モデル生成を用いた代謝ネットワークにおける極小活性パスウェイの列挙

Enumerating Minimal Active Metabolic Pathways by Model Generation

宋 剛秀
Takehide Soh

新領域融合研究センター
Transdisciplinary Research Integration Center
soh@nii.ac.jp, <http://research.nii.ac.jp/il/soh/>

井上 克巳
Katsumi Inoue

国立情報学研究所/総合研究大学院大学情報学専攻
National Institute of Informatics/Department of Informatics, The Graduate University for Advanced Studies
ki@nii.ac.jp, <http://research.nii.ac.jp/~ki/>

keywords: minimal model generation, metabolic pathway, propositional formula, encoding, Systems Biology.

Summary

In systems biology, identifying vital functions like glycolysis from a given metabolic pathway is important to understand living organisms. In this paper, we particularly focus on the problem of enumerating minimal active pathways producing target metabolites from source metabolites. We represent the problem in propositional formulas and solve it through minimal model generation. An advantage of our method is that each solution satisfies qualitative laws of biochemical reactions. Moreover, we can calculate such solutions for a cellular scale metabolic pathway within a few seconds. In experiments, we have applied our method to a whole *Escherichia coli* metabolic pathway. As a result, we found a minimal set of reactions corresponding to the conventional glycolysis pathway described in a biological database EcoCyc.

1. は じ め に

人間を含む生物は多くの化学反応の連鎖によりその生命を維持している。システム生物学において、そのような化学反応の相互作用は **パスウェイ** と呼ばれるネットワークによって表現される。近年パスウェイは活発に研究が行われており、様々な解析手法が提案されてきた [Jong 02, Terzer 09].

微視的な手法としては微分方程式を用いた手法がある [Tomlin 07]. これは詳細な解析が可能であり、例えばパスウェイに含まれるそれぞれの化合物濃度の時間変化をシミュレーションできる。一方、細胞全体程度の大きさのパスウェイに対する巨視的な手法も研究されている [Croes 06, Rahman 05, Planes 09]. これらの手法はパスウェイの構造に着目した離散的な解析であり、微分方程式系による手法よりもその入力が簡略化されている。それぞれの巨視的な手法は問題定式化とそれを解くための計算手法において異なっているが主目的は同じであり、与えられたパスウェイから生物学的に機能を持った部分を見つけることである。

Croes らの手法では、パスウェイを重み付きの有向 2 部グラフで表現し、二つの頂点間の重み付き経路を求める問題を解く [Croes 06]. 同じ問題を解く手法とし

ては他にも Planes と Beasley が制約プログラミングを用いた手法を提案している [Planes 09]. 彼らの手法の利点は重みによって出力された解の質の評価が行えること、既存の高速な経路探索手法を適用することで大きなパスウェイにも適用が可能であることである。しかしながら問題の出力は常に 2 頂点間の経路であり、複数の入出力を持つ化学反応の影響を考慮することができない。よって得られた解からは反応経路に必要な代謝物および化学反応などの重要な情報が欠落してしまう。

本研究では、与えられた初期代謝物の集合より目標代謝物の集合を生成するような化学反応からなるパスウェイを活性パスウェイと呼び、パスウェイの解析として、活性パスウェイを列挙する問題の定式化とモデル生成を用いた解法を提案する。特に我々は活性パスウェイの中のどの化学反応を削除しても、それが活性パスウェイにならないような活性パスウェイを列挙する **極小活性パスウェイ列挙問題** を提案する。この問題の解は目標代謝物を出力するために必要不可欠な化学反応のみを含むことになる。極小活性パスウェイ列挙問題の解を計算するために、まず命題論理式への符号化を行う。次に符号化された命題論理式の極小モデルを生成することで、最終的な解である極小活性パスウェイ

を得る．命題論理式の極小モデル生成器としては SAT ソルバーに変更を加えたものを利用した．近年の SAT 技術およびソルバーの研究は活発に行われており [井上 10, 鍋島 10]，これを極小モデル生成に利用することで大きなパスウェイに対しても高速な求解を期待できる．

提案手法を評価するために，大腸菌の代謝パスウェイに対して本手法を適用した．パスウェイのデータは論文 [Beasley 07] によるものと，大腸菌のデータベースとして良く知られる EcoCyc [Keseler 09] で公開されているものを用いている．近年，大腸菌は製薬の過程におけるアミノ酸合成に実際に利用されており，その解析の重要度が上がっている．本研究のように包括的に代謝パスウェイを解析することは，そのような化合物合成の効率改善などの実用上の価値が存在すると考えられる．

本論文の構成は以下になる．まず 2 章では命題論理式とその極小モデルについて簡単に説明を行う．3 章では極小活性パスウェイ列挙問題の定式化を行う．4 章では定式化した問題の命題論理式への符号化を説明する．5 章では実験結果の考察と関連研究についての議論を行い，6 章においてこの論文のまとめを行う．

2. 命題論理式と極小モデル

本章では必要な用語の定義を行う．リテラルとは命題変数またはその否定 ($\neg$) である．連言標準形 (conjunction normal form; CNF) の**命題論理式**はリテラルの選言である**節**の連言で与えられ CNF 式とも呼ばれる．ある命題論理式 Ψ への真偽値割当てとは，その命題論理式中の命題変数の集合 V から真偽値 $\{True, False\}$ への写像である．また Ψ を充足するような真偽値割当てを Ψ の**モデル**という．本論文ではモデルを，真 ($True$) に割り当てられた命題変数の集合としても表現することにする．命題論理式 Ψ と式中の命題変数の集合 V が与えられたとき，**極小モデル**は以下のように定義される [Niemelä 96]:

【定義 1】 I_1 と I_2 をそれぞれ命題論理式 Ψ のモデル， $V' \subseteq V$ を命題変数の集合とする．モデル I_1 が V' に関してモデル I_2 よりも小さいとは， $I_1 \cap V' \subset I_2 \cap V'$ が成り立つことをいう．ここで V' に関して I_1 より小さいモデルが存在しないときに， I_1 は Ψ の V' に関する**極小モデル**であるという．

明らかに極小モデルはモデルの要素数には依存しない．命題論理式の極小モデルを求める手法はいくつか提案されているが，本研究では SAT ソルバーを用いた極小モデル生成法を使用する [Koshimura 09]．

