

平成21年 6月 1日現在

研究種目：若手研究（A）

研究期間：2006～2008

課題番号：18681033

研究課題名（和文） 環境因子および二次代謝産物の反応経路予測法の開発

研究課題名（英文） Development of reaction prediction system for environmental chemicals and secondary metabolites

研究代表者

服部 正泰（HATTORI MASAHIRO）

京都大学・化学研究所・助教

研究者番号：60372554

研究成果の概要：

本研究では、代謝反応パスウェイのうち未解明のまま残されている部分、すなわち、薬物や環境因子などの細胞にとって通常外来物質と見られる化合物の代謝反応経路、および、植物の二次代謝産物の生合成経路などについて、代謝反応に関するケミカル情報を的確に表現する手法の開発、それら新しい知識をデータベースとして体系化する作業とデータの計算機的利用法の開発、酵素反応データに基づいた代謝経路の予測法の実現を行った。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	4,100,000	1,230,000	5,330,000
2007年度	3,700,000	1,110,000	4,810,000
2008年度	1,900,000	570,000	2,470,000
年度			
年度			
総計	9,700,000	2,910,000	12,610,000

研究分野：バイオインフォマティクス

科研費の分科・細目：ゲノム科学・応用ゲノム科学

キーワード：バイオインフォマティクス、ケモインフォマティクス、ケミカルゲノミクス、データベース、代謝パスウェイ解析、酵素反応機構、反応予測

## 1. 研究開始当初の背景

代謝化合物と酵素反応よりなる代謝反応パスウェイのうち、生化学的実験を経て既知となった部分パスウェイの情報は、たとえば国内ではKEGG PATHWAYに代表されるパスウェイデータベースに収録されていく。このようなパスウェイデータは、遺伝子の機能アノテーションのほか、細胞シミュレーションなどのポストゲノム解析に対して積極的に応用されるようになってきている。一方で、パスウェイデータベースに登録されるデータは、基

本的に生化学的実験の成果に頼っている部分が多く、収録されるデータの絶対量やその片寄りに関する問題を避けて通ることができない。たとえば、薬物や環境因子などの細胞にとって通常外来物質と見られる化合物が、如何に代謝され無害化されていくかという反応経路は、ほとんど未解明のままである。また、植物の二次代謝産物の生合成経路なども未知の部分が多い。このため、代謝反応に関するケミカル情報を的確に表現する手法の開発、それら新しい知識をデータベースと

して体系化すること、およびデータの計算機的使用法の開発、酵素反応データに基づいた代謝経路の予測、得られた結果の生物学的意義付けなど、既存のケミカルゲノミクス研究を発展させた新規解析手法の開発が必須である。

## 2. 研究の目的

ケミカルゲノミクスの流れに沿った、化学情報データベースの構築プロジェクトは、世界中でいくつも立ち上がっている。例えば、申請者が研究分担者となっている特定ゲノム4領域においても化学情報のデータベース化が行われており、酵素反応を含む全有機化学反応を原子レベルでのテンプレート化、およびケミカル情報の表現、分類、予測を行うための基盤となるデータベース構築が行われている。このような反応データベース自体がすでに国際的に見ても特異なものであるが、本研究課題では、特にこのデータベースを用いたより生物学的・薬学的な解析を行うための基礎技術の習得と応用解析手法の新規開発を目指す。このような研究課題は、学術的な必要性が高いにもかかわらず、国際的にも前例の少ない非常に独創性の高いものである。前例が少ない理由の一つは、利用可能なケミカルデータが限定的であるうえに、データベース化の作業のほとんどが手作業に依存する部分が多いことが原因であると思われる。本研究課題では、このような手作業でしか行えなかった処理工程の半自動化を目指しており、これにより、我が国発の最先端のケミカルゲノミクス解析を目指すものである。この研究課題によって得られる成果は、我が国内のケミカルゲノミクス研究コミュニティのみならず、広く国際社会にも多大なインパクトを与えると予想できる。本研究課題の具体的な内容は以下の通りとする。

(1) KEGG LIGAND データベースの一部として構築していた酵素反応毎の原子レベルでの表現手法に関連して、これらを計算機的に自動抽出するアルゴリズムの開発と、そのための酵素科学的知識の蓄積を行う。

(2) 以上の成果を基に、酵素反応機構を高精度に予測するシステムを構築する。

(3) 個別のルールとして定義した個々の酵素反応を複数組み合わせることによって、外来化合物や二次代謝産物を代謝し得る生体内パスウェイを、計算機的に再構築する手法の開発を行う。

