

酵素反応式中の化合物グラフマッチング 簡略化手法の開発

小寺正明^{†1} 時松敏明^{†1} 中川善一^{†1}
守屋勇樹^{†1} 金久實^{†1,†2} 五斗進^{†1}

多様な二次代謝化合物の生合成経路を推定するには、酵素反応式中で変化または保存する化学構造を求める必要がある。これは、反応式の両辺を構成する化合物をそれぞれグラフとして定義しマッチングを求めることに相当するが、酵素反応の性質を考慮することで一般的なグラフマッチングより簡略な解法が可能である。本研究では、与えられた完全または部分反応式から原子の移動を表す化合物ペアを自動抽出する方法を報告する。

Development of a simplified method for compound graph matching in enzyme reactions

MASAAKI KOTERA[†] TOSHIAKI TOKIMATSU[†]
ZEN-ICHI NAKAGAWA[†] YUKI MORIYA[†]
MINORU KANEHISA^{†,††} and SUSUMU GOTO[†]

Identifying conserved or changing chemical substructures in enzyme reactions facilitates predicting possible biosynthetic pathways of various secondary metabolites. This issue can be formulated as a matching problem where each side of the reaction equation is depicted as a graph. Graph matching is known to be NP-hard, but with the nature of enzyme reactions, simpler solutions are possible. In this study, we developed a simplified method to decompose a complete or partial equation into a set of compound pairs representing the flow of atoms.

^{†1} 京都大学化学研究所バイオインフォマティクスセンター
Bioinformatics Center, Institute for Chemical Research, Kyoto University

^{†2} 東京大学医科学研究所ヒトゲノム解析センター
Human Genome Center, Institute of Medical Science, University of Tokyo

1. 背景と目的

ゲノムにコードされている様々な酵素タンパク質は生物固有の代謝経路を構成することで、その生物の生存戦略に貢献している。現在知られている酵素反応は現在 KEGG データベース¹⁾ に登録されているだけでも約 8,400 反応に及ぶが、実験的に活性が確認された酵素のタンパク質配列が登録されているものは約 4,300 反応に過ぎず、大半の酵素反応は酵素タンパク質やその遺伝子との対応関係が不明なまま残されている。また、ガスクロマトグラフィーや高速液体クロマトグラフィー、質量分析法や核磁気共鳴分光法などの分析技術の向上により、細胞の活動によって生じる様々な分子を網羅的に分離・検出し分析するメタボロミクスが近年発達している。このことは、既に代謝経路が同定されている化合物についてその濃度変化を網羅的に調べることを可能にすただけにとどまらず、今まで存在の知られていなかった化合物が新たに検出される可能性を高めている。これによって、代謝経路が不明な化合物の数は大幅に増えることが予想される。

このような化合物の多くは、二次代謝産物と呼ばれる化合物群に分類される。二次代謝産物とは、生物が生合成する化合物のうち、その生物自身が成育する上で直接は必要がない（と考えられていた）化合物のことを指す。有名なものとして抗菌物質や色素がある。たとえばアオカビは抗菌物質ペニシリンを合成して周囲の細菌の発育を阻害する。しかし、もしも周囲に敵（細菌）が存在しなければペニシリンは不必要であると考えられる。このような二次代謝産物は他にも植物や菌類などに多く見られ、それらの代謝経路を明らかにし、それを構成する酵素タンパク質や遺伝子を同定することは、代謝医学・薬学的あるいは工学的に大きな価値がある。そのためには、実験手法の開発は勿論ではあるが、情報学的な予測手法の確立も欠かせない。

我々は生物体内で起こる可能性のある酵素反応を酵素タンパク質や酵素遺伝子と関連づけるために、酵素反応の化学構造情報に基づいて適切な酵素番号 (EC number)^{2),3)} などの反応分類に自動的に割り当てる方法の研究を行い⁴⁾、それをウェブプログラム E-zyme として実装した⁵⁾。ここで酵素番号とは国際生化学分子生物学連合 (IUBMB) の酵素委員会 (EC)²⁾ により管理されている酵素分類であり、代謝化合物の化学情報とそれを代謝する酵素タンパク質の配列情報とをつなぐ基本情報を提供している。