3. 極小活性パスウェイ列挙問題

本章ではまず問題を定式化するために必要な定義を行い，次に極小活性パスウェイ列挙問題の定義を行う．代謝物の集合を M ，化学反応の集合を R とする．ここで $M \cap R = \emptyset$ とする．以降，化学反応のことを単に反応と記述する．集合 M, R の要素間の有向辺の集合を $A \subseteq (R \times M) \cup (M \times R)$ とする．**パスウェイ**は代謝物と反応をノード集合とする有向 2 部グラフ $G = (M, R, A)$ で与えられる．パスウェイ $G = (M, R, A)$ に対し， $M_S \subset M$ を初期代謝物の集合， $M_T \subset M$ を目標代謝物の集合とする．ただし $M_S \cap M_T = \emptyset$ である．ある代謝物 $m \in M$ は有向辺 $(m, r) \in A$ が存在するときに，反応 r の**反応物**という．一方，代謝物 $m \in M$ は有向辺 $(r, m) \in A$ が存在するときに，反応 r の**生成物**という．パスウェイ上において有向辺は反応における反応物と生成物の関係を表している．例えばあるパスウェイ上において，二つの代謝物 m_1, m_2 を反応物とし，二つの代謝物 m_3, m_4 を生成物とするような反応 r_1 がある場合，有向辺の集合は $\{(m_1, r_1), (m_2, r_1), (r_1, m_3), (r_1, m_4)\} \subseteq A$ を満たす．このような反応物と生成物および反応の関係を表す写像を次に定義する．写像 $s: R \rightarrow 2^M$ ， $s(r) = \{m \in M \mid (m, r) \in A\}$ はある反応が与えられたときにその反応物の集合を返す．また写像 $p: R \rightarrow 2^M$ ， $p(r) = \{m \in M \mid (r, m) \in A\}$ はある反応が与えられたときにその生成物の集合を返す．これら二つの写像に対して写像 s' と p' を定義する．写像 $s': M \rightarrow 2^R$ はある代謝物が与えられたときにそれを反応物として使用する反応の集合を返す．同様に写像 $p': M \rightarrow 2^R$ はある代謝物が与えられたときにそれを生成物として生成する反応の集合を返す．先ほどの反応の例を用いると， $s(r_1) = \{m_1, m_2\}$ であり，またパスウェイ中において m_1 を反応物として使用する反応が r_1 のみであるときは $s'(m_1) = \{r_1\}$ となる．

ここで時刻を表す整数変数 t を導入する．1 単位時間は任意の一つの反応の反応時間である．本定式化では，時刻によって各反応と代謝物間の活性および供給の順序を表す．いまパスウェイ $G = (M, R, A)$ 上における代謝物の部分集合 $M' \subset M$ からの供給可能性および活性可能性を以下に定義する．パスウェイ G 上で代謝物 $m \in M$ が時刻 $t = 0$ において M' より**明らかに供給可能**であるとは $m \in M'$ となるとき，またそのときに限る．パスウェイ G 上で反応 $r \in R$ が時刻 $t > 0$ において M' より**活性可能**であるとは，時刻 $t - 1$ において任意の代謝物 $m \in s(r)$ が供給可能であるとき，またそのときに限る．パスウェイ G 上で代謝物 $m \in M$ が時刻 $t > 0$ において M' より**供給可能**であるとは，少なくとも一つの反応 $r \in p'(m)$ が時刻 t において M' より活性可能であるとき，またそのときに限る．

本研究では，パスウェイにおいて初期代謝物は目標

代謝物を生成するまでの間供給が継続するものとする。これは水分子など細胞を構成する物質は多くの量が存在すると考えられ、またそれらの物質の他に初期代謝物に指定される物質についても十分な供給を考えるためである。これより供給が継続する代謝物を使用する反応も一度活性可能になればその状態を保持し、その生成物についても供給が継続され別の反応で使用されても無くなることはない。これより代謝物と反応に関して次の二つの性質が成り立つ：ある反応 $r \in R$ が時刻 t において活性可能ならば時刻 $t+1$ においても活性可能であり、ある代謝物 $m \in M$ が時刻 t において供給可能ならば時刻 $t+1$ においても供給可能である。

ここで**パスウェイ問題**を5項組 $\pi = (M, R, A, M_S, M_T)$ で定義する。このときにパスウェイ問題 π の活性パスウェイは以下のように定義される。

【定義2】活性パスウェイ

パスウェイ $G = (M, R, A)$ の部分グラフ $G' = (M', R', A')$ が以下の条件を満たすとき、 G' を π の**活性パスウェイ**とよぶ。

- (A1) $M_T \subset M'$.
- (A2) $M' = M_S \cup \{m \in M \mid (m, r) \subseteq A, r \in R'\}$
 $\cup \{m \in M \mid (r, m) \subseteq A, r \in R'\}.$
- (A3) $A' = \{(m, r) \in A \mid r \in R'\}$
 $\cup \{(r, m) \in A \mid r \in R'\}.$
- (A4) $G' = (M', R', A')$ 上で、すべての $m \in M'$ は、 M_S より供給可能。

定義から分かるように活性パスウェイとは、初期代謝物 M_S から供給可能な代謝物集合および活性可能な反応集合のみから構成されており、かつ、すべての目標代謝物 M_T を供給可能にするものである。 M_S と M_T の組合せにも依存するが、一般的に代謝パスウェイにおいて、このような活性パスウェイは非常に数多く存在する。本研究ではその中でも重要なものとして極小活性パスウェイを提案する。定義は以下になる。

【定義3】極小活性パスウェイ

パスウェイ問題 π に対する二つの活性パスウェイ $G_1 = (M_1, R_1, A_1), G_2 = (M_2, R_2, A_2)$ が与えられたとき、 $R_1 \subset R_2$ であるとき、またそのときに限り活性パスウェイ G_1 は G_2 より**小さい**と呼び、 $G_1 \subset G_2$ と表される。さらにパスウェイ問題 π に対する活性パスウェイで G_1 より小さいものが存在しないときに、 G_1 を π の**極小活性パスウェイ**とよぶ。

定義より極小活性パスウェイ中のどの反応を削除しても、活性パスウェイにはならない。これは反応それぞれが必要不可欠であることを表している。極小活性パスウェイ列挙問題は次のように定義される。

【定義4】極小活性パスウェイ列挙問題

入力 パスウェイ問題 $\pi = (M, R, A, M_S, M_T)$.