(4) 上記の目的を達成するためには、膨大なグラフ計算が必要になるとと思われるので、それを回避するためのヒューリスティクスの探索と導入も行う。

## 3. 研究の方法

(1) KEGG LIGAND データベースの一部として構築した、酵素反応前後における化合物構造の原子変換パターンデータベース RPAIR に関連して、これらの情報を計算機的に自動抽出するためのアルゴリズム開発を行う。そのために、酵素科学的知識の蓄積を行い、これを自動抽出プログラムのヒューリスティクスとして導入する。この酵素科学的知識に関しては、申請者の従来の研究業歴によって得られた経験的知識をベースとし、これにさらに、酵素科学系のテキスト等を用いて、必要な情報を極力補完していく。なお、アルゴリズムの開発・改良は原則としてパーソナルコンピュータ上で行うことが可能であるので、安価なシステムにより開発環境を構築する一方、計算に必要なデータファイルの管理や本研究課題に特化したデータベースの構築作業、アルゴリズムの適合性を検証するための計算サーバーおよび作業環境などは、京都大学化学研究所の所有する各種ワークステーション/サーバー上で構築する。以上の方法により開発したアルゴリズムによって抽出した原子変換パターンは、酵素反応の基本的なルール群としてサーバー上でライブラリ化し、代謝パスウェイ全体に渡って統計を取ることににより、生体内において生物がより好んで用いる酵素反応セットを導出することに用いる。ここで導出された知識は連続する酵素反応のルールとして再びライブラリ化することが可能であり、反応経路予測の精緻化に応用可能である。

(2) 以上の研究成果を基にして、代謝パスウェイの予測を行う際のキーステップとなる酵素反応機構の予測を行うシステムの構築を行う。具体的には、上記の手法によって抽出される個別の酵素反応ルール（原子変換パターン）が、与えられた化合物構造に適用可能であるかどうかを、グラフ情報を基に判定して行く。この際、原子変換パターンのすべての情報を適用するのか、あるいはその一部のみを適用するのか、など様々なヒューリスティクスが考えられるので、それらを考慮した上でより高精度なシステムの構築を目指す。また、予測された酵素反応が各生物種のゲノムによって規定される反応空間と照らし合わせて妥当であるかどうかをユーザーが判断出来るようにするために、ゲノム情報と結びつけるためのリンクを付与出来るようにするなど、ユーザビリティの向上も行って行く。

(3) 上記の研究過程において酵素反応ルールとしてライブラリ化しておいた情報を、互いに組み合わせることによって、生体内パスウェイを再構成するためのアルゴリズム開発を行う。これは、酵素反応ステップをモデ

ル化し、計算機的に模倣することによって実現可能である。具体的には、与えられたクエリ化合物構造（特に外来化合物や二次代謝産物）に対して、前述の酵素反応ルール（原子変換パターン）を逐次適用し、次の候補化合物構造を得る一連のプロセスを自動化すればよい。これには膨大なグラフ計算を必要とするので、そのための対策として、計算サーバーの増強を適宜行う。

(4) 一般に上記のようなグラフの計算は組み合わせ的に解空間が広がることが分かっており、サーバーの増強だけで十分に対応できるものではない。そこで、同時にこの問題特有のヒューリスティクスの探索を行い、計算量の減少を可能にする手法を目指す。この探索は、既知の代謝パスウェイに似たパスウェイを候補パスウェイとしてなるべく採用するための新規スコアリング関数を開発することによって、ある程度達成可能であると思われる。

#### 4. 研究成果

(1) KEGG LIGAND データベースの一部として構築した酵素反応前後における化合物構造の原子変換パターンのデータベース RPAIR に関連して、これらの情報を計算機的に自動抽出するための新規アルゴリズムの開発を行い、RPAIR データベースそのものの改善を行った。この結果、RPAIR データベースに収録されていたエントリ数はほぼ倍増し、また、データ標記上の統一性/整合性も向上したため、より詳細な解析に用いることが可能となった。これらの結果を KEGG データベースの一部として公開した。次に、開発したアルゴリズムを用いて、酵素反応における基本的な反応知識ルールを抽出し、これらのルールを代謝パスウェイ全体に渡って統計を取るなどの処理によって、生体内において生物がより好んで用いる酵素反応のセットを見出した。

(2) 酵素反応メカニズムとして抽出されたルールを基に、酵素反応を予測するシステムを構築した。この際、予測に用いる原子変換パターンの情報をすべて利用するのか、一部のみ利用するのかについて任意性があるが、新たに構築した酵素番号予測システムでは、それらを組み合わせ、より最適なスコアの予測結果を出力出来るようにした。具体的には、まずすべての原子変換パターン情報を用いて予測を行い、その時点で予測結果が得られない場合には利用する情報を適宜減らして（すなわちパターンの一致条件を緩和して）予測を行っていく。それぞれの予測ステージにおけるスコアも相関係数を基にした重み付けスコア関数を用いており、従来型の

完全一致のみの予測システムよりも高精度化を達成出来た。さらに、データベースに登録されたデータのエラー（生化学的な不整合）を修正するにあたって、この予測システムを組み入れた修正プロトコルを内部向けに構築した。