E-zyme で用いた方法のうち、化学構造変換パターン抽出法を示したのが図 1 である。化合物はグラフ⁶⁾と呼ばれるデータ構造で表され、反応前後で得られる化合物の構造比較から、保存されている部分構造は共通部分グラフとして、反応により結合あるいは脱離した部分構造はそれ以外の部分グラフとして表される。その境界領域はその酵素反応の反応中心 (reaction center) であると見なせる。化合物グラフでは各原子は官能基などの情報を持つ頂点（たとえば N1a = アミノ基）として表され、反応中

心はこれらの情報を連結した RDM 化学構造変換パターン (RDM chemical transformation patterns) として計算機的に表現される. 生成された RDM パターンに対する酵素番号は, 既知の酵素反応のパターンとの類似性に基づき選ばれる.

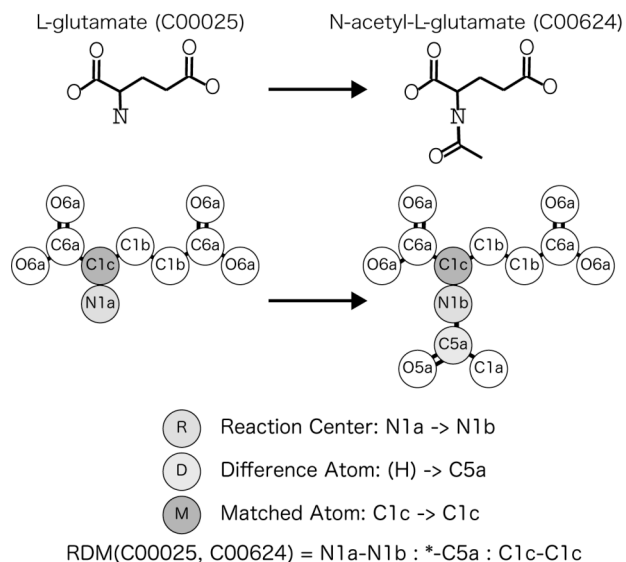


図 1 化合物構造比較と RDM パターン

Figure 1 Chemical structure comparison and the RDM pattern.

現バージョンの E-zyme では, このように構造比較を行うべき基質・生成物ペア (reactant pair) を入力時に指定する必要がある. 基質・生成物ペアは酵素反応式中の原子の移動を表した二部グラフ (ここでは reactant pair graph と呼ぶことにする) の辺として表現できる. 典型的な酵素反応の基質・生成物ペアを図 2 に示した (水素原子の流れに注目したペアは点線で表されているが, これらは省略され基質・生成物ペアとは見なされない). E-zyme の教師セットとなる RPAIR データベース⁷⁾を作成する際にも, 既知の酵素反応式から変換する作業を手作業で行っている. RPAIR データベース作成においては, 手作業で指定された各々のペアについて化合物構造比較プログラムである SIMCOMP⁸⁾を用いて化合物構造比較を行い, 得られた結果をさらに手作業で精査して反応中心を同定している.

酵素反応式を基質・生成物ペアの組み合わせ (reactant pair graph) として表現するこ

とは, メタボローム研究の視点からも, 情報学的な視点からも利点がある. まず, メタボローム研究の視点からは, 様々な技術によって直接検出されるのは化合物の集合であり, 酵素反応式の集合ではない. したがって, 代謝経路を予測したり酵素タンパク質や遺伝子との関係を解析したりするには, 化合物の集合から任意の 2 つを取り出し, その間をつなぐ反応が存在するかどうかを考える必要がある^{9),10),11),12),13),14)}. そのときに, 酵素反応を完全な反応式ではなく reactant pair graph として表現し reactant pair 単位で解析できるようにすると便利である. 実際に, 代謝経路予測ウェブツール PathPred⁹⁾ では既知の酵素反応の情報を reactant pair の形で蓄え, ユーザが興味のある化合物を入力するとそれがどのように代謝されるかを連続的に予想しツリー状の予測経路が描画される.

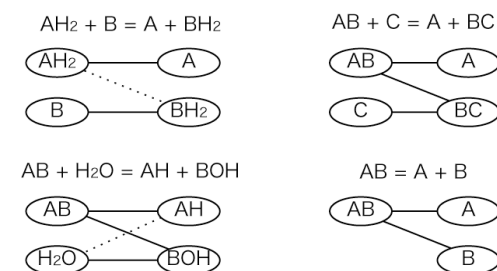


図 2 酵素反応式と基質・生成物ペアグラフ

Figure 2 Enzyme reaction equations and reactant pair graphs.