出力 π に対する活性パスウェイが存在すればその極小活性パスウェイを全て列挙し、存在しなければ

“なし”を返す。

ここである終了時刻 z および極小活性パスウェイを全て列挙するために時間 $[0, z]$ を考える。この時間内にすべての目標代謝物を生成できるような極小活性パスウェイを列挙する問題を**終了時刻 z に関する極小活性パスウェイ列挙問題**と呼ぶことにする。終了時刻を $z = 1$ から上限 $z = |R|$ まで一つずつ増加させ逐次的に問題を解くことで元の極小活性パスウェイ列挙問題の解を全て計算できる。実際に未知のパスウェイを求める際には、 $z = 1$ から開始し、ある z で解集合が得られればそれらの解析を行い、必要であれば z を増やして解列挙を行うことを繰り返す。なおここで上記と同様に活性パスウェイを列挙する問題を**活性パスウェイ列挙問題**および**終了時刻 z に関する活性パスウェイ列挙問題**と呼ぶことにする。

4. 命題論理式への符号化

本章では前章で定義した問題を命題論理式へ符号化し、その極小モデルを生成することによる求解手法の説明を行う。後の章にも記載するが、このように命題論理式へ符号化することの利点は生物学的な要因によって発生する制約の追加およびそれを考慮した求解を容易に行うことができることにある。

4.1 時刻に基づいた符号化

符号化にあたり整数 i, j をインデックスとし、代謝物集合の要素を $m_i \in M$ 、反応集合の要素を $r_j \in R$ のように表す。また i, j は代謝物、反応のどちらにも使用するものとする。整数定数 z を与えられた終了時刻とし、整数変数 $t (0 \leq t \leq z)$ を時刻を表す変数とする。符号化を行うにあたり2種類の命題変数を定義する。命題変数 $m_{i,t}^*$ は代謝物 $m_i \in M$ が時刻 t において生成されるときかつそのときに限り真である。命題変数 $r_{j,t}^*$ は反応 $r_j \in R$ が時刻 t において活性されるときかつそのときに限り真である。ここから実際の符号化を説明する。まず代謝物は一度供給可能になればそれ以降も供給可能であるという性質は ψ_1 で与えられる。また反応についての性質も同様に ψ_2 で与えられる。

$$\psi_1 = \bigwedge_{m_i \in M} \bigwedge_{0 \leq t < z} (m_{i,t}^* \rightarrow m_{i,t+1}^*)$$

$$\psi_2 = \bigwedge_{r_j \in R} \bigwedge_{1 \leq t < z} (r_{j,t}^* \rightarrow r_{j,t+1}^*)$$

パスウェイ中の任意の反応 $r_j \in R$ に対して、以下の論理式が与えられる。これは反応 r_j が時刻 t で活性されるのであれば、その反応物は $t-1$ においてすべて供

給可能でなければならないことを表している.

$$\psi_3 = \bigwedge_{r_j \in R} \bigwedge_{1 \leq t \leq z} \left(r_{j,t}^* \rightarrow \bigwedge_{m_i \in s(r_j)} m_{i,t-1}^* \right)$$

パスウェイ中の任意の反応 $r_j \in R$ に対して, 以下の論理式が与えられる. これは反応 r_j が時刻 t で活性されるのであれば, その生成物は時刻 t において供給可能になることを表している.

$$\psi_4 = \bigwedge_{r_j \in R} \bigwedge_{1 \leq t \leq z} \left(r_{j,t}^* \rightarrow \bigwedge_{m_i \in p(r_j)} m_{i,t}^* \right)$$

初期代謝物を除くパスウェイ中の任意の代謝物 $m_i \in (M \setminus M_S)$ に対して, 以下の論理式が与えられる. これはある代謝物 m_i が時刻 t で供給可能になっているのであれば, それは m_i が $t-1$ で供給可能になっている, もしくは少なくとも一つの $r_j \in p'(m_i)$ が t で活性可能になっていることを表している.

$$\psi_5 = \bigwedge_{m_i \in (M \setminus M_S)} \bigwedge_{1 \leq t \leq z} \left(m_{i,t}^* \rightarrow m_{i,t-1}^* \vee \bigvee_{r_j \in p'(m_i)} r_{j,t}^* \right)$$

問題の初期条件と目標条件は以下のように与えられる:

$$\begin{aligned} \psi_6 &= \bigwedge_{m_i \in M_S} m_{i,0}^* \wedge \bigwedge_{m_j \in M \setminus M_S} \neg m_{j,0}^* \\ \psi_7 &= \bigwedge_{m_i \in M_T} m_{i,z}^* \end{aligned}$$

ここで上記の式 $\psi_1, \psi_2, \dots, \psi_7$ の連言を Ψ_z とする. 次の命題は終了時刻 z に関する活性パスウェイ列挙問題に対する健全性と完全性を与える.

命題 1 $\pi = (M, R, A, M_S, M_T)$ をパスウェイ問題, 時間 $0 \leq t \leq z$ に対する符号化された論理式を Ψ_z とする. I が Ψ_z のモデルであるとき, またそのときに限り $G^{sol} = (M^{sol}, R^{sol}, A^{sol})$ は π の活性パスウェイである. ただし

$$\begin{aligned} M^{sol} &= \{m_i \mid m_{i,z}^* \in I\}, R^{sol} = \{r_j \mid r_{j,z}^* \in I\}, \\ A^{sol} &= \{(m_i, r_j) \in A \mid r_j \in R^{sol}\} \\ &\quad \cup \{(r_j, m_i) \in A \mid r_j \in R^{sol}\} \end{aligned}$$

である.

(証明)

($\Rightarrow$) G^{sol} が定義 2 の各条件を満たすことを以下で証明する. まず G^{sol} が $G = (M, R, A)$ の部分グラフであるのは明らかである. 次に $\Psi_z \models \psi_7$ より G^{sol} は条件 A1 を満たす. また $\Psi_z \models \psi_1 \wedge \psi_3 \wedge \psi_4$ より, $r_{j,z}^* \in I$ ならば, 任意の $m_i \in s(r_j) \cup p(r_j)$ に対して $m_{i,z}^* \in I$ であるので G^{sol} は条件 A2 を満たす. 条件 A3 を満たすことは G^{sol} の定義より明らか. 条件 A4 を時刻を表す整数変数 k に関する数学的帰納法によって証明する. た

だし $0 \leq k \leq z$ とする.