(3) 酵素反応ルールとしてライブラリ化しておいた情報を、互いに組み合わせることによって、生体内パスウェイを再構成するためのアルゴリズム開発を行った。具体的には、RPAIR データベースに蓄積された代謝反応ルールを用いて、薬物の生分解経路の予測システムを開発した。これは、酵素反応ステップを計算機的に模倣することによって実現可能であり、与えられたクエリ化合物構造に対して、前述の酵素反応ルール（原子変換パターン）を逐次適用し、次の候補化合物構造を得る一連のプロセスを実行する。このシステムでは、クエリとして与えられた化合物構造に対してデータベース中の酵素反応ルールのうち適合可能なものを適用し、新たな化合物構造を生成していくため、通常操作では予測される反応経路は爆発的に増えていく。

(4) 上記の(3)の反応経路の探索空間の爆発のために、高精度に反応経路を予測するためには膨大なグラフ計算を必要とすることが分かっている。そこで、同時にこの問題特有のヒューリスティクスの探索を行い、計算量の減少を可能にする手法として、本課題では、クエリ化合物の全体構造とデータベース中の化合物構造の構造類似性に着目し、一定の類似がないものを省くことによって、この爆発を抑えることに成功した。すなわち、化合物全体の構造類似性が低い場合には、そのような酵素反応も起こらないであろうという考えである。このような手法の有効性については、既知薬物の構造をクエリとして、それらの既知反応経路が予測できるかどうかをもって検証しており、一定の効果があることを示している。また、既知の代謝パスウェイに似たパスウェイを予測結果の候補パスウェイとしてなるべく採用するためのスコアリング関数の開発も行った。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計5件)

① Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M., "E-zyme: finding enzyme orthologues from the chemical transformation pattern of substrate-product pairs",

Bioinformatics, in press, 2009, 査読有

② Shimizu, Y., Hattori, M., Goto, S., and Kanehisa, M., "Generalized reaction patterns for prediction of unknown enzymatic reactions", Genome Informatics, 20, 149-158, 2008, 査読有

③ Muto, A., Hattori, M., and Kanehisa, M., "Analysis of Common Substructures of Metabolic Compounds within the Different Organism Groups", Genome Informatics, 18, 299-307, 2007, 査読有

④ Oh, M., Yamada, T., Hattori, M., Goto, S., and Kanehisa, M., "Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways", Journal of Chemical Information and Modeling, 47, 1702-1712, 2007, 査読有

⑤ Saigo, H., Hattori, M., and Tsuda, K., "Reaction graph kernels for discovering missing enzymes in the plant secondary metabolism", NIPS Workshop on Machine Learning in Computational Biology 2007, 2007, 査読有

[学会発表] (計7件)

① Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M., "E-zyme: finding enzyme orthologues from the chemical transformation pattern of substrate-product pairs", 17th Annual International Conference on Intelligent Systems for Molecular Biology & 8th European Conference on Computational Biology (ISMB/ECCB2009), Stockholm, 2009.6.30

② 清水祐吾, 服部正泰, 五斗進, 金久實, 「未知の酵素反応予測に向けた一般化反応パターンの開発」, 第31回日本分子生物学会, 神戸, 2008.12.12

③ Shimizu, Y., Hattori, M., Goto, S., and Kanehisa, M., "Generalized Reaction Patterns for Prediction of Unknown Enzymatic Reactions", The 8th International Workshop on Bioinformatics and Systems Biology (IBSB2008), Berlin, 2008.6.9

④ Shimizu, Y., Hattori, M., Goto, S., and Kanehisa, M., "Comprehensive Analysis of Enzymatic Reaction Mechanisms Based on the EC Classification and KEGG RPAIR Database", The 2007 Annual Conference of Japanese Society for Bioinformatics (JSBi2007), Tokyo, 2007.12.17

⑤ Saigo, H., Hattori, M., and Tsuda, K.,

"Reaction graph kernels for discovering missing enzymes in the plant secondary metabolism", NIPS Workshop on Machine Learning in Computational Biology, Whistler, BC, Canada, 2007.12.7

⑥ Muto, A., Hattori, M., and Kanehisa, M., "Analysis of Common Substructures of Metabolic Compounds within the Different Organism Groups", The 7th Annual International Workshop on Bioinformatics and Systems Biology (IBSB2007), Tokyo, 2007.8.2

⑦ Hattori, M., "Atomic transformation patterns in enzymatic reactions based on the comparison of chemical compound structures", Asia Hub for e-Drug Discovery Symposium 2006 (AHeDD2006), Seoul, Korea, 2006.5.23

[図書] (計2件)

① Hattori, M., and Kotera, M., "To predict enzymatic reactions and metabolic pathways", in "Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques" (Edited by Dr. Lodhi, H. and Dr. Yamanishi, Y., Published by IGI Global in 2009)

② Kashima, H., Saigo, H., Hattori, M., and Tsuda, K., "Graph Kernels for Chemoinformatics", in "Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques" (Edited by Dr. Lodhi, H. and Dr. Yamanishi, Y., Published by IGI Global in 2009)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

なし

6. 研究組織

(1)研究代表者

服部 正泰 (Hattori Masahiro)

京都大学・化学研究所・助教

研究者番号: 60372554

(2)研究分担者

なし

(3)連携研究者

なし