

化学構造変換パターンに基づいた酵素反応分類データベースの開発

Development of the database for enzyme classification based on the reaction patterns

京都大学化学研究所 小寺正明

背景と目的

生体内での存在が確認されているがその生合成・生分解経路の不明な化合物 (orphan metabolites) は多く知られており、近年のメタボローム関連技術の発達によりその数はますます増えることが予想される。その経路で働く酵素群を予測するには、起こる可能性のある酵素反応を推定し、その生物種のゲノム配列との関連を解析する必要がある。このとき、推定される酵素反応式はいくつもの可能性が考えられ、一意には決められないことが多い。酵素反応式と酵素タンパク質 (または遺伝子) を関連づける国際標準的な酵素分類体系として IUBMB の enzyme list (EC numbers) があるが、これは酵素の存在が実験的に立証された完全な反応式の分類であるため、「推定の」「不完全な」反応式には使えない。逆に言えば「推定の」「不完全な」反応式を分類する方法の開発が orphan metabolites の代謝経路予測に必要であり、我々の目的である。

データと方法

酵素反応のデータは KEGG データベース (<http://www.genome.jp/kegg/>) から取得した。KEGG では、配列類似性に基づいたタンパク質分類である KEGG Orthology (KO) が構築されているが、今回報告する Reaction Class (RCLASS) は化学構造変換パターンに基づいた酵素反応分類であり、これによって酵素は配列類似性 (KO) と反応類似性 (RCLASS) の二つの視点から分類したことになる。その手順を以下に示す。

メタボロームデータ中の代謝化合物群の間をつなぐ酵素反応を推定するには、「不完全な」反応式に相当する化合物ペア (reactant pair) [1] の定義から始まり、その化学構造比較から反応中心 (reaction center) が計算される。化合物はグラフと呼ばれるデータ構造で表され、反応前後で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。その境界領域が反応中心であると見なせる。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 パターン (RDM chemical transformation patterns) として計算機的に表現される (図 1)。

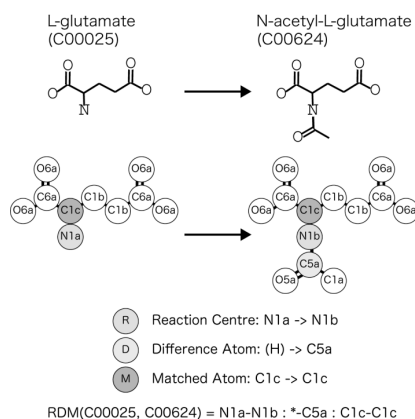


図 1、RDM パターン

結果

KO エントリの識別子 (K numbers) や反応式の識別子 (R numbers) にならい、RCLASS エントリにも識別子 (RC numbers) を与えた。ここで、RCLASS エントリの定義は「一つの化合物ペア (reactant pair) あたりの RDM パターンのセット」であり、ひとつの RDM パターンあたり、生成または消滅するひとつの化学結合を表現している。ひとつの化合物ペアあたり変化する化学結合はひとつとは限らないので、RCLASS エントリ中に含まれる RDM パターンの数もひとつとは限らない。これらは KEGG の BRITE 分類のひとつとして整備され、酵素遺伝子 (またはタンパク質) との対応付けを行っている。

2010年10月現在、KEGG データベース中の 12,334 RPAIR エントリから 2,302 RCLASS エントリを得た (図 2)。RCLASS エントリは RDM パターンを表す文字列だけでなく、その化学変換パターンを図示したグラフィクスを持っている。そこでは反応中心原子やその周辺が色づけで強調されている。また、様々な類似 (ではあるが同一ではない) 反応パターン