帰納法の基礎段階 $k=0$ の場合, $\Psi_z \models \psi_1 \wedge \psi_6$ より, $m_{i,k=0}^* \in I$ であるならば, それに対する $m_i \in M^{sol}$ は G^{sol} 上で M_S より明らかに供給可能である.

帰納段階 いま $m_{i,k}^* \in I$ ならば $m_i \in M^{sol}$ は G^{sol} 上で M_S より供給可能であると仮定する. このときに $m_{i,k+1}^* \in I$ ならば $m_i \in M^{sol}$ が G^{sol} 上で M_S より供給可能であることを証明する. $\Psi_z \models \psi_5$ より, $m_{i,k+1}^* \in I$ ならば次の二つの場合のいずれかが必ず成り立つ: (a) $m_{i,k}^* \in I$, (b) 少なくとも一つの反応 $r_j \in p'(m_i)$ に対して, $r_{j,k+1}^* \in I$. (a) の場合, 仮定より m_i は時刻 k において供給可能であり, 定義より m_i は時刻 $k+1$ においても供給可能である. (b) の場合, $\Psi_z \models \psi_3$ より, 全ての $m_a \in s(r_j)$ に対して, $m_{a,k}^* \in I$ である. 仮定より代謝物 m_a は時刻 k において供給可能であるので定義より r_j は時刻 $k+1$ において活性可能である. よって r_j の生成物である m_i は時刻 $k+1$ において供給可能. 両方の場合において, $m_i \in M^{sol}$ は時刻 $k+1$ において供給可能である. それ故, 数学的帰納法より任意の k において, m_i は G^{sol} 上で M_S より供給可能であり, よって G^{sol} は条件 A4 を満たす. 以上より, G^{sol} は π の活性パスウェイである.

($\Leftarrow$) いま π が与えられたときに, 上限の時刻 z までの活性パスウェイが存在するならば, それに対応する Ψ_z のモデルが存在することを証明する.

π の時刻 z までの任意の活性パスウェイを $G^a = (M^a, R^a, A^a)$ とし, それを構成するような命題変数集合 I_a とする. G^a は活性パスウェイであることから全ての代謝物 $m_i \in M^a$ は M_S より供給可能であり, また $M_T \subset M^a$ より $I^a \models \psi_5 \wedge \psi_6 \wedge \psi_7$. また定義 2 の条件 A2 と A3 より $I^a \models \psi_3 \wedge \psi_4$ である. 活性パスウェイの代謝物と反応は時刻 t において供給可能かつ活性可能であるのであれば $t+1$ においても同様なので $I^a \models \psi_1 \wedge \psi_2$. 以上より $I^a \models \Psi_z$ であり, 活性パスウェイが存在するならばそれを構成するモデルが存在する. $\square$

ここで $V^z = \{r_{i,z}^* \mid r_i \in R\}$ を上限の時刻における反応変数の集合とする. このとき次の命題は終了時刻 z に関する極小活性パスウェイ列挙問題に対する健全性と完全性を与える.

命題 2 $\pi = (M, R, A, M_S, M_T)$ をパスウェイ問題, 時間 $0 \leq t \leq z$ に対する符号化された論理式を Ψ_z とする. もし I_{min} が Ψ_z の V^z に関する極小モデルであるとき, またそのときに限り $G^{sol} = (M^{sol}, R^{sol}, A^{sol})$ は π の極小活性パスウェイである. ただし

$$\begin{aligned} M^{sol} &= \{m_i \mid m_{i,z}^* \in I_{min}\}, R^{sol} = \{r_j \mid r_{j,z}^* \in I_{min}\}, \\ A^{sol} &= \{(m_i, r_j) \in A \mid r_j \in R^{sol}\} \\ &\quad \cup \{(r_j, m_i) \in A \mid r_j \in R^{sol}\} \end{aligned}$$

である.

(証明)

($\Rightarrow$) 命題 1 より, G^{sol} は π の活性パスウェイである.

いま $G' \subset G^{sol}$ となるような、活性パスウェイ $G' = (M', R', A')$ が存在すると仮定する. このときに定義より $R' \subset R^{sol}$ である. 命題 1 より, そのような活性パスウェイ G' を構成するような Ψ_z のモデル I' が存在し $I' \cap V^z \subset I_{min} \cap V^z$ を満たす. しかしこれは極小モデルの定義と矛盾するので, このような活性パスウェイは存在しない. よって G^{sol} は極小活性パスウェイである.

($\Leftarrow$) 命題 1 より, 任意の極小活性パスウェイはそれに対応する Ψ_z のモデルを持つ. いま与えられた極小活性パスウェイ G^{sol} を構築するモデル I_{min} に対して $I' \cap V' \subset I_{min} \cap V'$ となるモデル I' が存在したとする. しかしモデル I' から構成される活性パスウェイ G' は $G' \subset G^{sol}$ となり矛盾する. よって極小活性パスウェイであれば対応する Ψ_z の V' に関する極小モデルが存在する. $\square$

ここまで終了時刻 z に関する極小活性問題の命題符号化方法とその正当性を示した. 3 章で記述したように実際の解析では終了時刻 z を一つずつ増加させながら逐次的に計算を行うが, 時間 $0 \leq t \leq z$ に対する計算を行う際には時間 $0 \leq t \leq z-1$ で得られた結果から阻止節 [長谷川 10] を生成して利用することで冗長な計算を回避している.

4.2 制約の追加と生物学における応用

ここまで述べた極小活性パスウェイ列挙問題は初期代謝物からの反応の活性可能性のみを考慮しているが, 対象となるパスウェイの生物学的制約が事前に分かっている場合は, それらを節で表現し追加することができる. 以下には代表的な制約を挙げる.

