P$
r_{13}	PrsDeoB	$R5P \rightarrow R5P_{ex}$

のフィードバック反応が表れているところが他のモデルと異なる. このフィードバックは冗長なものではなく反応 r_7 の活性に必要な ADP の生成に寄与したものであるが, 実際に起こり得るかどうかは専門家の意見を聞く必要があると考えられる.

7. 結論と今後の課題

本論文では SAT ソルバを用いたフィードバックを含むパスウェイを解析するための方法について述べた. SAT 問題に変換することの利点は, 出力されるモデルが必ず全ての制約を満たしていること, また極小解を求めることで冗長なモデルを出力せずに適切なモデルを出力できることである. 提案手法を解糖系のパスウェイに適用した結果, 適切だと認知されているものを含む 3 つのモデルを出力することができた. しかしパスウェイより深く理解するには文献 [Ray *et al.*, 2009] にみられるように酵素反応等も考慮しなければならない. パスウェイの機能に関わりのあるこれらの要因を含めて変換し, 解析することは今後の重要な課題である.

参考文献

- [Schuster *et al.*, 2000] S. Schuster, D. A. Fell and T. Dandekar: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nature Biotechnology*, 18, pp.326–332, NPG, 2000.
- [Tamura *et al.*, 2009] T. Tamura, K. Takemoto and T. Akutsu: Measuring structural robustness of metabolic networks under a Boolean model using integer programming and feedback vertex sets, *IIBM2009*, pp.819–824, March, 2009.
- [Tiwari *et al.*, 2007] A. Tiwari, C. Talcott, M. Knapp, P. Lincoln and K. Laderoute: Analyzing pathways using SAT-based approaches, in: *2nd Intl. Conf. on Algebraic Biology*, LNCS, 4545, pp.155–169, 2007.
- [Ray *et al.*, 2009] O. Ray, K. Whelan and R. King: A non-monotonic logical approach for modelling and revising metabolic networks, *IIBM2009*, March, 2009.
- [Koshimura *et al.*, 2008] 越村三幸, 鍋島英知, 藤田博, 長谷川隆三: 極小モデル生成とジョブショップスケジューリング問題の解法, 日本ソフトウェア科学会 第 25 回大会, 東京, 2008.12.12.