また, 情報学的な観点からは, 基質・生成物ペアに分解することによりグラフマッチングに必要な計算量が少なく済む利点がある. 例えば図 2 の右上にある “ $AB + C = A + BC$ ” という酵素反応式は, 化合物 AB から化合物 C へ部分構造 B が転移して化合物 A と化合物 BC を生成する反応を表しており, このときペア AB-A は部分構造 A を, ペア AB-BC は部分構造 B を, ペア C-BC は部分構造 C をそれぞれ保存しているため結線が引かれるが, ペア C-A には引かれない. このことは, 酵素反応式中で変化した化学結合を同定するにあたって, 化合物 C と化合物 A の化学構造を比較する必要がないことを示している. したがって, 正しいペアを選ぶ効率を高めることで, 反応式全体を直接扱うよりも少ない計算量でマッチングできるようになると考えられる.

本研究では, この基質・生成物ペアを, 与えられた酵素反応式から自動的に予測する方法の開発を行った. 2 つの基質が反応し 2 つの生成物が生じる反応 $C1 + C2 = C3 + C4$ を例に説明すると, 可能な基質・生成物ペアは図 3 に示したように全 7 通

り存在する。本研究で扱う問題は、その中でどれが実際の反応を反映したものであるかを予測する問題に相当する。ペアの予測は、与えられた反応の両辺に含まれる化学構造をそれぞれグラフとして表現したあと、複数の部分構造に切断し、両辺の部分構造式の組み合わせ最適化問題を解くことにより実現した。計算結果は手作業による基質・生成物ペアのデータベース RPAIR との一致度により評価し、約 75% の精度を得た。今後は予測精度の向上および、ウェブツールへの実装を行う予定である。これによって、予測された酵素反応から酵素タンパク質や遺伝子への対応付けがより効率化することが期待できる。

2. データと方法

本研究ではまず、KEGG データベースに登録されていた 8,421 反応のうち、原子の情報のない糖鎖を含む 396 反応を除去し 8,025 反応を得た。その中で両辺のうち一方でも化合物数が 1 である反応 (図 2 の右下に相当) は 2167 あり、これらは基質・生成物ペアが自明であるため本法は不要であると判断し除去した。最終的に残った 5,858 反応を本研究の対象として用いた。

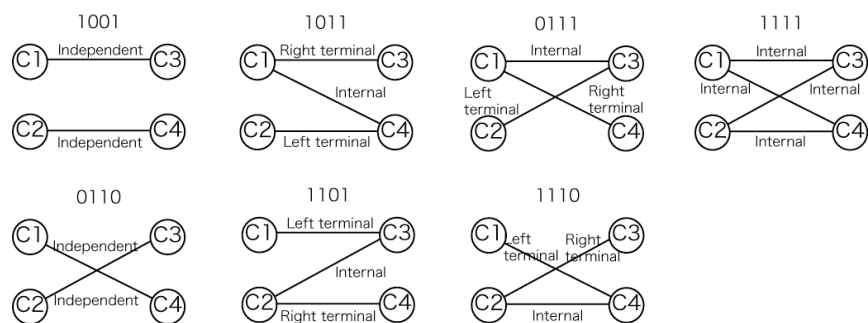


図 3 酵素反応 $C1 + C2 = C3 + C4$ を構成する基質・生成物ペア

Figure 3 Possible reactant pair graphs forming the equation $C1 + C2 = C3 + C4$.