i. 代謝物の生成・増加の抑制

ある代謝物を生成する際に, 副産物として生成したくない代謝物が存在する場合がある. そのような代謝物の集合 M' を一切生成しないような極小活性パスウェイを計算するにはそれぞれの時刻 t ($0 \leq t \leq z$) に対して以下の論理式を加えることで実現できる:

$$\bigwedge_{m_i \in M'} \neg m_{i,t}^*$$

また実際の生物の代謝パスウェイを考慮する際には, 生成は許されるがそれが増加・蓄積していくことが許されないような場合が存在する. すなわちそれらの代謝物集合 M' は生成された後に消費されることが要求される. この制約はそれぞれの時刻 t ($0 \leq t < z$) に対して以下の論理式を記述することで表現できる:

$$\bigwedge_{m_i \in M'} m_{i,t}^* \rightarrow \bigvee_{r_j \in s'(m_i)} r_{j,t+1}^*$$

ii. 遺伝子に関わる制約

上記の代謝物の生成に関する制約の他に, 代謝パスウェイにおける反応に重要な役割を果たす酵素に関する

制約がある. 酵素は DNA から該当する部分が転写・翻訳という過程を経て生成される (遺伝子発現). よって遺伝子制御ネットワークにおいて排他関係にある遺伝子に対応する酵素に触媒される反応は同時に活性されることがない. この制約はそれぞれの時刻 t ($1 \leq t \leq z$) に対して以下の論理式を記述することで表現できる:

$$\neg r_{i,t}^* \vee \neg r_{j,t}^*$$

ここで反応 r_i と r_j は抑制関係にある酵素を介して活性するものとする. このような抑制関係の知識を加えることによって, 生物学的には生成されないような活性パスウェイの出力を避けることができる. この他に遺伝子と酵素の対応が事前に知られている場合は, 生物における野生型とその遺伝子の一部分が変異したような変異型の代謝パスウェイの違いを解析することも可能である. 例えばある遺伝子が削除された変異型における代謝パスウェイを計算するには削除された遺伝子に対応する酵素反応が活性できないようにそれぞれの時刻 t ($1 \leq t \leq z$) および対象の反応に対して単位節 $\neg r_{i,t}^*$ を追加すればよい.

5. 実験結果と考察

5.1 計算環境

評価を行うために大腸菌の代謝パスウェイに対して提案手法を適用した. 計算機実験にあたっては, 二つの反応データベースを用いた. 一つは [Beasley 07] で用いられているものであり, もう一つは生物学分野においてよく知られた大腸菌のデータベース EcoCyc [Keseler 09] である. 後者は過去の生物学における実験データや論文の情報を集積したものであり, 今回は論文執筆時点の最新版である 13.6 を使用した. EcoCyc データベース中の可逆反応は方向の異なる二つの反応として反応名の後ろに $_a, _b$ を付加して区別する. また以下の実験において, 生物学分野における実験によって検証され一般的な知識として定着している代謝パスウェイを参照パスウェイと呼ぶことにする. パスウェイ問題から命題論理式への符号化は Perl で実装したプログラムを使用し, 極小モデル生成器は C++ で書かれた SAT ソルバー Minisat 2.0 [MiniSat] のコードを変更したものを利用した. OS は Ubuntu Linux を使用し, CPU が 1.84GHz, メモリは 1GB の計算機を用いて計算を行った.

5.2 論文 [Beasley 07] のデータを用いた評価

まず論文 [Beasley 07] の反応データベースを用いて提案手法を二つの従来手法と比較した. 一つはパスウェイ解析を最適化問題にモデリングしている [Beasley 07] であり, この手法の入力は化学量の情報を含んだ反応データベースである. もう一つの手法は制約プログラミングを用いて重み付き経路探索を行う手法 [Planes

09] であり、この手法の入力は提案手法と同じく化学量の情報を除いた反応データベースである。この二つの手法は論文 [Planes 09] においても比較が報告されている。今回の実験では、論文 [Beasley 07, Planes 09] と同じ反応データベースを入力のパスウェイとし、初期代謝物、目標代謝物等の条件設定は全てこの論文と同一である。参照パスウェイとの一致については対象二つの手法が比較された文献 [Planes 09] を踏襲し、計算された解中に含まれる反応が全て参照パスウェイ上に存在するか否かを確認している。

結果を表 1 に示す。1 列目は実験を行った 10 のパスウェイ問題を表しており、2 列目は解を求めた終了時刻 z の値、3 列目は解の数、4 列目は $z = 1$ から 2 列目に記載された z の値までの合計計算時間を表している。また 1 列目の番号は次のパスウェイに対応している: #1: Gluconeogenesis, #2: Glycogen, #3: Glycolysis, #4: Proline Bio-synthesis, #5: Ketogluconate Metabolism, #6: Pentose Phosphate, #7: Deoxythymidine Phosphate, #8: Krebs's Cycle, #9: NAD Biosynthesis, #10: Arginine Biosynthesis. 今回実験は一致する解が見つかるまで終了時刻 z を一つずつ増加させた。5 列目から 8 列目はそれぞれの手法が参照パスウェイと一致するパスウェイを見つけることができたかどうかを示している。比較 (1) と表記した 6 列目と 7 列目はそれぞれ論文 [Beasley 07] における 2 種類の最適化関数を用いた結果を表しており (a) は反応数の最小化 (b) は ATP 生産の最大化の結果である。また比較 (2) と記載した 8 列目は [Planes 09] の結果を表している。

結果として提案手法は参照パスウェイに一致する全てのパスウェイを $z = 9$ までに全て発見することができた。解の数はパスウェイ #3, #8, #9 を除いて 10 以下である。二つの従来手法と提案手法は入力情報、問題定式化、また求める解の数などに違いがあるため、三つの手法を直接に比較することは困難であるが、化学量などの情報を使わずに全ての参照パスウェイを計算によって求解可能であることは我々の手法の利点であるといえる。次に計算時間の比較を行う。論文 [Beasley 07] では 3GHz の CPU を用いて行った 2 種類の最適化関数に対する 10 のパスウェイの計算機実験、つまり合計 20 の各実験において、平均の計算時間が 11 秒であり最長は 85 秒であることが報告されている。提案手法は 1.84GHz の CPU を用いて実験したうえで、平均の計算時間は 1 秒、最長で 2.7 秒であり提案手法はより高速であるといえる。なお文献 [Planes 09] には計算時間については記載がないため比較を行っていない。

5.3 EcoCyc のデータを用いた評価

大腸菌における解糖系のパスウェイを列挙するためにグルコースー 6ーリン酸 (GLC-6-P) を初期代謝物とし、ピルビン酸 (PYRUVATE) を目標代謝物とした。括