の向きができるだけ揃うように調整してあるので、反応中に構造がどう変化したのか、あるいは異なる反応パターンを人の目で把握し理解するのが容易になっている。

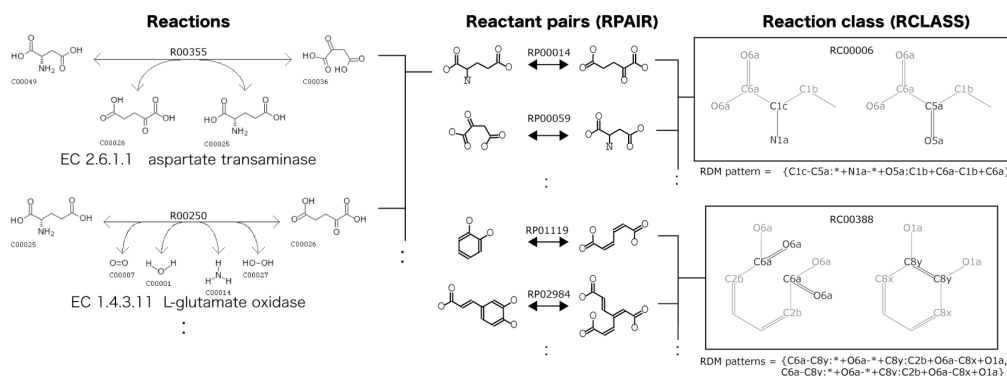


図2、RCLASS データベースと RCLASS エントリ

応用

本研究のように化学変換パターンと酵素遺伝子との対応関係を整理することにより、我々が以前に開発した E-zyme [1,2] を拡張し、メタボロームデータから予測される反応に対しゲノム中の酵素遺伝子と対応づけることを我々は考えている。また、当研究室の守屋らが開発した PathPred [3] では、この反応分類や既知の化合物ペアの情報を利用して、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のかを自動的に多段階予測することができる。

結辞

「omics」と総称される生物学的大規模データ解析の鍵となる行程のひとつにアノテーション (annotation) がある。たとえばゲノムアノテーションとは、ゲノム配列解析で得られた塩基配列 (またはアミノ酸配列) に生物学的情報を予測 (または実験的に同定) し、両者を関連づけることである。言い方を変えれば、ゲノムアノテーションとはゲノム情報と機能情報を結びつける行程である。酵素の場合、機能として化学反応を触媒する能力がある。現在のゲノムアノテーションは、主に配列類似性に基づき、「既知の反応」を結びつけるものである。RCLASSは「推定の反応」を取り扱えるようにすることで新しいタイプのアノテーション方法を作る可能性があると考えている。

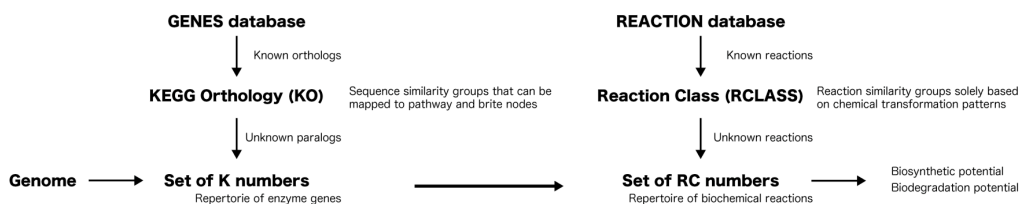


図3、酵素化学的な視点を加えたゲノムアノテーション

参考文献

- [1] Masaaki Kotera, Yasushi Okuno, Masahiro Hattori, Susumu Goto, Minoru Kanehisa. J. Am. Chem. Soc. 2004, 126, 16487-16498. [PMID: 15600352]
- [2] Yoshihiro Yamanishi, Masahiro Hattori, Masaaki Kotera, Susumu Goto, Minoru Kanehisa. Bioinformatics 2009, 25(12): i179-i186. [PMID:2687977]
- [3] Yuki Moriya, Daichi Shigemizu, Masahiro Hattori, Toshiaki Tokimatsu, Masaaki Kotera, Susumu Goto and Minoru Kanehisa. Nucleic Acids Research 2010, doi:10.1093/nar/gkq318. [PMID: 20435670]