今回の方法の正解データとして、基質・生成物ペアを格納している RPAIR データベースを用いた。ここで、基質・生成物ペアについていくつかの単語を定義しておく。酵素反応 $C1 + C2 = C3 + C4$ を例にとると、可能な化合物ペアは C1-C3, C1-C4, C2-C3, C2-C4 の 4 通りである。実際の酵素反応式を表現する際に、これらのペアの有無をビット列で表すと、0000 - 1111 まで全 16 通りの reactant pair graph が得られ、次数 0 の頂点を含むものを除去すると 7 通りになる (図 3)。このうち 1001 と 0110 は典

型的な酸化還元反応に見られ (図 2 の左上に相当)、グラフのトポロジーとしては同形であり、2 つの辺は互いに離れて存在する。これらの辺を “Independent pair” と呼ぶことにする。一方、1011, 1101, 0111, 1110 も互いに同形であり、S字型の reactant pair graph として表せる。このうち、グラフ 1011 の C1-C3 のように右辺の頂点の次数が 1 であるものを “Right terminal pair”, C2-C4 のように左辺の頂点の次数が 1 のものを “Left terminal pair”, C1-C4 のように両辺の頂点が 2 以上のものを “Internal pair” と呼ぶことにする。この S 字型の reactant pair graph は典型的な転移反応に見られる (図 2 の右上に相当)。最後の 1111 は 8 の字型のグラフであり、転移反応に見られるが数は少ない。

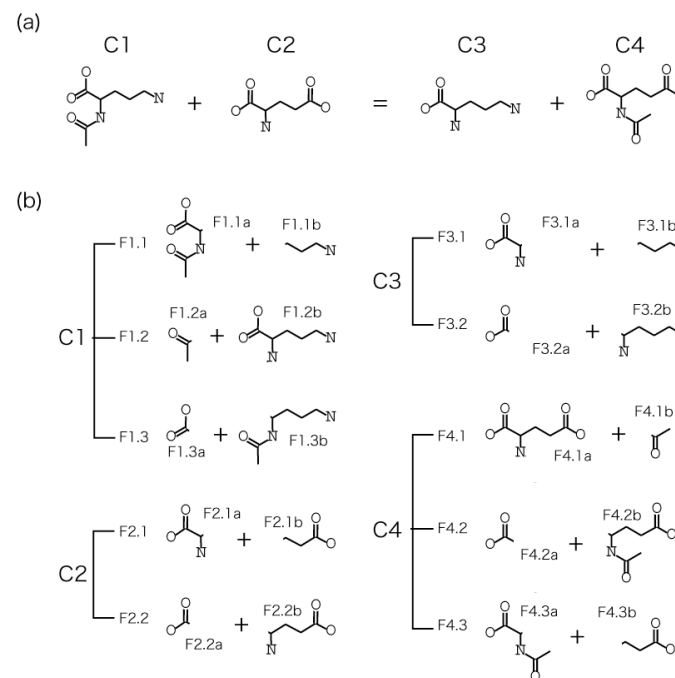


図 4 部分構造フラグメントへの分解

Figure 4 Fragmentation of compounds.

今回用いた方法を、酵素番号 2.3.1.35 に分類される酵素反応 $C1 + C2 = C3 + C4$ を例に説明する (図 4a)。まず化合物 C1, C2, C3, C4 それぞれについて、任意の結合

を切断して生成した仮想的な断片を考えた. 化合物に含まれる結合の数を b とすると, 切断の方法は全部で $b!$ 通り存在するが, ここでは説明を簡単にするため, 例えば化合物 $C1$ からは3通りの断片セット (F1.1, F1.2, F1.3) しか生じず, 同様に化合物 $C2$, $C3$, $C4$ からはそれぞれ 2, 2, 3 通りの断片セットしか生成しないものとする (図 4b).