表 1 論文 [Beasley 07] のパスウェイに対する解析結果

番号	提案手法				比較 (1)		比較 (2)
	z	#S	秒	一致	一致 (a)	一致 (b)	一致
#1	9	6	1.53	✓	✓		
#2	3	1	0.36	✓	✓		✓
#3	9	35	2.57	✓	✓	✓	
#4	4	1	0.49	✓	✓		
#5	5	6	0.76	✓			✓
#6	6	7	0.95	✓	✓		✓
#7	4	1	0.53	✓	✓		✓
#8	8	35	2.71	✓		✓	
#9	5	16	0.69	✓	✓		✓
#10	8	1	0.98	✓	✓		✓
一致数の合計				10	8	2	6

表 2 極小活性パスウェイと構成される反応

反応名 (by [Keseler 09])	計算されたパスウェイ									EC
	a	b	c	d	e	f	g	h	i	
2.7.1.121-RXN				x						
2PGADEHYDRAT-RXN				x	x	x	x	x	x	x
3PGAREARR-RXN				x	x	x	x	x	x	x
6PFRUCTPHOS-RXN		x						x		x
6PGLUCONOLACT-RXN	x									
ADENYL-KIN-RXN							x			
DLACTDEHYDROGNAD-RXN		x	x							
F16ALDOLASE-RXN		x						x		x
F16BDEPHOS-RXN										x
GAPOXNPSPHOSPHN-RXN				x	x	x	x	x	x	x
GLU6PDEHYDROG-RXN										
GLYOXIII-RXN		x	x							
KDPGALDOL-RXN	x									
METHGLYSYN-RXN		x	x							
NAD-KIN-RXN	x									
NADPYROPHOSPHAT-RXN					x					
PEPDEPHOS-RXN					x			x		x
PEPSYNTH-RXN						x	x		x	x
PGLUCISOM-RXN		x	x	x	x	x	x	x	x	x
PGLUCONDEHYDRAT-RXN	x									
PHOSGLYPHOS-RXN				x	x	x	x	x	x	x
RXN0-313			x	x	x	x	x		x	
RXN0-4401									x	
TRIOSEPISEMERIZATION-RXN				x						x
反応の数	5	6	6	7	7	8	8	8	8	11

弧内は EcoCyc 内での表記である。以下、代謝物および反応名の表記は EcoCyc から引用したものをを用いる。また初期代謝物として GLC-6-P の他に反応データベースにおいて最も出現頻度の高いものから 6 つを選択している。なおこの出現頻度の計算は [Beasley 07] を基に行っている。具体的には 1073 ある代謝物の中から出現頻度に従って PROTON, WATER, ATP, [Pi], ADP, NAD を初期代謝物とした。なお本実験においては 3 章に記載した指針に従い $z = 1$ から $z = 8$ まで一つずつ増加させ、また基数制約 [Sinz 05] を用いて活性できる反応数を 8 以下に制約している。

以上の設定を用いて、大腸菌の代謝パスウェイ全体に対して極小活性パスウェイを計算し、最終的に結果として 9 つの極小活性パスウェイを列挙した (表 2 参照)。なお計算時間は $z = 1$ から $z = 8$ までを合計して

図1 EcoCyc のパスウェイに対する解析結果

図2 計算された解糖系パスウェイ h (拡大図)

3.6 秒である。表の 1 列目は EcoCyc 表記による反応名を示している。表の 2 列目から 10 列目は計算された 9 つのパスウェイをアルファベットで示しており、“x”印はパスウェイを構成している反応を表している。例えば計算されたパスウェイ a は 5 つの反応で構成されている。11 列目にはパスウェイ EC が記載されているがこれは EcoCyc に記載されている解糖系の参照パスウェイである。表から分かるように計算によって求めた極小活性パスウェイ h の反応は全て参照パスウェイの反応と一致していることが分かる。大腸菌の代謝パスウェイ全体におけるパスウェイ h を図 1 に、またその拡大図を図 2 に示す。これらの図は Java 言語のライブラリである Jgraph を用いて出力した。図 2 において楕円形の頂点は代謝物、長方形の頂点は反応を表す。また代謝物のうち灰色と黄色で表されたものは初期代謝物、赤色のものは目標代謝物を表す。図から初期代謝物より目標代謝物が生成される過程を読み取ることができる。今回の実験では参照パスウェイに一致するものが見つかるまで z を一つずつ増加させたが、論文や過去の実験の結果により推測可能であれば z の値を直接指定しても良い。例えば今回実験を行った解糖系では論文 [Freedberg 71, Ferguson 98] などにも記載されているように一般的に、GLC-6-P から FRUCTOSE-6P, FRUCTOSE-16-DIPHOSPHATE, GAP, DPG, G3P, 2-

PG, PHOSPHO-ENOL-PYRUVATE, PYRUVATE を経た 8 段階で記述されることから z の値を推測することができる。

表 2 では、参照パスウェイにおけるいくつかの反応が計算されたパスウェイ h に含まれていなかった。これは参照パスウェイがしばしばある代謝物からある代謝物へのバイパス経路を含んでいることが理由である。例えば今回の実験では計算された極小部分パスウェイ h は参照パスウェイと比較して、反応 TRIOSEPISEMERIZATION-RXN が含まれていない。この反応は代謝物 DIHYDROXY-ACETONE-PHOSPHATE から代謝物 GAP を生成するものであるが、解糖系パスウェイでは既に反応 F16ALDOLASE-RXN a が代謝物 GAP を生成しているため、極小活性パスウェイの定義よりこのような反応は計算された解の中には含まれない。しかしこのような反応は一見冗長であり、効率的ではない。大腸菌の解糖系パスウェイにこの反応が含まれる理由として、代謝物 DIHYDROXY-ACETONE-PHOSPHATE が生成されたまま蓄積すると生育障害が発生すること [Freedberg 71], そしてそれを避けるために生成されると同時に消費する反応経路が必要であることが報告されている [Ferguson 98]。このような知識は、パスウェイの構造からだけでは判断ができないが、事前に知ることができれば 4.2 節に記載したように元の問題に制約として追加することができる。実際に代謝物 DIHYDROXY-ACETONE-PHOSPHATE が供給可能であるならばそれは少なくとも一つの反応によって消費されなければならないという制約を追加したところ、参照パスウェイと同様に反応 TRIOSEPISEMERIZATION-RXN が追加されたパスウェイ h が解として計算された。他に二つの反応が含まれていないが、これらの反応も PYRUVATE 生成には必須でないように見える。目標代謝物である PYRUVATE 生成後の処理に関するものである可能性もあり、これらを考察することは今後の課題の一つである。