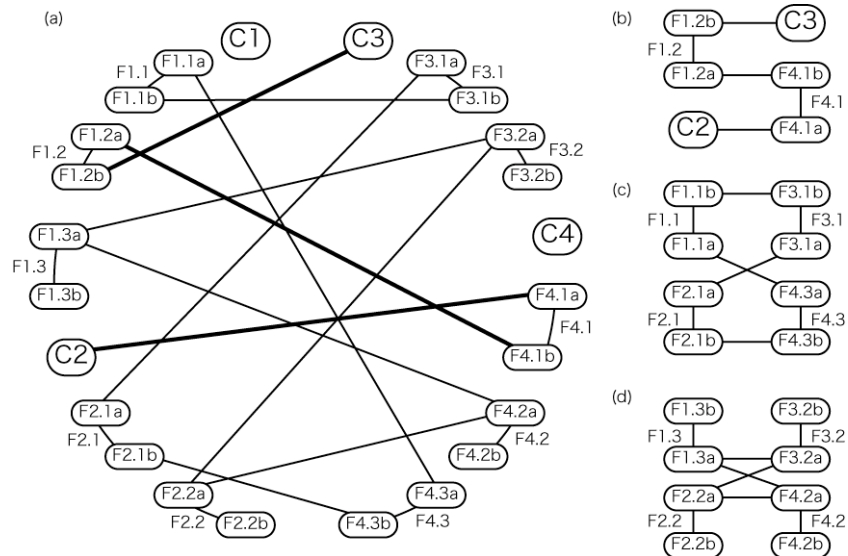


図 5 基質・生成物ペアを探索するための部分構造組み合わせグラフ
Figure 5 Fragment matching graph to find optimal reactant pairs.

次に, 左辺と右辺の化合物グラフ, および断片グラフをそれぞれ頂点とする新しい Ac グラフを考えた (図 5a). 両辺の化合物および断片グラフをそれぞれ左右に配置し, ある切断法により生じる断片グラフ間 (たとえば F1.1a と F1.1b) を辺で結んだ. また, 左右の辺で同形の化合物または断片グラフ間も辺で結んだ. たとえば左辺の化合物 $C1$ から生じた断片 F1.1a は右辺の化合物 $C4$ から生じた断片 F4.3a と同一であるので, 辺で結ばれた. このようなグラフを定義することにより, 左右で原子の過不足が極小になるような部分グラフを発見する問題として表現することができ, 選ばれた最大部分グラフが基質・生成物ペアを, 選ばれたフラグメント対が化学構造アラインメントを表す問題となる.

以上に示したような方法をそのまま適用すると, たとえば2つの化合物が反応して2つの化合物を生成する反応においてそれぞれの化合物の結合数をそれぞれ b_1 , b_2 , b_3 , b_4 とすると $(b_1 + 1)!(b_2 + 1)!(b_3 + 1)!(b_4 + 1)!$ 通りの中から1つを選ばなければ

ならないため, 大変効率が悪い. そこで本法では, この問題の効率を高めるために様々なヒューリスティクスを用いた.

まず使ったのは, 既知反応から得られる reactant pair graph のトポロジーの偏りである. RPAIRデータベースに登録されているペアから各反応を表す reactant pair graph を構築すると, $C1 + C2 = C3 + C4$ の形をとる 3,020 反応のうち最初のトポロジー (1001 または 0110, 図3) の reactant pair graph は 226 反応, S 字型の reactant pair graph は 2411 反応, 8 の字型は 165 反応存在した (残りの 39 反応と 179 反応は化学構造が曖昧であるなどの理由で, ペアがそれぞれ 0 個または 1 個しか定義されていない反応であった). 続いて reactant pair graph 中に現れる頂点 (化合物) の次数の分布を調べたところ, 次数 1 の頂点は全体の 65.8%, 次数 2 の頂点は 32.5% であり, 次数 3 以上の頂点は全体の 1.7% に過ぎなかった. このように, 可能な二部グラフのうち実際に見られる reactant pair graph のトポロジーには偏りがあることが分かった.

この知見を使うことで, (既に図 5a で例示したように) 1 つの化合物から得られる断片の数最大 2 つになるように切断する結合を制限でき, 扱える酵素反応の数をほとんど減らすこと無く計算量を大幅に減らした. また, reactant pair graph の頂点の次数は 1 が多いことを利用して, Left-terminal または Right-terminal pair から探索する方法を利用した. 図 5a に示したグラフから結合成分を取り出すと図 5b, 5c, 5d が得られるが, そのうち Left-terminal または Right-terminal pair を含むものは図 5b だけである. そして実際にこの図 5b が正しく原子の移動を表す reactant pair graph であった.

実際の手順は次の通りである. まず, 酵素反応中で指定されたプロトン (H^+) 以外の化合物を 2 通りの方法で表現した. 1 つめは水素以外の原子を頂点とし結合を辺としたグラフであり, もう 1 つは水素原子を無視した構造式である (例えば図 4 の $C1$ は $C_7N_2O_3$ と表した). 次に, 両辺の分子の構造式を比較して, 同じ構造式を持つペアを抜き出し Independent pair とした. たとえば $C1 + C2 + NAD^+ = C3 + C4 + NADH + H^+$ のような酵素反応式があった場合, まず H^+ が無視され, NAD^+ と $NADH$ がどちらも同じ $C_{21}N_7O_{14}P_2$ と表現されるので Independent pair として取り出され, 残された部分反応式 $C1 + C2 = C3 + C4$ だけを取り扱えばよいことにした.

続いて, 残された部分反応式の両辺の分子を総当たりで比較し, Left terminal または Right terminal pair になる可能性を検討した. まず両辺の分子を構造式で比較し, 一方が他方に含まれるか調べた. たとえば図 4a のように化合物 $C1$ と $C3$ の構造式がそれぞれ $C_7N_2O_3$ と $C_5N_2O_2$ だった場合, 共通部分は最大で $C_5N_2O_2$ であり, 両辺から $C_5N_2O_2$ を除くと左辺には C_2O_1 が余り, 右辺には何も余らない. このようとき大きい化合物 $C1$ のグラフ構造をチェックし, 最小の切断で $C_5N_2O_2$ と C_2O_1 の断片を得るような結合を幅優先 (深さ優先でもよい) 探索し (図 4b の F1.2 がそれに

あたる), そのような結合が存在する場合, そのペアは **Right terminal pair** になる可能性のあるペアとして記録した. 同様に, C2 (C₅N₁O₄) と C4 (C₇N₁O₅)を比較すると C₅N₁O₄を共通部分とし, 右辺に C₂O₁ が余り左辺に何も余らない **Left terminal pair** 候補を得た. 一方 C1 と C4 を比較した場合は, 共通部分 C₇N₁O₃ が得られるが, 左辺に N₁ 右辺に O₂ が余るので **terminal** には成り得ない. 同様に C2 と C3 も右辺に O₂ 右辺に N₁ が余るので **terminal** にはなりえない. つまりこれらのペアは, もしも **reactant pair graph** 中に組み込まれるとしても **internal pair** にしかなりえない.

このようにして, 酵素反応式中の化合物ペアの総当たりに対して, **Independent pair**, **Terminal pair**, **Internal pair** としてラベル付けした. このうち **Terminal pair** と **Internal pair** をそれぞれ頂点とする新しいグラフを考え, ある頂点の「余り」が他の頂点の「共通部分」と一致すれば辺を結んだ. そして, 共通部分が最も大きい頂点から深さ優先探索を行うことで連結成分を取り出し, それを **Independent pair** と合わせることで **reactant pair graph** を得た.

3. 結果と考察

結果は RPAIR データベースで定義された化合物ペアとの一致度で評価した. 本研究との比較として, まずはランダムに予想した場合の的中率を以下のように計算した. 対象とする反応中, 左右の化合物の可能な組み合わせは 32,349 ペアであった. その中で RPAIR データベース中で定義されているのは 16,087 ペアだった. そこで, 酵素反応式と与えられた場合に 16,087 / 32,349 = 0.497 の確率でランダムに基質・生成物ペアを割り当ててみたところ, 36% の割合で **reactant pair graph** が一致した. 一方, 本研究で用いた方法では 75% の割合で一致した. ペア単位で見た場合, RPAIR データベース中で定義されていたペアのうち本法で予測できたペアは 85%, 両辺の化合物の組み合わせを総当たりで取った場合に出現するが実際には RPAIR 中で定義されていないペアのうち, 本法でも解答として出力しなかったペアは 90% であった. ペア単位で見た場合に比べグラフ単位で見た場合に正解率が著しく落ちているのは, ペアが 1 つでも間違っていればグラフが間違っていると判定しているためである.

今回の研究の第一の目的は, 酵素反応から反応中心を発見し酵素タンパク質や酵素遺伝子と関連づけるにあたって, 与えられた反応式を基質・生成物ペアの組み合わせに分解することによって, 必要な化学構造アラインメントの回数を減らすことにある. 言い換えれば以前に開発した **E-zyme** を行うための前処理なので, この前処理を行うことによってかえって計算時間が増えたり予測精度が落ちたりしないようにする必要がある. そこで本法では化合物グラフマッチングを直接は行わず, 基本的には構造式のマッチングだけで済ませ, 必要が生じた場合だけ化合物グラフを分解し, 断片のマッチングも構造式のマッチングだけで済ませるだけにして必要な計算時間を極力

節約した. その結果, 1.33 GHz PowerPC G4 (1GB DDR SDRAM) で計算して 1 反応あたりの平均計算時間は約 0.11 秒まで縮まった. ただし **reactant pair graph** 単位での正解率はまだまだ改善の余地は残されていると思われる, 酵素番号などの予測に与える影響についてはまだ調査中である.