5.4 関連研究

Schuster らは反応流束に着目した解析手法として Elementary flux mode (EFM) 解析を提案している [Schuster 00]。初期代謝物が与えられたときに EFM 解析によっても目標代謝物を生成するパスウェイが求められるが、それらのパスウェイ内の代謝物は供給されると必ず消費される必要がある点で、極小活性パスウェイとは異なる。また EFM 解析では反応流速を整数値で与えるため解の数が多くなる傾向にある。例えば 100 の反応を含むようなパスウェイに対しても 2 万以上の解を出力することが報告されており、その計算時間については最長で 1027 秒と報告されている [Klamt 02]。本研究では最大 1777 の反応を含むパスウェイに対して実験を

行った結果、平均の計算時間は1.3秒であり最長は3.6秒である。計算機環境や定式化等の相違により直接的な比較は困難であるが、本手法は十分に高速であることが分かる。またEFM解析と比較して4.2節で記載したように制約を柔軟に追加できることも本手法の利点である。

Handorfらは、Inverse Scope問題を提案している[Handorf 08]。この問題は目標代謝物と初期代謝物が与えられたときに、目標代謝物を生成するような初期代謝物の極小の部分集合を求める問題である。また極小集合ではなく要素数が最小の集合を求める最適化問題がAnswer Set Programming (ASP)を用いた手法によってSchaubとThieleにより解かれている[Schaub 09]。これらの問題と極小活性パスウェイ列挙問題の違いはこれらの問題の解はパスウェイではなく代謝物の集合であるということである。与えられた初期代謝物の集合から目標代謝物へ至る極小の活性パスウェイを列挙することは重要な経路を選択し判断するために重要であり、本研究の特徴の1つであるといえる。ASPを用いた他の解析手法としてRayらは用いた反応の抑制関係を考慮したパスウェイの解析手法を提案している[Ray 09]。しかし実際のパスウェイに対する評価はされておらず、どのような解が求められるかについては述べられていない。

重み付き経路発見を用いた手法[Rahman 05, Croes 06, Planes 09]のメリットは、既存のグラフ探索アルゴリズムを活用でき高速に解を求めることである。これらのアルゴリズムはスケーラビリティの面で代謝パスウェイ全体を解析するのに十分であるが、複数入出力の反応の制約は考慮されていない。Figueiredoらは経路発見手法[Croes 06, Rahman 05]に関する問題点を具体的な例を用いて報告している[Figueiredo 08]。Küffnerらは複数の基質と生成物の反応制約を考慮したペトリネットを用いた手法[Küffner 99]を報告しているが、解の極小性については述べられていない。

Tiwariらは重み付きMax-SATソルバを用いた手法を提案している[Tiwari 07]。彼らはソフト制約を用いてパスウェイ上の制約を重み付きMax-SAT問題で表現している。この手法は解の重みによって順位付けを行うことが可能であるが、その順序はしばしば生物学的に受け入れることができない。その理由の一つは、それが実際には頻繁に起こり得るにも関わらず一つの化合物を二つの反応が同時に使用することを禁止していることである。

パスウェイ中に含まれる閉路は、しばしばそれを除去するための前処理や後処理を必要とする[Beasley 07, Tamura 09]。それに対して目標代謝物を生成するのに寄与しない閉路を含む活性パスウェイは、それを除去しても活性パスウェイであるため定義より極小活性パスウェイになり得ない。これより提案手法ではそれら

の前後処理なしに必要な解を求めることができる。

KautzとSelmanはSTRIPS形式の古典的プランニング問題をあるプラン長におけるプランの存在を判定する問題としてSAT問題へ符号化し、短いプラン長の問題からプラン長を一つずつ増加させながら充足可能性を判定していく解法を提案している[Kautz 92]。この解法では初めて充足可能になった問題のプラン長で実行可能なプランが解になる。それに対して提案手法は終了時刻 z に関する極小活性パスウェイ列挙問題を命題論理式へ符号化すること、また極小モデルの列挙を行う点で彼らの手法と異なる。極小モデルを求めることによって冗長な閉路を含まない解を列挙できることは本手法の利点の一つであるが、このような制約をSTRIPS形式で表現できるかどうかを考察することは今後の課題である。

6. おわりに

本研究では初期代謝物集合から目標代謝物集合を生成するために必要な反応集合を列挙する極小活性パスウェイ列挙問題を定式化し、命題論理式への符号化を用いた求解方法を提案した。評価のために大腸菌の代謝パスウェイに対して本手法を適用し、結果として実験を行った11のパスウェイ全てにおいて参照パスウェイと一致する解を計算によって求めることに成功した。本手法の利点は細胞程度の大きさのパスウェイに対しても従来手法と比較して高速に計算を実行できる点、そして柔軟に制約を追加できる点である。これは今後より大きな代謝パスウェイを解析する際、またシグナル伝達パスウェイや遺伝調節パスウェイのような異なったパスウェイを横断した解析を行う際に重要な特徴である。今後の課題として、出力された解を生物学的視点から解析し、専門家の意見を取り入れながらより精度の高いパスウェイを出力できるように手法を改良することが挙げられる。

謝 辞

本研究の一部は、平成20～23年度科学研究費補助金基盤研究(A)「推論による仮説発見とシステム生物学への応用」(No. 20240016)および平成22～23年度科学研究費補助金特別研究員奨励金「SAT変換を用いた制約充足問題の解法とシステム生物学への応用」による。本論文における実験結果については、新領域融合研究センター・馬場知哉特任准教授に貴重なコメントを頂いた。

◇ 参 考 文 献 ◇

- [Beasley 07] Beasley, J. E. and Planes, F. J.: Recovering metabolic pathways via optimization, *Bioinformatics*, Vol. 23, No. 1, pp. 92–98 (2007)