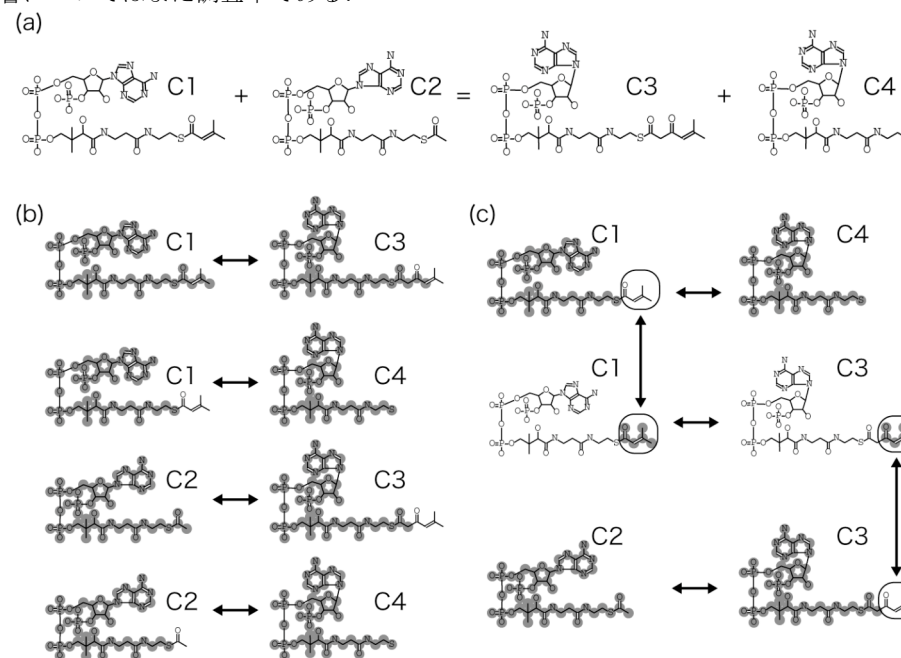


図 6 今回の方法により改善した例
Figure 6 An example improved matching.

しかしながら, 本法の利用による好影響がありそうな例がいくつか見られる. 現在のバージョンの **E-zyme** は, 与えられた基質・生成物ペアが複数であってもそれぞれ独立に化学構造比較を行うため, 得られた複数の化学構造アラインメント間に矛盾が生じる場合がある. 今回の方法の応用により, 基質・生成物ペア中でマッチング可能な左辺の断片と右辺の断片が限定されるため, 酵素反応式として与えられた両辺の化合物を全体として矛盾なく構造比較することが可能にできると考えられる. たとえば図 6a に示したような反応の両辺を総当たりで構造比較した場合, 図 6b に示した 4 つのペアでグレーで示したマッチングが得られる. これらのうち実際の原子の移動を反映したマッチングは C1-C4 と C2-C3 の 2 つだけである. 本法では, C1-C3 のよ

うなペアは両辺に余りが存在するため Terminal pair とは見なされず後回しにされ、次に C1-C4 と C2-C3 が Terminal pair として同定された。続いて、それぞれの余りを補うペアとして、図 6b 中のものとは異なるマッチングをした C1-C3 (図 6c) が得られた。このように本法は化学構造比較の回数を減らしながらも、反応式全体を眺めた効果的なマッチングを実現している。

また、図 7 に示したように、不完全な反応式でも十分なマッチングが得られる例が見られた。完全な酵素反応式 $C1 + C2 + H_2O = C3 + C4 + H^+$ (図 7a) を入力したとき、プロトン H^+ は水素原子しか含まないので基質・生成物ペアを形成せず無視され (図 7b)、Left terminal pair が 2 つと Independent pair が 2 つの reactant pair graph が形成された (図 7c)。このグラフは水素原子以外の原子の移動を正しく表している。一方、不完全な酵素反応式 $C1 + C2 = C3 + C4$ を入力した場合、図 7d のような reactant pair graph が得られた。水分子 H_2O のマッチング情報が欠落しているが、E-zyme で酵素番号を予測する際には水分子を含む基質・生成物ペアは必要としないので、ここで得られたペアで十分である。

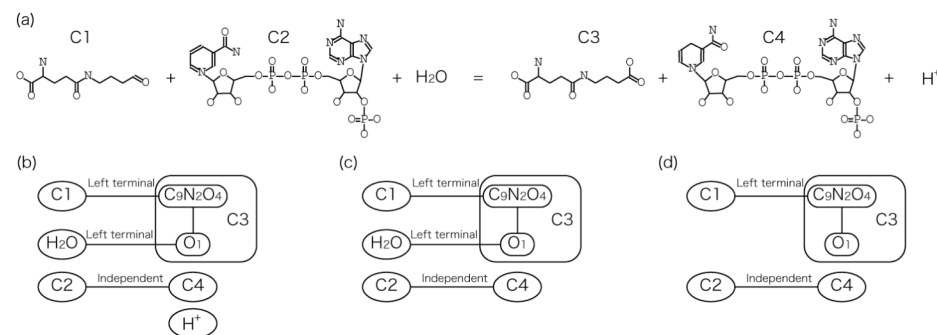


図 7 不完全な反応式のマッチング例
Figure 7 An example matching of incomplete equations.

4. 結論

本研究では、酵素反応式から基質・生成物ペアを予測する方法を開発することで、EC 番号予想プログラム E-zyme の機能向上を試みた。今後は計算時間を極力犠牲にせずペア予測精度を改善する方法を検討し、また酵素番号予測精度に対する影響などを調査し、実用化につなげる予定である。

謝辞 本研究は文科省統合データベースプロジェクトのサポートを受けています。

また、KEGG は JST バイオインフォマティクス推進機構のサポートを受けています。

参考文献

- 1) Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M.: KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Res*, Vol.38, D355-D360 (2010).
- 2) Tipton, K.F., and Boyce, S.: Enzyme Classification and Nomenclature. In: *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, Chichester: <http://www.els.net> (2005).
- 3) McDonald, A.G., Boyce, S., and Tipton, K.F.: ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res*. Vol.37, D593–D597. (2009).
- 4) Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M.: Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions, *J. Am. Chem. Soc*, Vol.126, pp.16487-16498 (2004).
- 5) Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M.: E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs, *Bioinformatics*, Vol.25, i79-i86 (2009).
- 6) Biggs, N.; Lloyd, E. and Wilson, R.: *Graph Theory*, 1736-1936. Oxford University Press (1986).
- 7) Kotera, M., Hattori, M., Oh, M., Yamamoto, R., Komono, T., Yabuzaki, J., Tonomura, K., Goto, S., and Kanehisa, M.: RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. *Genome Informatics*, Vol.15, P062 (2004).
- 8) Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M.: Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways, *J. Am. Chem. Soc*, Vol.125, pp.11853-11865 (2003).
- 9) Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.: PathPred: an enzyme-catalyzed metabolic pathway prediction server, *Nucleic Acids Res*, Vol.38, W138-W143 (2010).
- 10) Langowski, J., and Long, A.: Computer systems for the prediction of xenobiotic metabolism. *Adv. Drug Deliv. Rev.*, Vol.54, 407-415 (2002).
- 11) Klopman, G., and Tu, M.: Structure-biodegradability study and computer automated prediction of aerobic biodegradation of chemicals. *Environ. Toxicol. Chem.* Vol.16, 1829-1835 (1997).
- 12) Faust, K., Croesa D., and van Helden, J.: Metabolic Pathfinding Using RPAIR Annotation. *J. Mol. Biol.*, Vol.388, pp390-414 (2009).
- 13) Ellis, L. B. M., Roe, D., and Wackett, L. P. The University of Minnesota biocatalysis/biodegradation database: the first decade. *Nucleic Acids Res.*, Vol.34, D517-D521 (2006).
- 14) Kotera, M., McDonald, A.G., Boyce, S., and Tipton, K.F.: Eliciting Possible Reaction Equations and Metabolic Pathways Involving Orphan Metabolites, *J. Chem. Inf. Model*, Vol.48(12), pp.2335-2349 (2008).