- [Croes 06] Croes, D., Couche, F., Wodak, S. J., and Helden, van J.: Inferring meaningful pathways in weighted metabolic networks, *Journal of Molecular Biology*, Vol. 356, No. 1, pp. 222–236 (2006)
- [Ferguson 98] Ferguson, G. P., Totemeyer, S., MacLean, M. J., and Booth, I. R.: Methylglyoxal production in bacteria: suicide or survival?, *Archives of Microbiology*, Vol. 170, No. 4, pp. 209–218 (1998)
- [Figueiredo 08] Figueiredo, de L. F., Schuster, S., Kaleta, C., and Fell, D. A.: Can sugars be produced from fatty acids?: A test case for pathway analysis tools, *Bioinformatics*, Vol. 24, No. 22, pp. 2615–2621 (2008)
- [Freedberg 71] Freedberg, W. B., Kistler, W. S., and Lin, E. C. C.: Lethal synthesis of methylglyoxal by *Escherichia coli* during unregulated glycerol metabolism, *Journal of Bacteriology*, Vol. 108, No. 1, pp. 137–144 (1971)
- [Handorf 08] Handorf, T., Christian, N., Ebenhö, O., and Kahn, D.: An environmental perspective on metabolism, *Journal of Theoretical Biology*, Vol. 252, No. 3, pp. 530–537 (2008)
- [長谷川 10] 長谷川 隆三, 藤田 博, 越村 三幸: モデル列挙とモデル計数, 人工知能学会誌, Vol. 25, No. 1, pp. 96–104 (2010)
- [井上 10] 井上 克巳, 田村 直之: SAT ソルバーの基礎, 人工知能学会誌, Vol. 25, No. 1, pp. 57–67 (2010)
- [Jong 02] Jong, H. D.: Modeling and simulation of genetic regulatory systems: A literature review, *Journal of Computational Biology*, Vol. 9, pp. 67–103 (2002)
- [Kautz 92] Kautz, H. and Selman, B.: Planning as satisfiability, in *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI '92)*, pp. 359–363 (1992)
- [Keseler 09] Keseler, I., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R., Johnson, D., Krummenacker, M., Nolan, L., Paley, S., Paulsen, I., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A., and Karp, P.: EcoCyc: A comprehensive view of *Escherichia coli* biology, *Nucleic Acids Research*, Vol. 37, pp. D464–D470 (2009)
- [Klamt 02] Klamt, S. and Stelling, J.: Combinatorial complexity of pathway analysis in metabolic networks, *Molecular Biology Reports*, Vol. 29, No. 1–2, pp. 233–236 (2002)
- [Koshimura 09] Koshimura, M., Nabeshima, H., Fujita, H., and Hasegawa, R.: Minimal model generation with respect to an atom set, in *Proceedings of the 7th International Workshop on First-Order Theorem Proving (FTP '09)*, pp. 49–59 (2009)
- [Küffner 99] Küffner, R., Zimmer, R., and Lengauer, T.: Pathway analysis in metabolic databases via differential metabolic display (DMD), in *German Conference on Bioinformatics*, pp. 141–147 (1999)
- [MiniSat] MiniSat, : <http://minisat.se/MiniSat.html>
- [鍋島 10] 鍋島 英知, 宋 剛秀: 高速 SAT ソルバーの原理, 人工知能学会誌, Vol. 25, No. 1, pp. 68–76 (2010)
- [Niemelä 96] Niemelä, I.: A tableau calculus for minimal model reasoning, in *Proceedings of the Theorem Proving with Analytic Tableaux and Related Methods (TABLEAUX 1996)*, Vol. 1071 of *LNCS*, pp. 278–294 (1996)
- [Planes 09] Planes, F. J. and Beasley, J. E.: Path finding approaches and metabolic pathways, *Discrete Applied Mathematics*, Vol. 157, No. 10, pp. 2244–2256 (2009)
- [Rahman 05] Rahman, S. A., Advani, P., Schunk, R., Schrader, R., and Schomburg, D.: Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC), *Bioinformatics*, Vol. 21, No. 7, pp. 1189–1193 (2005)
- [Ray 09] Ray, O., Whelan, K. E., and King, R. D.: Logic-based steady-state analysis and revision of metabolic networks with inhibition, in *Proceedings of the 2nd International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2009)*, pp. 661–666 (2009)
- [Schaub 09] Schaub, T. and Thiele, S.: Metabolic network expansion with answer set programming, in *Proceedings of the 25th International Conference on Logic Programming (ICLP 2009)*, pp. 312–326 (2009)
- [Schuster 00] Schuster, S., Fell, D. A., and Dandekar, T.: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nature Biotechnology*,

Vol. 18, pp. 326–332 (2000)

- [Sinz 05] Sinz, C.: Towards an optimal CNF encoding of Boolean cardinality constraints, in *Proceedings of the Principles and Practice of Constraint Programming (CP 2005)*, pp. 827–831 (2005)
- [Tamura 09] Tamura, T., Takemoto, K., and Akutsu, T.: Measuring structural robustness of metabolic networks under a Boolean model using integer programming and feedback vertex sets, in *Proceedings of the 2nd International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2009)*, pp. 819–824 (2009)
- [Terzer 09] Terzer, M., Maynard, N. D., Covert, M. W., and Stelling, J.: Genome-scale metabolite networks, *Systems Biology and Medicine*, Vol. 1, No. 3, pp. 285–297 (2009)
- [Tiwari 07] Tiwari, A., Talcott, C. L., Knapp, M., Lincoln, P., and Laderoute, K.: Analyzing pathways using SAT-based approaches, in *Proceedings of the Second International Conference on Algebraic Biology (AB '07)*, Vol. 4545 of *LNCS*, pp. 155–169 (2007)
- [Tomlin 07] Tomlin, C. J. and Axelrod, J. D.: Biology by numbers: mathematical modelling in developmental biology, *Nature Reviews Genetics*, Vol. 8, No. 5, pp. 331–340 (2007)

[担当委員: 桜井 成一郎]

2011 年 2 月 8 日 受理

著者紹介

宋 剛秀 (正会員)

2004 年神戸大学工学部電気電子工学科卒業。2006 年同大学院自然科学研究科電気電子工学専攻修了。サントリー(株)を経て、2011 年総合研究大学院大学複合科学研究科情報学専攻博士後期課程修了。同年より新領域融合研究センター融合プロジェクト特任研究員。博士(情報学)。SAT 問題を含む制約充足問題の研究に従事。2009 年度人工知能学会全国大会優秀賞受賞。

井上 克巳 (正会員)

1982 年京都大学工学部数理工学科卒業。1984 年同大学院工学研究科数理工学専攻修士課程修了。松下電器産業(株)、(財)新世代コンピュータ技術開発機構、豊橋技術科学大学、神戸大学を経て、2004 年より国立情報学研究所。現在、同情報学プリンシプル研究系教授および総合研究大学院大学複合科学研究科情報学専攻教授。京都大学博士(工学)。人工知能、システム生物学などの研究に従事。